



**Investigating the Washback Effect of a Revised EFL  
Public Examination on Teachers' Instructional Practices,  
Materials and Curriculum**

**Thesis submitted to University of Leicester for the degree of**

**Doctor of Education**

**in**

**Applied Linguistics and TESOL**

**By**

**Abdulhamid Mustafa El-Murabet Onaiba**

**School of Education**

**University of Leicester**

**November 2013**

## **Abstract**

### **Investigating the Washback Effect of a Revised EFL Public Examination on Teachers' Instructional Practices, Materials and Curriculum**

By

Abdulhamid Mustafa El-Murabet Onaiba

The phenomenon of how tests influence teaching and learning is commonly described in language education as "washback". The purpose of this study was to investigate how English teachers in Libyan schools were influenced by introducing a reformed EFL public examination, called the Basic Education Certificate Examination (BECE), in terms of their instructional practices—the *how*; and teaching materials and curriculum—the *what*. The study also sought to examine the effect of any other teacher and context-dependent variables on washback.

Three main aspects were studied: teachers' perception of the exam (perception washback); classroom teaching and testing practices (methodology washback); and teachers' choice, selection and use of teaching materials (curriculum washback). To address these issues, a mixed methods approach was utilized. Descriptive and inferential statistics were used to analyse quantitative data obtained from surveying 100 teachers. Content analysis was then conducted to interpret qualitative data elicited from documents, observations of two teachers and interviews with 11 teachers and 7 inspectors.

The study found that teachers expressed negative rather than positive views towards the exam. Findings indicated that the introduced exam did exert washback on teachers' instructional practices. However, washback was noticed in issues related to classroom testing practices rather than in teaching practices. As the new BECE did not represent the current curriculum, negative washback was observed on the content of the curriculum: some teachers tended to rely on the "hidden syllabus", while others narrowed the syllabus to meet the content of the exam. While some desired aims were achieved through the exam, others were not. The intensity and direction of washback was shown to be influenced by mediating variables such as the teacher and the context: the data indicated differences between veteran and novice teachers, and their level of education also affected their response to the exam. Gender was irrelevant. The grade level and class size taught was also associated with impact.

The findings of the study indicate how examination reform can be used as leverage of pedagogical change and to dictate the *how* and the *what* of teaching, but to different degrees. The study made it crystal clear that washback does exist, but that it operates in a complex manner associated with other variables besides the exam *per se*. Thus the study suggests that washback cannot be a theory because it seemed not to have predictive power. This study also suggests guidance for future policies for improvement of the examination system in Libyan schools, arguing for alignment of examinations and the English curriculum, with some recommendations as to how this may be achieved.

## **Acknowledgements**

Firstly, I am forever thankful to my God (Allah) who gave me the ability to complete my study. I sincerely wish to give my deepest gratitude to my supervisor Professor Glenn Fulcher for his kind and constant support and encouragement throughout the whole course of my research. He painstakingly read and commented on numerous texts of the thesis and offered invaluable feedback and stimulating discussions.

I would like to thank the professional services staff, at the School of Education, mainly Sharon Simpson and Leigh Blair, for providing all necessary help, with special thanks going to the director of school, Professor Janet Ainley, for her assistance in gaining the extension period of study.

My thanks also go to all teachers and inspectors who participated in the fieldwork of this thesis despite the difficult circumstances encountered. Their cooperation has immensely contributed to complete this study. I would also like to thank my friend Dr. Ilyas Khan for his help and suggestions on some texts of this work.

Also, I want to thank my family members, brothers and sisters, especially my late mother, whose endless love, support and prayers were the source of my energy, and helped keep me sane during undertaking this work. Also, unforgettably, I wish to express my endlessly deepest gratitude, thanks, love and respect to my beloved wife Khiria who gave unstintingly of her time and efforts to care for me and our children during my doctoral study. Without her patience and encouragement, it would have not been possible for me to reach this destination.

## **Dedication**

*To the souls of my late father and mother*

*To my beloved wife, sons and daughters*

*To all my brothers, sisters and dearest friends*

*To my home country, Libya*

## Table of Contents

Abstract .....	i
Acknowledgements .....	ii
Dedication .....	iii
Table of Contents .....	iv
List of figures .....	xi
List of tables .....	xii
List of Acronyms .....	xiv
<b>Chapter 1 Introduction: .....</b>	<b>1</b>
1.1 Background to the Research Problem .....	1
1.2 Research Aims and Questions .....	4
1.3 Context of the Study .....	6
1.3.1 The Education System in Libya .....	6
1.3.1.1 Overview, Schools and Stages .....	6
1.3.1.2 English Curriculum and the Applied Method of Teaching .....	9
1.3.2 The Examination System in Libyan Schools .....	15
1.3.2.1 Overview .....	15
1.3.2.2 Basic Education Certificate Examination (BECE) in English .....	18
1.3.2.3 The New Examination Policy .....	22
1.3.2.3.1 Exam Purposes and Aims (Intended Washback) .....	22
1.3.2.3.2 Exam Design and Format .....	24
1.3.2.3.3 Exam Content .....	26
1.3.3 EFL Teachers in Libyan Schools .....	28
1.3.4 EFL Inspectors in Libyan Schools .....	32
1.4 Rationale and Significance of the Study .....	34

1.5 Organization of the Thesis .....	36
<b>Chapter 2 Literature Review .....</b>	<b>38</b>
2.1 Introduction .....	38
2.2 Defining and Describing Washback.....	40
2.2.1 Definitions of Washback.....	40
2.2.2 Similar Concept Terms to Washback.....	44
2.3 Directions of washback—Negative or Positive.....	45
2.3.1 Negative Washback.....	46
2.3.2 Positive Washback .....	48
2.4 Mechanisms of Washback – Hypotheses and Models .....	50
2.4.1 Alderson and Wall Washback Hypothesis.....	51
2.4.2 Hughes’ Model.....	52
2.4.3 Bailey’s Model.....	53
2.4.4 Conceptual Framework of the Study.....	54
2.5 Washback Impact of Public Examination Change on Aspects of Classroom-related Areas.....	56
2.5.1 Washback Effects on Teachers' Feelings and Attitudes Towards Exams (Perception Washback) .....	56
2.5.2 Washback Effects on Instructional Practices (Methodology Washback) .....	61
2.5.3 Washback Effects on Teaching Materials and Curriculum (Curriculum Washback) .....	66
2.6 Factors Influencing Washback .....	70
2.6.1 Teacher-related Factors.....	70

2.6.1.1 Teachers Beliefs and Attitudes.....	71
2.6.1.2 Teaching Experience.....	72
2.6.1.3 Teacher Training on and Awareness of an Examination Reform.....	74
2.6.1.4 Academic Qualification .....	75
2.6.2 Exam-related Factors .....	76
2.6.3 Context-related Factors.....	78
2.7 Research Questions .....	80
2.8 Summary of the Chapter.....	83
<b>Chapter 3 Research Design and Methodology .....</b>	<b>86</b>
3.1 Introduction .....	86
3.2 Research Design.....	87
3.2.1 Research Philosophy and Paradigm.....	88
3.2.1.1 Paradigm .....	88
3.2.1.2 Ontology.....	89
3.2.1.3 Epistemology .....	89
3.2.1.4 Methodology .....	90
3.2.2 Key Philosophical Paradigms (Research Paradigm).....	91
3.2.2.1 Positivism.....	91
3.2.2.2 Interpretivism .....	92
3.2.3 Justification for Adopting a Mixed Approach to Research Paradigm .....	95
3.3 Research Methodology.....	98
3.3.1 Research Type and Strategy.....	99
3.3.1.1 Qualitative and Quantitative Approaches.....	99
3.3.1.2 Mixed Methods and Triangulation.....	101
3.3.1.3 Methodological Considerations from Washback Research.....	102
3.3.2 Sample and Methods of Data Collection .....	104
3.3.2.1 Sampling Frame and the Selection of Participants of the Study .....	104
3.3.2.2 Survey Strategy .....	111

3.3.2.2.1 Teacher Questionnaire.....	111
3.3.2.3 Case Study Strategy .....	116
3.3.2.3.1 Interviews.....	118
3.3.2.3.1.1 Examination Board Director Interview.....	121
3.3.2.3.1.2 Teacher Interviews.....	121
3.3.2.3.1.3 Inspector Interviews.....	122
3.3.2.3.2 Documentary Study .....	123
3.3.3 Methods of Data Analysis.....	124
3.3.3.1 Quantitative Data .....	125
3.3.3.2 Qualitative Data .....	126
3.3.4 Research Phases and Timeline.....	128
3.3.5 Pilot Study and Implications for the Main Study.....	131
3.3.5.1 Participant Selection and Sampling Frame.....	131
3.3.5.2 Piloting Teacher Questionnaires .....	132
3.3.5.3 Piloting Interviews .....	134
3.3.5.3.1 Teacher Interview.....	134
3.3.5.3.2 Inspector Interview .....	136
3.3.5.4 Piloting Classroom Observation.....	137
3.3.5.4.1 Rationale.....	137
3.3.5.4.2 Procedure.....	138
3.3.5.4.3 Instruments .....	139
3.3.5.4.4 Classroom Observation in the Pilot Study .....	140
3.4 Validity and Reliability Issues .....	142
3.5 Ethical Considerations.....	143
3.6 Accessibility of Research Site and Participants .....	145
3.7 Turmoil and Unrest in the Context of the Study .....	147
3.8 Data Collection Procedures, Challenges and Difficulties .....	150
3.9 Summary of the Chapter.....	157
<b>Chapter 4 Presentation of Results .....</b>	<b>159</b>
4.1 Introduction .....	159



4.2 Results and Analysis .....	159
4.2.1 Teachers' Perceptions of and Reaction to the New Exam .....	160
4.2.1.1 Teachers' Perceptions of the Reasons behind Implementing the New Exam .....	160
4.2.1.2 Teachers' Attitudes towards the New Exam and Its Quality.....	165
4.2.1.3 Teachers' Perceptions of the Exam Concerning Their Classroom Practices.....	170
4.2.2 Washback on Teachers' Instructional Practices.....	172
4.2.2.1 Teachers' Teaching Practices.....	172
4.2.2.1.1 Teachers' Teaching Methods and Techniques.....	173
4.2.2.1.2 Teachers' Classroom Teaching Activities .....	180
4.2.2.2 Teachers' Classroom Testing Practices.....	182
4.2.3 Washback on Teaching Materials and Curriculum .....	187
4.2.3.1 Teachers' Choice of Teaching Materials, Learning Resources .....	188
4.2.3.2 Teachers' Arrangements and Selection of Teaching Content and Materials.....	190
4.2.4 The Degree of Washback Effect vis-à-vis teacher and context factors.....	197
4.2.4.1 Qualification effect.....	201
4.2.4.1.1 Teachers' teaching practices, teaching methods and techniques.....	201
4.2.4.1.2 Teachers' Classroom Testing Practices and Evaluation, On-going Tests .....	203
4.2.4.1.3 Teachers' Choice, Selection and Use of Teaching Materials and Curriculum.....	204
4.2.4.2 Experience Effect .....	206
4.2.4.2.1 Teachers' Teaching Practices, Teaching Methods and Techniques .....	207
4.2.4.2.2 Teachers' Classroom Testing Practices and Evaluation, On-going Tests .....	211
4.2.4.2.3 Teachers' Choice, Selection and Use of Teaching Materials and Curriculum.....	213
4.2.4.3 Grade Effect .....	215
4.2.4.3.1 Teachers' teaching practices, teaching methods and techniques.....	215
4.2.4.3.2 Teachers' Classroom Testing Practices and Evaluation, On-going Tests .....	217
4.2.4.3.3 Teachers' Choice, Selection and Use of Teaching Materials.....	218
4.2.4.4 Classroom Size Effect .....	220
4.2.4.5 Gender Effect .....	222
4.2.5 Documentary Study.....	228
4.3 Summary of the Chapter.....	236
<b>Chapter 5 Discussion .....</b>	<b>241</b>
5.1 Introduction .....	241

5.2 Synthesis of the Research Findings.....	243
5.2.1 Discussions on Findings to Research Question One.....	243
5.2.2 Discussions on Findings to Research Question Two .....	247
5.2.3 Discussions on Findings to Research Question Three .....	252
5.2.4 Discussions on Findings to Research Question Four .....	256
5.2.5 Discussions on Findings to Research Question Five .....	260
5.2.6 Discussions on Findings to Research Question Six .....	263
<b>Chapter 6 Conclusions.....</b>	<b>266</b>
6.1 Recapitulation.....	266
6.2 Implications and Contributions .....	269
6.2.1 Local Context.....	269
6.2.2 Relevant Literature.....	271
6.3 Limitations of the Study .....	275
6.4 Recommendations and Suggestions for Future Research .....	277
References .....	280
Appendices .....	297
Appendix 1 A sample of the Old BECE in English .....	297
Appendix 2a A Sample of the New BECE in English .....	301
Appendix 2b The Answer Sheet Attached to the New BECE in English .....	307
Appendix 3 An Overview of most Washback Studies Addressed in this Study ..	308
Appendix 4 Teacher Questionnaire for the Main Study.....	312

Appendix 5 Teacher Questionnaire for the Pilot Study .....	317
Appendix 6 Teachers' Interview Questions .....	323
Appendix 7 Inspectors' Interview Questions .....	324
Appendix 8 COLT Classroom Observation Scheme.....	325
Appendix 9 Observation Note-taking Sheet.....	326
Appendix 10 Ethical Approval to Conduct and Pursue the Thesis Stage .....	327
Appendices in the accompanying CD .....	328

## **List of figures**

Figure 1.1 Libyan education system, schools and stages.....	7
Figure 2-1 Correspondence between Hughes' trichotomy model and Alderson and Wall's washback hypotheses .....	53
Figure 2-2 A basic model of washback.....	54
Figure 3-1 The research "onion" model .....	97
Figure 3-2 A road map of the research: overall methodology. ....	98
Figure 3-3 Selection of participants and data triangulation .....	106
Figure 3-4 Research stages and timeline.....	130

## **List of tables**

Table 3-1 Comparison of the basic elements of positivism and interpretivism .....	94
Table 3-2 Demographics of surveyed teachers .....	108
Table 3-3 Characteristics of interviewed teachers .....	110
Table 4-1 Teachers' perceptions of the reasons for implementing the new exam....	161
Table 4-2 Teachers' awareness of key differences between new and old exams .....	164
Table 4-3 Teachers' beliefs and attitudes towards the new exam.....	166
Table 4-4 Extra work or pressure involved in teaching towards the new exam .....	171
Table 4-5 Language areas in which teachers think it is better to assess students .....	172
Table 4-6 Alterations to teaching methods and techniques due to exam change.....	174
Table 4-7 Teacher talk .....	175
Table 4-8 Degree of similarity between teachers' teaching in grade 9 and in other grades .....	179
Table 4-9 Teachers' classroom teaching activities .....	181
Table 4-10 Learning strategies in the context of the new exam .....	183
Table 4-11 The consistency between the content of teachers' on-going tests and the content of the new exam .....	185
Table 4-12 The consistency between teachers' on-going tests to grade 9 and other grades .....	186
Table 4-13 Teaching and learning resources and aids .....	189
Table 4-14 Teachers' arrangements and planning of teaching content and materials .....	190

Table 4-15 Learning strategies in the context of the new exam .....	191
Table 4-16 Teachers' selection of the content of teaching regarding the current textbooks .....	194
Table 4-17 Results obtained from one-way ANOVA with eta squared values for qualification in relation to teaching practices. ....	203
Table 4-18 Results obtained from one-way ANOVA with eta squared values for qualification in relation to teachers' classroom testing practices.....	204
Table 4-19 Results obtained from one-way ANOVA with eta squared values for qualification in relation to teaching materials and curriculum use. ....	206
Table 4-20 Results obtained from one-way ANOVA with eta squared values for experience in relation to teachers' teaching practices.....	209
Table 4-21 Cross-tabulation: Academic Qualification - Teaching Experience .....	210
Table 4-22 Results obtained from one-way ANOVA with eta squared values for experience in relation to teachers' classroom testing practices .....	212
Table 4-23 Results obtained from one-way ANOVA with eta squared values for experience in relation to teachers' selection of teaching materials and curriculum .....	214
Table 4-24 Results obtained from one-way ANOVA with eta squared values for grades effect in relation to teachers' classroom teaching practices .....	217
Table 4-25 Results obtained from one-way ANOVA with eta squared values for the grade factor in relation to teachers' classroom testing practices.....	218
Table 4-26 Results obtained from one-way ANOVA with eta squared values for the grade factor in relation to teachers' selection of teaching materials and curriculum use. ....	219
Table 4-27 The descriptives generated from the one-way ANOVA for grade effect in relation to teachers' selection of teaching materials and curriculum use.....	220

### **List of Acronyms**

<b>ASL</b>	Arabic as a second language
<b>BA</b>	Bachelor of Arts
<b>BE</b>	Bachelor of Education
<b>BECE</b>	Basic Education Certificate Examination
<b>BERA</b>	British Educational Research Association
<b>DT</b>	Diploma of Teaching
<b>EFL</b>	English as a Foreign Language
<b>ESL</b>	English as a Second Language
<b>ET</b>	English Test
<b>HKCEE</b>	Hong Kong Certificate of Education Examination
<b>L1</b>	First/Mother Tongue Language
<b>IELTS</b>	International English Language Testing System
<b>ME</b>	Ministry of Education, the highest authority of education in the context
<b>NATO</b>	North Atlantic Treaty Organization
<b>SUEE</b>	Spanish University Entrance Examination
<b>SPSS</b>	Statistical Package for the Social Science

**TEFL** Teaching of English as a Foreign Language

**TESOL** Teaching of English to Speakers of other Languages

**TOEFL** Test of English as a Foreign Language

**UAE** United Arab Emirates

**UN** United Nations

**UEE** University Entrance Examination

**WASL** Washington Assessment of Student Learning



## Chapter 1 Introduction:

---

### **1.1 Background to the Research Problem**

[W]e need to investigate washback for every context in which test use is hypothesized to impact upon the process and content of teaching and learning (Fulcher and Davidson, 2007: 229).

Many theoreticians and practitioners in the field of language education have emphasized the impact of tests, especially public examinations and high-stakes tests. This means that these tests are designed to impact on teaching and learning. This influence, a phenomenon known in language education as washback (Alderson and Wall, 1993) is bidirectional—positive or negative—depending largely on several mediating factors, one of which is the quality of the test being used, as reported in forthcoming chapter.

Two types of high-stakes English language tests are administered worldwide: international proficiency English tests, such as IELTS or TOEFL; and standardized achievement public examinations introduced in EFL/ESL contexts of which the latter is the main concern in this thesis. Washback from examinations has intrigued educators and researchers, and has been investigated widely over the past two decades. However, this phenomenon had not been investigated empirically until Alderson and Wall (1993) published their seminal paper with a titular question: “Does washback exist?”, based on their research on the impact of high-stakes public examinations in Sri Lanka, which posed 15 hypotheses of washback (see section 2.4.1). References to washback research have been made ever since by various researchers in different contexts; employing diverse methods (see section 2.5).

Public examinations have been exploited as “levers for change” (Pearson, 1988: 107) in order to induce positive washback and promote curricular innovation (Wall and Alderson, 1993; Qi, 2004; Cheng, 2005), though they may have negative repercussions for teaching and learning (Shohamy *et al.*, 1996). Thus, recent washback studies in language education concentrate on the effects of public examinations on their stakeholders, particularly when the examinations undergo changes, in aspects such as learning, teaching methods, curriculum and attitudes towards tests (Qi, 2004; Saif, 2006; Tsagari, 2009a). In the area of instructional change, it was suggested that adopting valid and reliable tests is one of the main remedies of the widely-spread entrenched/traditional instruction, which is commonly based on rigid ways of teaching and learning, such as memorization, rote recall and citation of discrete facts and texts.

Moreover, public examinations can contribute to the process of educational innovation, influencing classroom procedures (Shohamy, 2001; Brindley, 2008). Empirical studies reveal varied findings regarding the influence of tests on different aspects of classroom instruction: some tests may have greater consequences for some areas of teaching and learning than others (*washback intensity*; Cheng, 1997: 43). However, it has been asserted that washback is a complex and multidimensional phenomenon (Alderson and Wall, 1993; Choi, 2008; El-ebyary, 2009, among others), and should not be considered an automatic or direct effect of exams (Bailey 1999; Spratt, 2005). The literature, as detailed in section 2.6, claims that some intervening variables beyond the exam *per se* may contribute to determining or precluding the amount and kind of washback impact. Some of these variables relate to teachers' characteristics, learners' characteristics and context (Spratt, 2005). These varied findings may indicate that each introduced exam needs tailor-made research to identify its washback.

Research also shows that high-stakes public examinations may from the outset affect some stakeholders' perceptions and attitudes (Cheng, 2005), and hence, they may be capable of changing teachers' classroom behaviour and teaching content (Shohamy *et al*, 1996). However, this change in the *how* (methodology) and the *what* (teaching content) is sometimes superficial rather than substantial, and the change may occur in the form of teaching and not in its substance (Cheng, 1997; Qi, 2004). In this respect, Shohamy (2001) asserts that public examinations are powerful enough to influence teachers' classroom behaviour. However, the scope and nature of this influence are still uncertain, and still under investigation.

## 1.2 Research Aims and Questions

Since “our understanding of the role of high-stakes tests as an agent for change is limited due to the small number of empirical studies available on this issue” (Qi, 2007: 51) and that “any given test needs research tailor-made to examine its washback” (Shih, 2007: 137), this research aimed to investigate the degree of change (the *how* and the *what*) Libyan teachers of English may make as a result of introducing a new high-stakes exam. Thus, the overarching aim of this study was first to explore teachers’ perceptions of the introduced exam in relation to their classroom practices, then to explore the actual influence of this examination on teachers' teaching practices, materials and curriculum in order to see what kind of washback, if observed to take place, this exam has brought about. A further aim was to discover how teacher and context-related factors may contribute to determining or precluding the amount and kind of washback impact.

Additionally, an important purpose was to explore whether *test retrofit* (Fulcher and Davidson, 2009: 123) i.e. exam change for one purpose or another would warrant the fulfilment of the desired consequences for an educational innovation, or whether unintended/negative consequences would be observed.

In the light of the description of the context of the research problem and aims of the study, this research will attempt to answer the following questions:

1. What is the nature and scope of the washback effect on grade 9 teachers’ perceptions of aspects of teaching in relation to the new examination?

2. What is the nature and scope of the washback effect on teachers' classroom instructional behaviours as a result of implementing the new revised examination?
3. What is the nature and scope of the washback effect on the content of the curriculum of English, as the change has been only on the testing syllabus?
4. To what extent do the teacher and exam context characteristics have an influence on the kind and degree of washback that materialized?

The answers to the above four questions will contribute towards answering two further questions, question five and question six.

5. To what extent did the new revised exam exert the intended washback envisaged by its constructors, and are there any other unintended consequences?
6. What quality of washback effect—negative or positive—did the new revised exam induce on EFL classrooms in Libyan schools?

Based on reflections coming from literature review and on methodological considerations discussed in section 3.3.2.3, the research questions are further outlined in detail in Chapter Two section 2.7.

To achieve the research aims and find possible answers to the research questions, a mixed-method research design was utilised.

Before embarking upon investigating these issues which are reported in Chapter Three, it is important to acquaint the reader with background information about the context in which this study was conducted, as this is deemed an important element in washback studies (Cheng, 1997; Wall, 2000; Watanabe 2004). This introduction,

therefore, first describes the context for this study in detail, including the education system and the examination system. Then, it provides background information about EFL teachers and EFL inspectors in Libyan schools. Further comments will detail the rationale for the research study, and then the organisation of the thesis will be outlined.

### **1.3 Context of the Study**

This study was conducted in Libyan Basic Education public schools, specifically Grade 9 classes. This section discusses the education system, including education stages and the curriculum prescribed, its objectives and content. It then explores the examination system, reviewing the new exam in terms of objectives and content, in order to allow conclusions to be drawn on the relationship between the two.

#### **1.3.1 The Education System in Libya**

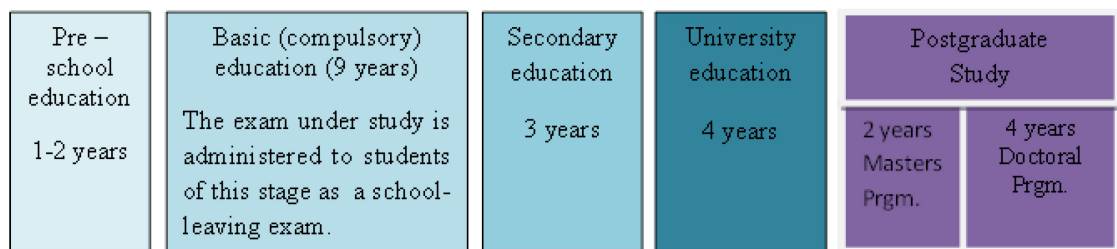
##### **1.3.1.1 Overview, Schools and Stages**

The education system in Libya is centralised, following top-down policy, i.e. is operated under the purview of the Authority of Education (i.e. Ministry of Education, henceforth as ME). This includes the preparation, distribution and monitoring of teaching and examination materials. All state schools have comparable educational facilities and resources. Schools are run on a term basis, the first term commencing in September and ending in early January; the second term starts in February and ends in May. Students have a two-week holiday between the two terms. Students attend schools six days a week (from Saturday to Friday), studying approximately six sessions a day, of forty minutes a session. English is taught as a foreign language

(EFL) four sessions a week. The teaching of English, which formally starts from the 5th grade of Basic Education at age 10, is practiced within a context-restricted environment, in which the determiners of language learning depend on classroom activities instructed by the classroom teacher but based on teaching materials prescribed by the ME (see 1.3.1.2). The ME compiles, develops and publishes textbooks and teaching materials for schools nationwide, thus the curriculum in this setting is a uniform one across all school subjects, especially in the Basic and Secondary Education Stages.

The Libyan educational system, as shown in figure 1.1, comprises five stages: a) pre-school education, which is optional and involves children from 4-5 years old; b) basic education, for students aged from 6-15; c) three-year secondary education, which involves students from 16-18 years old who have successfully finished basic education; d) university education for students who have finished secondary education; and e) postgraduate study. The academic year in all these education stages is usually eight months, from September to May. In Libya, education is free for all in basic, secondary and university institutions.

Figure 1.1 Libyan education system, schools and stages



The Basic Education generally “aims at providing the pupil with necessary principles, behaviour, knowledge expertise and practical skills” (Otman and

Karlberg, 2007: 101), and “learning foreign languages [i.e. English] to communicate with the world” (ME, 2008: 6). In this stage, students are expected to study a range of subjects one of which is English language. By the end of this stage all students sit national public examinations for all subjects they have studied, all of which they must pass with a grade of at least 50% in order to obtain the Basic Education Certificate (ME, 2008). The Basic Education Certificate Examination in English, which is the focus of this study, is one of these exams.

After gaining the Basic Education Certificate, students are eligible for three years of Secondary Education. The ME in Libya promulgated multiple schemes for students who successfully completed their basic education to enter secondary education. Secondary education is divided into various divisions of study: basic sciences; engineering sciences; life sciences; economic sciences; and a division of languages which includes disciplines of Arabic and English (ME, 2008). In English language, for instance, students study various subjects and skills, including listening, speaking (both supplemented with language laboratory sessions), phonetics, reading, writing and grammar. As it is given high social prestige, English language specialization as a school major is encouraged by many parents. However, students must successfully pass the Basic Education Certificate Examination (see section 1.3.2.1), and gain at least an average score of ‘good’ (65%) in English language in order to study English specialization in secondary education (as shown in appendix 11, under number 2, F). This requirement force students to excel to achieve this grade.

At the end of secondary education, students sit another high-stakes public examination to gain the Secondary Education Certificate to enable them to register at



university to study their chosen discipline or division. For example, students who choose English specialization as a major in secondary education can enter the department of English; in the faculties of Arts or Education, in which they study a range of English subjects (see section 1.3.3). Those students wanting to enter Faculties of Education must achieve 75% in their secondary education examination, and those who wanting to attend Faculties of Arts must achieve at least 65% (see appendix 12, point 6 - b)

### **1.3.1.2 English Curriculum and the Applied Method of Teaching**

The purpose of this section is to explore any concordance between the objectives and the content of the curriculum of the basic education schools and the objectives and the content of the new examination, utilized as document analysis.

It is important to note that the eleven years of work as a teacher of English in Libyan basic and secondary schools working and dealing with the current English curriculum provided the current researcher with an intimate knowledge of the Libyan education and examination system. Thus, in essence, it contributed to obtaining a clear review of the current English curriculum and a clear analysis of and distinction between the old and new examination patterns.

Due to the shift in language teaching, focusing on the functional use of language, in 2000 a new English curriculum was introduced to Libyan Basic Education and Secondary schools (Orafi and Borg, 2009). A perusal of the content of the new curriculum shows that it encompasses a series of textbooks called *English for Libya*. For the basic education schools, the syllabus comprises five levels of English language written specifically for Libyan students from Grades Five to Nine

(Grade Nine textbooks were previously known as *Preparatory 3*). Other levels of the same series are assigned for Grades One, Two and Three of secondary education. The syllabus was published by Garnet Ltd, Reading U.K., and written by Lucy Frino *et al*, under the supervision of The Libyan Authority of Education. For both levels, i.e. basic education and secondary education, the prescribed English syllabus comprises: a full-colour course-book, a black and white workbook, a teacher's book and a class cassette

The course-book, the main resource for implementing instruction, is divided into ten units based around themes such as the environment, achievements and technology, which present new vocabulary, structures and functions and contain texts to develop language skills (Appendix 34 shows a unit example copied from the course-book). Each unit comprises eight lessons, most of which begin with activities which are accompanied by related tasks in the workbook. "The final lessons of each unit are designed to stretch students by giving them more 'roughly tuned' texts to read or listen to, or more demanding speaking and writing activities" (Frino *et al.*, 2008: 6). The last unit, unit ten, revises language from the previous units and lessons, and offers students a chance to use the target language with more self-reliant tasks, such as writing at paragraph level.

The workbook, which supplements the course-book, contains consolidation and practice exercises to cover the language lessons studied in the course-book. The students need both books for most lessons. "The Workbook exercises also enable the teacher to check how far the students have mastered language items and allow further skills development" (Frino *et al.*, 2008: 6). According to the authors, the tasks

become more challenging as the students' overall language mastery progresses. The teacher's book contains a variety of lesson plans and lesson-by-lesson guidance for teachers. It outlines the aims of each lesson/unit and shows the language constructs to be covered, and how they are to be dealt with by teachers. The cassettes are mainly utilized for the oral and aural activities stated in the course-book and workbook.

The objectives of the current curriculum as outlined by Orafi (2008:15) are:

- For the students to leave school with a much better access to the world through the lingua franca that English has become.
- To create an interest in English as a communication tool, and to help students develop the skills to start using this tool effectively.
- To help students use the basic spoken and written forms of the English language.
- To help students learn a series of complex skills: these include reading and listening skills that help get at meaning efficiently, for example, skimming and scanning and interpreting the message of the text; they also include the speaking and writing skills that help the students organize and communicate meaning effectively.

Additionally, the target language is organized and introduced to students according to topic rather than structure, to encourage students to link language and functions to familiar topics and situations (Frino et al., 2008). The course-book sections are "dedicated to reading, vocabulary and grammar, functional use of language, listening, speaking and writing" (Orafi and Borg, 2009: 245). The curriculum recommends that English be used as much as possible in class with the aim of enabling students to communicate effectively and fluently (Orafi and Borg, 2009: 245) (Appendix 13 outlines the course general objectives and the textbooks prescribed according to the ME at this stage of language learning).

At the reading stage, students are introduced to the themes of the units through the reading of texts. These include information texts from web sites and magazines, a biography or autobiography, a newspaper article, diary entries, letters and e-mails. Vocabulary related both to the reading text and the theme of the unit is introduced at this stage, either before or after the text is read. Students are asked to discuss ideas, and work to answer previously text-based comprehension questions.

For the vocabulary and grammar stage, the course-book builds on the lexical syllabus from the course-books studied in the previous two years. “It introduces sets of key vocabulary in each unit and focusses on correct pronunciation and spelling as well as emphasising word grammar and collocation. It also emphasizes the categorisation of vocabulary to help students organise their lexical knowledge” (Frino, 2008: 6), and enables students to build up areas of vocabulary arising from the studied passages. Work on grammar follows, by introducing revisions to previous rules studied, and the presentation of new language structure which often derives from the patterns stipulated in the listening and reading texts studied (i.e. the grammar is taught functionally) through the Look! grammar box feature, highlighting grammatical patterns deductively (Frino, et al., 2008: 7). “The grammar is always presented in context and is recycled throughout the book in other contexts [with an aim for students] to notice the grammatical patterns repeatedly and practice them frequently before students can acquire the grammatical rule” (Frino et al., 2008: 6). Next, activities that concentrate on communicative functions of language comprise the function and listening stage. This is followed by aural activities to expand the previous activities. For example, students are asked to complete a range of different tasks such as listening or matching conversations to pictures, or listening to verify

ideas or answers after discussion of a question or problem. The teacher's role is to facilitate the process.

Finally, the production stage, speaking and writing, takes students into more demanding language tasks. Students are encouraged to use interactive activities and pair work, and practice speaking through simple oral tasks and activities, such as role-plays, discussions, information gap and problem solving. These activities "involve students using a prompt or model to produce their own sentences or short exchanges to opportunities for freer speaking, where students create their own conversations or discussions, generally using similar language patterns to the model or example in the book" (Frino et al., 2008: 7). At this stage, teachers are recommended to be patient regarding learners' errors during these activities. At the end, students do the writing section and, with the assistance of their teachers, carry out writing tasks to topics such as "the weather in Libya", "days off school" and "health and food". At this stage students are encouraged by the language being situated in contexts that they will be studying in other school subjects, such as science, history and geography. These tasks are conducted either in class, as homework, or both.

In light of the above discussion, it is evident that the current English curriculum in Libyan schools, particularly at the Basic Education Stage, is communicatively oriented. The intention is to promote communicative language teaching, and as a result it is highly recommended for teachers of English in Libyan schools to adhere to the principles and assumptions of teaching upon which the communicative approach to language teaching is based. This was further evidenced by

Ghuma (2011: 35), who affirms that the “Communicative approach is the current teaching method employed in these textbooks. Using communicative language teaching implies using activities that encourage communication and urging students to communicate”. This curriculum is generally welcomed by teachers and inspectors (Shihiba, 2011); however, these “positive views [specifically from teachers] towards communicative activities were not translated into classroom practices” (Orafi and Borg, 2009: 249).

Despite the establishment of new trends in classroom teaching and learning settings by introducing the new curriculum, Grammar-Translation method of language teaching seems still to hold the ground. The teachers, especially senior and untrained teachers, have not been able to come out of the influence of the method implied in the old curriculum.

According to some related research studies conducted in the Libyan context, some barriers challenge the implementation of this curriculum, relating to individual, contextual and socio-cultural considerations (Ali, 2008; Orafi, 2008; Orafi and Borg, 2009; Shihiba, 2011; Elabbar, 2011). For example, Orafi (2008), who investigated the implementation of the current English curriculum by five Libyan EFL teachers based on classroom observations and pre- and post-observation interviews, reported minimal alignment between curricular innovation intentions and pedagogical realities pertinent to teachers’ instructional practices. He identified three factors influencing teachers to implement the curriculum properly: a) teachers’ beliefs about language teaching and learning, students and teachers; b) teachers’ teaching experience in English language teaching and learning being shaped by their prior experiences of

English language teaching and learning; and c) contextual factors relating to classroom expectations, the examination system, teacher training and the development of the inspector's role. Ali (2008) concludes that his “study points very clearly to a limited uptake in the English language curriculum being implemented in the Libyan school system” (227) (For more discussion see Orafi, 2008: 204-223; Ali, 2008: 269-270; Orafi and Borg, 2009: 252).

### **1.3.2 The Examination System in Libyan Schools**

This section discusses the examination system, focussing on the Basic Education Certificate Examination (BECE) in English. It outlines the reasons for replacing the old BECE with a new one, the exam studied in this research. It also offers a brief description of both exam patterns, with a special emphasis on the new BECE in terms of aims, format and content.

#### **1.3.2.1 Overview**

According to their report to the International Conference on Education (2008) in Geneva, the ME (known previously as The General Peoples' Committee of Education) specified general goals of the examination system, that “the examination system is a tool used by which to assess the output of the educational process and to determine whether the student is capable of apprehending the curriculum taught during the school year ...” (ME, 2008: 11) (see section 1.3.2.3).

Assessment in the Libyan education system is managed by the ME through the administration of curricula which determine the criteria for evaluating students' learning and progress throughout the school year (ME, 2012a: 30). Schools receive

an official document every year outlining instructions, including students' appointed mark divisions and distributions. Appendix 14 shows a copy of this document assigned for the Grade 9 English classes at the Basic Education Stage. The document indicates the summative orientation of the assessment process in the education system. Hence, examination marks and grades, both formative and summative, represent the most common assessment strategies used in Libyan basic and secondary education. Examinations which largely test recall based on the memorization of grammar and vocabulary appear to shape classroom practices (Alhmali, 2007; Orafi and Borg, 2009).

English Language examinations in Libyan schools are very important, because students' educational futures depend heavily on their scores in these examinations. Two major high-stakes public examinations are administered to schools by the end of each school year as end-of-stage exams. One exam given to grade 9 basic education students (the Basic Education Certificate Examination, or BECE), is the focus of this research. The other exam is administered to grade 3 students of secondary education (i.e. Secondary Education Certificate Examination, or SECE) (it is important to note here that these examinations are not only restricted to assessing students' English language as a subject matter, but also all other school prescribed subjects). Passing these examinations is a prerequisite for students to proceed to the subsequent stages of education, the former to proceed to secondary education and the latter to enter university. Section 1.3.2.2 is devoted to the first type, particularly in English, as it is the main concern in this research.



The ME draws up implementation policies for entering secondary schools. Basic education school graduates can be assigned to secondary schools via special selections (based upon the students' special talents and desires) and, most importantly, test scores (based upon the BECE). Those who fail the exam have another chance to pass it by taking a reset exam administered within two weeks of the distribution of the main exam results. Students who then fail the reset exam must repeat the year in order to retake the exam and pass it, if they want to proceed to the next education stage.

Examinations of English (and of all other subjects) in Libyan schools are achievement tests, in that their content are solely based on the content of the materials taught, i.e. the existing curriculum of English. Regarding grade 9 of basic education and grade 3 of secondary education, these exams are classified into two main schemes.

***Scheme One, internal exams:***

These are in-class tests designed and administered by teachers. Usually, these are class-size quizzes or tests administered autonomously by teachers. They are formative tests—"short tests that are administered throughout the school year [either on weekly, monthly or periodically bases] to help instructors respond quickly to student weaknesses" (Toch, 2006: 13). Students sit at least five monthly tests and two periodical tests (the latter are also called first session and second session exams) during the school year based on the materials they studied before the time of each test took place. According to ME (2012b: 30), student marks in these tests, plus marks gained from homework and participation in classroom activities shown in

appendix 14, are allocating 40 marks out of 200, hence 20% of the total mark assigned for the subject in the school year, the remaining 80% of the total mark is devoted to the second scheme, discussed in the next subsection.

***Scheme Two, external exams:***

These are final written examinations, large-scale public exams, prepared and administered summatively by the relevant personnel of the ME only to grade 9 of the basic education and grade 3 of the secondary education. They are achievement tests, quantifying students' achievement in the teaching materials they studied during the whole school year. In basic education schools, this exam is considered very important for students as it represents the majority of their marks (i.e. 160 out of 200, or 80%) for the school year. Importantly, for a student to be counted successful, he or she must sit the final examination and achieve at least: a) 40% of the required mark of the final examination (BECE), i.e. 64/160, and; b) 50% of the total required mark in the subject, i.e. 100/200 (see appendix 14). Students who want to study English specialization in secondary education (see 1.3.1.1 above, and appendix 11) must achieve at least 65% in English, i.e. 130 marks out of 200, (see appendix 14). Therefore, concerned students as well as their parents consider the BECE in English essential regarding successful graduation, as it qualifies students to complete their basic education and obtain a certificate which allows them to proceed to secondary education (ME, 2008).

**1.3.2.2 Basic Education Certificate Examination (BECE) in English**

The BECE in English is the national teaching and textbook-oriented examination administered to Grade 9 basic education students as a stage-leaving exam. The exam

takes place in May annually, and is a pencil-and-paper exam. For technical and confidentiality purposes, different versions of the exam are constructed by some selected experienced inspectors. A panel of English subject experts at the ME select one exam for the first session and another for the second, as a reset exam. The panel ensure the exams are identical in terms of format, design and length, and similar in terms of difficulty and content. In this respect, exam writers are governed by a set of procedures and regulations assigned by the concerned parties of the ME, which are stemmed from test objectives (as for the new examination syllabus, see section 1.3.2.3 below).

Before the new BECE pattern was introduced in 2009, English tests in Libyan schools were traditional in nature, making use of mainly textbook-based questions on reading, vocabulary, grammar and composition (appendix 1 contains a sample of the old exam). The reading comprehension passage, which was mostly seen, usually consisted of a short story or a factual paragraph with open-ended, *wh*-questions, which could be answered by lifting sentences from the given paragraph. The exam questions were introduced in a randomized order, in that the exam was not split into sections in terms of the constructs it assessed.

With the old examination pattern, students answered exam questions on the same exam paper, and did not need an answer sheet. Examination papers were rated and the results were prepared and published manually by recruiting large numbers of teachers and inspectors.

This examination system, particularly that relating to English language, was criticized by Onaiba (2006) and Alhmali (2007) in terms of its content, and for its

emphasis on “the rote recall of information and holds great power over the learners at key times of the year” (Alhmali, 2007: ii). These exams were based on both constructed-response and selected-response testing items, albeit the former outweighed the latter. Students were encouraged to memorize huge chunks of material and then regurgitate them for their final exams. Students were evaluated on what they were able to recall and reproduce rather than on what they integrated and produced (Onaiba, 2006).

These examinations were described as being a hybrid of both integrative construct-response items and discrete-point/selected-response items (Onaiba, 2006; Orafi and Borg, 2009). The old exams pattern (see appendix 1) mostly introduced question items of reading comprehension, vocabulary, grammar and writing. Responses to these items, as stated by Orafi and Borg (2009: 252), were "based on memorisation of grammar and vocabulary". Also, subjective test items were heavily used as a test method to assess students' receptive skills, particularly reading comprehension, vocabulary and grammar, rather than productive skills (Onaiba, 2006), posing problems in marking exam papers, which was subjective. This diminished the validity of test items as the items set did not seem to measure what they were intended to measure (Brown, 2006). Additionally, two important skills—listening and speaking—were completely ignored in the previous testing syllabus, which in turn made teachers ignore teaching them in their classroom instruction – although the prescribed English syllabus is communicatively based (Orafi and Borg, 2009). It therefore appears that there was a mismatch between testing and teaching syllabi, which in turn had detrimental effects on education processes (Hughes, 2003: 1-2), as the primary objectives of the curriculum were largely unfulfilled.

Moreover, rating methods and the way in which students' results were issued were questionable as these examinations seemed to suffer from a lack of credibility in terms of marking (which was done manually, and many questions were marked subjectively), and the lack of adequacy and transparency in terms of preparing and publishing the results, all of which raised disquiet among students and parents. Consequently, it was noticeable that some students occasionally lodged appeals seeking scrutiny of their marks (Onaiba, 2006). These exams were also criticised in terms of cheating:

The examinations system clearly poses many problems, including relationships to the curriculum as well as cheating. The students want less reliance on recall, less reliance on end of year examinations and they feel that they are being undermined by the ease of cheating. The system is dominated by the reward of accurate recall (Alhmali (2007: ii).

Also, student-to-student cheating and teacher-to-student cheating could easily occur because students of the same grade were tested on the same subjects with the same exam papers and sat the exam quite close to one-another in the same venues. Also some teachers/markers may recognise their schools' exam papers and give them higher marks.

As a result of this, many educators, who studied issues related to examination and curriculum implementation in the context, called for a change in the examination system (Onaiba, 2006; Alhmali, 2007; Orafi, 2008; Ali, 2008). They recommended that the testing syllabus should be recast, suggesting it be replaced with more effective tools of assessment which consider the objectives upon which the curriculum is based.

### 1.3.2.3 The New Examination Policy

This section describes briefly the new BECE pattern, aims and purposes. It also reviews the exam in terms of its format and content. More detailed analysis of the exam questions and items in relation to its objectives and the objectives of the current curriculum of English is dealt with in Chapter Four, section 4.2.5.

#### 1.3.2.3.1 Exam Purposes and Aims (Intended Washback)

As mentioned in the previous section, the old exam pattern has received a lot of critique regarding its construction and content (Onaiba, 2006; Alhmali, 2007; Orafi and Borg, 2009). So, following the quotation extracted above in section 1.3.2.1:

... and to develop this tool a decision was issued by the General People's Committee of Education [i.e. the Ministry of Education] No. (6) for 2007 to adopt the *electronic examinations* .... [T]hese exams aim to develop ways and methods of examinations to go in line with modern scientific developments and to introduce computers in examinations to monitor grades, issue certificates and [I]t also enables students to review their results automatically (ME, 2008: 11).

It is important to mention here that “electronic examinations” are not online exams, but they are paper-pencil exams marked (and the results issued) electronically by computers. These new exams were administered as school-leaving exams to two student cohorts: grade 9 students at the Basic Education Stage and grade 3 students in secondary schools – the former being the main concern in this study (Appendices 2a and 2b contain a sample of this type of exam and its answer sheet).

Regarding test administration, the new examination is administered following a diligently typical top-down structure to ensure uniformity and security. The General Administration of Examinations and Inspection of the Ministry of Education oversees the general guidelines of the test. Its implementation is conducted by the

executive offices of the Authority of Education through Examination Boards. Each Examination Board supervises the administrative practices of each municipal authority and is responsible for technical testing-related practices.

The new exam was first introduced to Grade 9 students of basic education in May 2009. The theoretical part of this research coincided with the implementation of the new examination, and at the time of conducting the empirical research, the exam had only been in practice for two academic years. This contributed to the significance and originality of this research as it was the first study in this context to conduct such research on this type of examination.

The aims of the new exam, apart from its primary function as a disciplinary tool, as outlined in the decree of the ME No.6 – 2007 (see appendix 15), are:

- to facilitate the way in which candidates answer the exam questions;
- to cover, comprehensively and equally, all the components and contents of the curriculum [as previously there has been a mismatch between exam content and curriculum content, as reported above];
- to provide students with a gauge of their language learning achievement as far as the material of the prescribed syllabus is concerned;
- to minimize the risk of cheating [a prevalent phenomenon as reported by Alhmali, 2007: ii];
- to score the answer sheets mechanically, and hence to disseminate the results quickly, adequately and as transparently as possible.

However, there was no empirical evidence available before this research was conducted as to whether the new exam helped to achieve the desired aims (the intended washback), and whether there were any unintended consequences. Therefore, these issues were of particular interest in this study as they contribute to the significance and originality of this research.

#### **1.3.2.3.2 Exam Design and Format**

A perusal of a sample of the new examination pattern, with reference to the old one as discussed in section 1.3.2.2 (also see appendices 1 and 2a and 2b), shows that the former has been changed noticeably in terms of format and style. Unlike the previous exam pattern, candidates transfer their answers to a provided answer sheet. As marking the exam is conducted electronically by computers, it is possible for students to get their results automatically online with no need to go to their schools and ask for the result, as with old examinations (ME, 2008). In addition, unlike the previous examination, the new pattern of the BECE paper is exclusively based on selected-response items, such as multiple choice, true-false and matching items, where candidates are not required to compose their answers, they simply need to choose the answer they perceive is correct (Hogan and Murphy, 2007), and with a pencil they shadow its number or letter as indicated on the answer sheet (see appendix 2b). The answer sheet is then inserted into computer software to calculate the total score, gaining reliable, quick and transparent results. This kind of marking based on scanning of answer sheets is deemed a form of “automated scoring” (Bejar, 2011: 336), which has often been cited for their positive impact on the quality of marking. However, the literature showed widespread evidence for the perception that such auto-sored formats of tests do not yet command



public/widespread support as they are deemed to reduce standards and undermine the quality of assessment (Haggie, 2008).

According to exam constructors, in order to facilitate the way of dealing with the new pattern of the BECE, unlike the old exam pattern, a cover sheet is attached to the new exam paper which contains information such as student name, city, school, subject matter and the time allotted to answer the exam. More importantly, as the exam format is new for students, and in order to make it easily understood, the cover sheet also includes information and instructions for students on how to answer the exam and the means by which the student can transfer their answers to the answer sheet. All this information is written in Arabic to ensure students are aware of how to deal technically with the exam, reducing the risk of misunderstanding on the part of the students. This could contribute to the increased ease of exam administration; and hence boosts exam practicality. Moreover, the instruction of each type of question (true/false, multiple choice, or matching) is translated into Arabic, to make it easier for students to know what they are asked to do and to reduce the possibility of students making haphazard responses, a process which may improve the face validity of the exam (Beglar and Hunt, 1999).

Furthermore, each candidate has his/her own exam paper with their name on it. Importantly, for test fairness which is deemed an aspect of test validity (Xi, 2010a: 147), all exam papers are identical; however, given that whenever there is testing, there is cheating (Fulcher, 2011), the exam paper for each student within the same region of exam room is different from his or her neighbour in terms of

question ordering, i.e. sequencing of questions. For example, question 4 for X student may be question 18 for Y student, and so forth.

Additionally, as cheating is associated with many forms of large-scale state-wide assessments (Amrein-Beardsley *et al.*, 2010), better security measures and identification checks are carried out. The purpose of this is two-fold: to prevent exam impersonation, and to diminish the risk of cheating. As noticed by Alhmali (2007: ii) the practice of cheating was a common behaviour in the context of study among students and some teachers with the old examinations patterns. Students with the new exam pattern are less likely to cheat because to do so would take too much time in finding what question in their paper matches the other student's exam paper. Also exam invigilators are usually not aware of English matters so that they are not able to assist students; regardless, if they did so they would make noise and it would take time for them to recognize the answers to be leaked to the student(s), a process likely to catch the attention of the exam invigilation and monitoring committee in the venue. Moreover, these committee members will not condone such practices as they are meticulously and selectively appointed by the examination boards in the region to monitor the examination process. So, all efforts are being made to ensure the integrity of test scores by eliminating chances for candidates to attain marks by fraudulent means, which in turn may threaten test validity (Fulcher, 2011).

#### **1.3.2.3.3 Exam Content**

The exam content is totally based on the content of the prescribed teaching materials/textbooks. In other words, what is included in the exam stems from what students studied during the school year, making the exam one version/form of other

achievement tests. Similar to the old exam, the new exam items appear to be constructed with the objective to test students' reading comprehension abilities and grammar competence. However, with the new exam pattern, the emphasis seemed to be on questions of reading comprehension, which are cloned verbatim from texts stipulated in the students' textbook. For example, the nature of the questions seemed to be virtually based on specific information of some events and dates taken from the reading texts in the prescribed course-book. Thus, it is claimed that this would likely encourage students to memorize these chunks of language rather than understanding or developing the skills they represent (Hughes, 2003), a strategy may make the exam harmful rather than useful to the educational process in the context. The current researcher during the empirical stages attempted to verify this from teachers' perspective (also by analysing the content of the exam vis-à-vis the content of curriculum, see section 4.2.5) and explore the impact of this drawback on teachers' classroom instruction.

Consequently, given that what is untested will likely be untaught or unlearned (Hughes, 2003), it could be hypothesised that instructors would neglect the untested aspects such as speaking and listening in their classroom teaching. This claim needs to be evidenced empirically. Hence, this issue was a top priority to investigate at the empirical stage of this research, in order to explore whether this was the case as far as teachers' perceptions of and reactions to the new exam are concerned (see Chapter Four, section 4.2.5 which reviews and analyses the new examination in more detail). It could be hypothesised, thus, that EFL teaching and learning in the context may be directed from syllabus-oriented teaching to test-oriented teaching, test alignment instead of curriculum alignment. These issues are worth investigating in this research

study from teachers' and inspectors' perspectives, as the aim is primarily to explore the extent to which this examination has impacted upon teachers' instructional practices, materials and the curriculum.

As the participants of this research are EFL grade 9 teachers and EFL inspectors, it is important to acquaint the reader with an overview about these participants, including their qualifications and educational backgrounds. These issues are succinctly discussed in the following two sections.

### **1.3.3 EFL Teachers in Libyan Schools**

The current study's participants hold different qualifications in terms of their educational and pedagogical backgrounds. As this may impact upon their actions and reactions to examination change, it was of particular interest during the empirical part of this study to explore the degree of impact, *washback intensity* (Cheng, 1997), the exam had on grade 9 teachers regarding their academic qualifications, years of teaching experience and gender. Thus, this section will provide some background on the main participants of this study, grade 9 teachers of English.

Teaching in Libya is a popular profession among both women and men; however, it is more common among women (ME, 2008: 22), because female teachers feel more comfortable in teaching as a profession, which in turn, may have positive effects on their efficacy in teaching. In this regard, studies showed that there is a significant effect of gender on teacher instructional performance, where female teachers have better self-efficacy than male teachers (Isiksal and Cakiroglu, 2005 cited in Atta *et al.*, 2012: 253), especially, "Efficacy to Influence Decision Making, Efficacy

to Enlist Parental Involvement and Efficacy to Enlist Community Involvement subscales” (Atta *et al.*, 2012: 259).

“Teaching” as a female profession is described as a global phenomenon (Drudy, 2008; Sari, 2012). Nonetheless, in some cases it is attributable to religion and social grounds. In Libyan society, where boys’ schools and girls’ schools are usually run separately according to students’ gender, teaching is considered as a female profession (especially in those girls’ schools), because female teachers, their parents and husbands find teaching a more secure profession. In addition, teachers in Libyan schools deliver teaching on a session basis according to their timetables given to them by school principals, and hence, they are not usually required to be present at school beyond the timetable assigned to them. Thus, most female teachers due to their gender roles (being mother and/or spouse) find teaching as a career more feasible for them in the perspective of time constraint (Sari, 2012), which may have strengthened their belief that they can perform well if they choose teaching as a profession, which again may positively affect their performance in class. This was described by Shazadi *et al.* (2011) as an efficacy belief. The authors noted that: “if a person has a belief that he can do a given task, he performs it much better than that person who has some fears or reservations about that task” (392). The study, thus, will endeavour to look into the effect of teacher genders on their classroom instructional practices vis-à-vis the washback effect of the new BECE (the same will be applicable to the academic qualification teachers hold and the teaching experience they have).

In Libyan schools, until the late 1980s and early 1990s English language education was not a graduate profession; a Diploma of Teaching (DT), an undergraduate qualification rather than a degree, was the highest teaching qualification, therefore Libyan teachers of English in basic and secondary education schools overwhelmingly were not graduates. In recent years, however, there has been an increasing trend for language teachers to obtain a degree qualification from departmental universities. Up to the time this study took place, no postgraduate degree holders had been appointed to work in Basic or Secondary schools as teachers; instead, they work for higher institutions and universities.

The academic qualifications of teachers employed for teaching English language in Libyan basic education schools can be mainly classified into three types. The first are graduates of English departments in faculties of Arts, Bachelor of Arts holders (BA). These faculties prepare students for further studies and research but not for undertaking teaching tasks. Literature, translation and theoretical linguistics represent the core modules of the curriculum. Students in these faculties receive four years of English language study. Although the curriculum of the English departments in these faculties does not include any teaching practice or teaching methodology modules, graduates from this department represent the majority of teachers of English teaching in basic education schools because faculties of education, which are orientated more towards such purposes, have recently been adopted in Libyan universities.

The second type is graduates of English departments from faculties of education, Bachelor of Education holders (BE). These graduates receive four years of training in

TEFL. The curriculum of the English departments in these faculties is quite different from those in faculties of Arts, as it includes both theoretical and practical modules. The theoretical modules are concerned with developing student teachers' understanding of the linguistic components of English language through subjects such as grammar, phonetics, reading comprehension and writing. These modules are also concerned with introducing theories of psychology and their application into education through subjects such as general psychology, which is taught in Arabic. The practical modules, on the other hand, are also concerned with training student teachers to implement English language teaching methods in actual teaching tasks through subjects such as teaching methodology and teaching practice. Besides this, a language testing module has recently been introduced in some departments of English in some faculties of education. These student teachers usually have four weeks of teacher training during which they teach English classes in a Libyan school. Graduates under this scheme are assumed to be better prepared and trained to carry out the task of teaching English in basic and secondary schools than graduates from faculties of Arts. So, considering their academic and educational background, their instructional reactions to the introduced exam may differ from other teachers (i.e. BA teachers and DT teachers).

However, a common feature shared by the majority of graduates from the English departments of Libyan universities is their undeveloped oral and aural skills. In this respect, Orafi and Borg (2009) reported, as did Akle (2005) and Alhmali (2007), that "English language teachers in Libya typically graduate from university with undeveloped spoken communication skills in English" (251). Nevertheless, teachers in Libyan schools represent the main source for providing students with information

and language input (Saleh, 2002). Alhmali (2007) studied a sample of 1939 students from Libyan basic and secondary schools and reported that the teachers were seen by these students as “suppliers of information which has to be recorded and reproduced accurately in examinations” (69), and as authority figures whose instructions and knowledge should be beyond students’ questioning (Ahmali, 2007: 173).

The third kind of English instructors are teachers who hold a “Diploma of Teaching ” (DT) in English, previously gained after the completion of five years of secondary education: two years studying general subjects and three years studying English specialisation. This is the oldest scheme of teacher education in the Libyan education system in basic education schools. This scheme has not existed for many years, since the opening of the English departments in faculties of Arts and Education. Some of these teachers are still teaching in basic education schools, and others are working as inspectors of English. This qualification is professionally considered the lowest among the types discussed above, particularly in terms of readiness and language competence; however, those teachers have the advantage of many years of teaching experience, in some cases more than twenty years.

#### **1.3.4 EFL Inspectors in Libyan Schools**

As their role in this research study is deemed pivotal in the sense that they have direct contact with teachers (main research participants) in schools and classrooms during the school year, some brief background on the second type of research participants, i.e. inspectors of English, is worthy of attention in this section.



The process of inspection in the Libyan educational system is regulated under the purview of the ME by the Administration of Pedagogical Inspection. The responsibilities of this administration have been outlined by Shihiba (2011: 23) as:

- undertaking the tasks of pedagogical inspection and evaluating the performance of basic and secondary school teachers;
- supervising the process of the application of teachers' work load and providing proposals for redistributing them across the schools to fill the gaps and to ensure the adherence to the decided workloads;
- supervising the inspection process and inspectors' performance and preparing written reports about them through the offices of inspection of this administration;
- conducting periodical inspectorate visits to the educational institutions and the institutions of cooperative education and identifying any violations and treating any aspects of weakness, and participating in studying and analysing these aspects through coordination with the other relevant administrations.

Subject inspectors are selected on the basis of certain criteria. According to Abdulali (1986), a certain number of years of teaching experience are often required, as well as an average of 'very good' for the evaluation of the last three years of teaching the relevant subject (it is important to mention here that most current inspectors of English in the context of this study hold "teacher's diploma" who had had long years of teaching experience). Those who meet these criteria can apply for inspection competitions, which are held under the supervision of the Authority of Education. The competition is an examination-like process in which the candidates have to pass written and oral tests in their subject specialisation and in psychology (Abdulali, 1986). Then, those candidates who are successful in the competition are assigned as subject inspectors in the inspection offices of their regions, where they are responsible for monitoring and evaluating a certain number of teachers.

A standard annual teacher's assessment form outlining the criteria used by all Libyan inspectors has been issued by the ME (see appendix 16). This form should be filled by the inspectors after completing their classroom observation with a statement outlining their judgement of the teacher, ranging from 'weak, to 'excellent'. This final statement is often decisive, as other actions and decisions (rewards and punishments) taken by the ME are based on it (Abdulali, 1986).

Having mentioned the above, it is clear that the inspector of English carries out important work regarding monitoring and evaluating the educational and pedagogical process in schools and inside classrooms. Inspectors, therefore, have a vital role in this research in terms of their views of the exam and their reports on its impact on teachers' curriculum planning and instruction.

#### **1.4 Rationale and Significance of the Study**

The study is important for several reasons. First and foremost, in the present context of inquiry, in Libya, there has been a dearth of empirical research investigating issues pertinent to language testing, specifically focusing on the influence of language tests on EFL classrooms, washback. Existing information, which is documented in unpublished MA dissertations or PhD theses, is primarily based on studies whose loci were exclusively about investigating curricula implementation and uptake, teaching methods applied, difficulties encountered by learners in learning the language, and broadly investigating some issues related to evaluating language tests.

Secondly, since a new reformed public examination was introduced in 2009 in Libyan schools, little is known about this exam and its impact on EFL teaching and learning. As the introduced exam is a high-stakes test, this research endeavoured to

find out what was happening (teachers' perceptions and actions) under the influence of this exam; evaluate it; identify whether or not washback took place and, if so explore its nature; and determine the *washback intensity* (Cheng, 1997) on teachers' instructional planning and curriculum. Also, the study aimed to explore whether the revised exam induced the washback intended by its designers, and whether any unintended consequences emerged, making statements about the links between policy, test and other variables such as contexts, teacher beliefs and experience.

Thirdly, and importantly, it was hoped that the findings of the study will provide valuable information to encourage the involved parties in English education in Libya to improve the testing system, in order to better assess the goals and objectives of the language programme and thus promote second language learning. Thus, the results of this study may have important implications for English examination system reform, and the roles public examinations play in shaping EFL classroom practices in Libyan schools. In addition, this study may be a valuable source (base-line data) for future/follow-up pertinent studies either in the local or broader educational community.

Above all, this study will educate test writers and researchers about washback, providing on-site discoveries about the nature of washback in the context of the research. So, this study is significant because it adds to the literature new insights about EFL teachers' perceptions and practices as a result of public examination change. In addition, it may increase our understanding of what directions some factors/variables other than the exam *per se*, relating to teacher and context may push washback, which may address the question: "... is washback a theory?" (Fulcher and

Davidson, 2012: 4). This in turn will contribute to broaden our understanding of the inquiries raised by researchers since the beginning of 1990 seeking more clarifications of the intricate nature of test washback in the realm of language testing and teaching (e. g. Alderson & Wall, 1993; Bailey, 1996; Wall, 1996; Watanabe, 1996; Shohamy *et al.*, 1996; Bailey, 1999, Cheng, 2005; Wall, 2012; Fulcher and Davidson, 2012).

## **1.5 Organization of the Thesis**

This thesis consists of seven chapters. Following this introductory chapter, Chapter Two develops a conceptual framework for the research by reviewing literature pertaining to the subject of this study—washback literature—including definitions, directions and mechanisms of washback. The review also discusses washback impact of public examination change on aspects of classroom-related areas, and outlines some variables other than the exam *per se* that may come into play in determining the degree and direction of washback. Based on the literature reviewed and considering the aims of the study, this chapter sets the theoretical framework of the study by posing the research questions and the aspects they cover.

Chapter three describes the research methodology and design adopted, and presents and justifies the appropriate philosophical stance taken. Then the chapter discusses the methodological processes, encompassing the approaches and strategy used for this research, including the research sample, research instruments used in collecting and analysing research data, and deals with the data collecting procedures, documenting the difficulties and challenges encountered.

Chapter four reports the research data elicited by quantitative and qualitative methods—questionnaires, interviews observations and document study. The chapter presents the findings of the study thematically based on the research aims and questions.

Chapter Five is a discussion of the main findings guided with reference to the research questions, compared to and linked with the pertinent reviewed literature. Then the discussed findings are succinctly recapitulated in Chapter Six, enabling the researcher to draw appropriate conclusions. This chapter also outlines the implications and contributions of the study, noting its limitations, and therefore, presenting some recommendations and suggestions for future research.

## Chapter 2 Literature Review

---

### **2.1 Introduction**

It was stipulated in the previous chapter that the main aim of this study is to investigate the influence of introducing into Libyan schools a revised EFL public examination (BECE) on aspects of classroom teaching, particularly on teachers' instructional practices and the curriculum. It is believed that an important step towards achieving this aim is to form and develop an appropriate conceptual framework for the study. So, this chapter aims primarily to approach this. To do so, principally, I consulted and reviewed the pertinent literature to see what was (or was not) revealed in terms of investigating and determining washback intensity. For the purpose of the study, the focus is devoted to the extent to which public examinations,

especially when they undergo changes, have an influence on their stakeholders' (teachers in this case) feelings and attitudes—perceptions, teaching practices and the curriculum. The intention is to determine *washback intensity* in these particular areas.

Thus, this chapter focuses on the theoretical underpinnings that shaped and guided the present study. It reviews the debate associated with the washback impact of public examinations on teaching, particularly on second and foreign language education. The response within the language testing realm to the concerns of eminent scholars (such as Hughes, Alderson, Wall, Shohamy and Messick) about the consequences (i.e. washback) of test use on aspects of teaching and learning has been vigorous, as evidenced in the upsurge of interest in the area over the last 15 years. For instance, *Language Testing* devoted an entire issue (13, 3) to the question of washback (Alderson and Wall, 1993), in which Messick (1996) wrote the lead article. It also contained two discussions of the nature and scope of washback and four case studies. Another example, in addition to Wall (2005) and Cheng (2005), is the volume, *Washback in Language Testing: Research Contexts and Methods*, edited by Cheng, Watanabe and Curtis (2004). The second part of the book, particularly, presents a range of chapters which address a collection of empirical washback studies carried out in different education settings (Australia, China, Hong Kong, Japan, Israel, New Zealand and the United States).

In this chapter, reviewing the pertinent literature starts with an exploration of the definitions of the concept of washback, and discusses the directions and types of washback, positive or negative, followed by an illustration of the mechanism of washback. The following section will expound on how and to what degree the

washback impact of public examinations on some aspects of teaching and learning can be induced, and what intervening variables beyond the exam itself are commonly associated with the impact of public examinations on teachers' curricular planning and instruction, which in turn may affect the degree and kind of washback. In light of the literature reviewed and the research aims, this chapter will rephrase the research questions to be addressed in this study. The final section will provide a summary of the chapter.

## **2.2 Defining and Describing Washback**

This section starts with a review of the various definitions of washback, then my own view of the term will be described. Next, the section examines the similar concept terms referred to by other researchers, and expounds on how and when washback effects from examination—directions of washback—can be harmful or beneficial. It then deals with the mechanisms through which the washback effect may operate.

### **2.2.1 Definitions of Washback**

Definitions of washback are nearly as numerous as the people who wrote about it (Bailey, 1999: 3).

Generally speaking, the educational phenomenon that describes the influences of tests on teaching and learning is referred to as “washback” (Alderson and Wall, 1993; Messick, 1996; Bailey, 1996) or “backwash” (Hughes, 1989; Spolsky, 1995; Biggs, 1995). The former has been widely used in language education—applied linguistics—whilst the latter in general education. Hughes (2003: 1) succinctly states: “the effect of testing on teaching and learning is known as backwash”. Messick (1996: 241) further states: “washback ... refers to the extent to which the



introduction and use of a new test influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning”. He notes an additional important dimension: “evidence of teaching and learning effects should be interpreted as washback ... only if that evidence can be linked to the introduction and use of the test” (243).

Although washback is sometimes seen to be confined only to the unforeseen but unintended effects and not to the intended effects of tests (Spolsky, 1995), there seems to be a consensus among educators that washback is described as any effects—positive or negative, intended or unintended—that are induced on teaching and learning as a result of administering examinations (Alderson and Wall, 1993; Bachman and Palmer, 1996; Hughes, 2003; Cheng *et al.*, 2004; Cheng, 2005; Bachman and Palmer, 2010; Hung, 2012). For example, Cheng (2005: 112) states that washback indicates “an intended or unintended (accidental) direction and function of curriculum change on aspects of teaching and learning by means of a change of public examinations”. Thus, for this study, any impact associated with the introduction of the studied exam, positive or negative, intended or unintended, will be deemed washback. Also, as noted by Messick (1996), the effects are only washback if they can be linked to the introduction and use of the targeted test.

The present study endeavours to explore whether or not the exam constructors have achieved their announced aims (outlined in section 1.3.2.3), i.e. intended consequences via the introduced exam, and whether or not any other unintended consequences have emerged. In case of the exam not producing the intended washback or inducing unintended and negative consequences, a further product of

this study, as recommended by Wall (2012), is to understand and reveal the reasons behind this failure to feed back into the process of redesigning or improving the exam.

Washback is described as the effect of the use of tests on what happens in schools (at the micro level) and society as a whole (at the macro level) (Andrews, 2004). However, it is argued that the term *washback* is used to refer to the effect of tests on teaching and learning at the micro level, whereas effects at the macro level are referred to as *impact* (Wall, 1997; Hamp-Lyons, 1997; Bachman and Palmer, 2010; Brown and Abeywickrama, 2010). Washback is, thus, deemed to be one dimension or subset of test impact. Nevertheless, the term *washback* in this study may be used interchangeably with other terms such as effect or washback effect, impact or washback impact, and influence or consequence; besides this, Wall's (1997: 291) distinction between test impact and test washback is not adopted in this study.

Furthermore, washback can be defined according to the research conducted, the researcher and the context in which it is investigated. For example, Cheng (2005), in her Hong Kong washback study of HKCEE, used the term to “indicate an intended direction and function of curriculum change on aspects of teaching and learning by means of a change of public examination” (7). Similarly, as this work is a washback case study in the Libyan context, the term ‘washback’ is used to refer, specifically, to the extent to which a revised EFL public examination generates changes in teachers’ teaching behaviour—classroom teaching and testing practices—and teaching materials and curriculum use. That is, as noted by Messick (1996), the effects are only washback if they can be linked to the new BECE. Lam (1995: 84 – 85) points

out various forms of washback in terms of the facets explored. Among those, three main forms are dealt with in this study. These include:

- Attitude/perception washback. This refers to the effect of the introduced exam on teachers' attitudes and perception of the exam vis-à-vis instruction.
- Methodology washback. This implies how the new exam influences the way teachers plan and implement their classroom instruction, teaching and testing.
- Curriculum/textbook washback. This indicates the effect of the exam on the content of teaching as far as the current curriculum/textbook of English is concerned, i.e. teachers' choice and selection of teaching material resources.

These issues are explicated in section 2.5 with reference to the pertinent literature.

Some researchers have linked the kind of washback generated by administering a revised or new examination with test validity. For instance, Messick (1996) points out that washback is "only one form of testing consequence that needs to be weighed in evaluating validity, and testing consequences are only one aspect of construct validity needing to be addressed" (243). Likewise, it is argued that washback is not the only manifestation of the consequential aspects of test validity (Wall, 2012). For Messick, the key to test validity is that the test avoids two main flaws: a) construct under-representation, where, for instance, neither speaking nor listening skills are tested in a communicative competence-based test; and b) being construct irrelevant, where, for instance, teachers pay unduly marked attention to grammar points in a communicative-based test. Hamp-Lyons (1997) seconds this view, adding that: "[t]hese failures of test development can be identified by judgments made on the test itself [sic], and their consequences can be sought in classrooms, in teachers' and

learners' behaviours, and in textbook materials" (299). She elaborates that the effects of a test can be noticed in learners' use of the language being tested outside the classroom environment and in the role of this language in society at large. Interweaving the above views, the present study proposes that one form of the studied exam's validity be linked with and judged by the kind of washback the exam generates. That is, the more beneficial washback induced from the exam on teaching and curriculum, the more valid the test will be considered, and vice versa. Further, the exam's validity will be measured, as argued by Winke (2011: 628), from the angle of "the overall quality and acceptability" of the test among different stakeholders, teachers and inspectors.

### **2.2.2 Similar Concept Terms to Washback**

In addition to the concept of washback, researchers have used other terms associated with the influences of tests on the field of education. This section throws light on these terms. These include: "test impact" (Bachman and Palmer, 1996; Wall, 1997; Andrews, 2004), where tests can have far-reaching effects in educational systems and societies. The assumption that "tests or examinations can or should drive teaching, and hence learning" (Cheng and Curtis, 2004: 4), is often referred to as "measurement-driven instruction" (Popham, 1987) or "washback effect", as noted by Shohamy (1992).

Shohamy (1992) contends that "the use of external tests as a device for creating [a positive] impact on the educational process is often referred to as the washback effect or measurement driven instruction" (513). To accomplish this goal—guiding instruction via sound tests—a match between the content and format of the

curriculum or instruction and the content and format of the test is promoted, and this is referred to as “curriculum alignment” (Shepard, 1993; Shohamy *et al.*, 1996). This alignment, in which a new or revised test is commonly introduced into the educational system with the aim of enhancing instruction and tests, is referred to by Frederiksen and Collins (1989) as increasing “systemic validity”. The introduced test, thereby, becomes a part of a dynamic process in which alterations in the educational system occur according to feedback elicited from the test, referred to as “consequential validity” (Messick, 1996), returning us back to the concept mentioned above “test impact”, where “tests have various influences both within and beyond the classroom ...” (Pan, 2009: 258). These various influences are referred to by Cheng (1997) as *washback intensity*, “the degree of washback effect in an area or a number of areas of teaching and learning affected by an examination” (43).

### **2.3 Directions of washback—Negative or Positive**

Wiseman (1961:159-61) asserts that tests have *debts* as well as *credits* (cited in Wall, 2005: 34). Since then, there has been a consensus among researchers in language testing and education that washback is bi-directional (Alderson and Wall, 1993; Bailey, 1999; Cheng and Curtis, 2004; among others), contingent upon whether this washback has beneficial or deleterious effects on the educational process (Hughes, 1989). Hung (2012) states that washback from examinations can be referred to as the positive or negative effects tests have on teaching and learning. Therefore, as one of the aims of this study is to explore what kind of washback (positive or negative) the targeted exam will induce, it is important to know, by reference to the literature, when washback is positive and when negative, and what promotes or inhibits beneficial washback; these are the aims of this section.

### 2.3.1 Negative Washback

The use of public examinations may have pernicious effects on an educational system at the micro level—classroom settings. Shohamy *et al.* (1996) state that policy makers usually use tests to promote their political agendas—gate keeping—and to seize control of educational systems. For example, Choi (2008) concludes that EFL testing has had a big impact on EFL education in Korea across the three education stages: in elementary education, students are less motivated in their language learning; in secondary education, where students are compelled to practice test-taking strategies to prepare for exams; and in higher education "where obtaining high scores on EFL tests is deemed a prerequisite to successful graduation and employment" (55).

Alderson and Wall (1993) and Abu-Alhija (2007) argue that tests will have bad effects if they induce anxiety among teachers (see 2.5.1). At the micro level, as a consequence of inappropriate test-preparation practices, a test will also have negative effects on teaching and learning when students' scores increase without a concomitant increase in learning, i.e. *test score pollution*, as pointed out by Haladyna *et al.* (1991) (Andrews *et al.*, 2002; Ferman, 2004; Choi, 2008). A further deleterious effect that might be caused by examinations is that they may promote traditional ways of delivering instruction on the part of teachers. Traditional or dull teaching is described by Gorsuch (1999) as "a) teacher centered; b) teacher-to-whole-class oriented; c) focused on the learning of discrete facts; d) product oriented in that students are expected to repeat facts through recitation and written tests ..." (25). Traditional teaching, thus, prompts teachers to get students to learn via low cognitive

processes—memorization and rote learning—rather than comprehension and meaningful learning, and inhibits teaching, making teachers focus on the quantity rather than quality of learning, and on grade performance, rather than giving advice for improvement, which may demoralize the lowest achievers (Black and Wiliam, 2006).

Another possible negative effect associated with the use of public examinations in classroom settings is test-preparation practices. In this regard, Popham (1991) proposed two evaluative standards through which it could be decided whether a particular way of preparing students for a test is appropriate. The first is professional ethics, which highlights the avoidance of any test-preparation practice that is unethical and may “involve violation of general ethical canons dealing with theft, cheating, lying, and so on”, and any involvement by teachers in such behaviours is likely to lead to “potential personal repercussions (e.g. loss of credential) and professional repercussions (e.g. reduced confidence in public schools)” (13). The second is educational defensibility, which stresses that students' scores on a test and their mastery of the content domain being tested should increase concomitantly. Otherwise, scores will provide a deceptive picture of students' achievement, which is considered educationally indefensible, and likely to generate negative washback.

The other side of the coin related to test-preparation practices is *teaching to the test*, which is considered as a test-related incidence of cheating that some teachers carry out (Amrein-Beardsley *et al.*, 2010: 7). Studies have shown that most public examinations impose restrictions on curricula, teachers and students. For example, teachers tailor classroom practices to meet exam requirements and improve students'

scores. This impairs educational quality by distorting the curriculum, and trivializes some important aspects of language learning, i.e. narrowing the curriculum (Shohamy, 2001; Cheng and Curtis, 2004; Saif, 2006). These repercussions will not only have pernicious consequences on classroom settings at the micro levels, but also will reverberate negatively across the education system and society at large—the macro levels.

Furthermore, some washback studies have also noted that classroom instructional time has been usurped by tests: teachers spend a lot of time on test-oriented activities. Andrews *et al.* (2002) found that two-thirds of classroom instructional time was spent working with exam-related materials. However, this should not be perceived as a negative effect if the time allocations for test preparation were spent on more meaningful learning tasks, as in the case of Shohamy (2001), Andrews *et al.* (2002) and Ferman (2004), all of whom noted a focus on oral skills.

### **2.3.2 Positive Washback**

At the micro level—classroom settings—washback of a test can be beneficial if it promotes effective teaching, and hence prompts productive/creative learning. Appropriate use of tests can encourage learners to learn and teachers to exert efforts in classrooms, use effective techniques and methods of teaching, and pay as much attention to weak students as to strong ones (Wall, 2005). The literature indicates that tests impact teachers and students in myriad ways. Conclusions drawn from Shohamy's (1993) and Shohamy *et al.*'s (1996) studies of two language tests introduced in Israel revealed that the tests were capable of improving educational behaviour. For the Arabic (ASL) test, there was an alteration in teaching practices



evidenced in the placing of some aspects of language at the top of the teaching agenda, such as Arabic vocabulary; however, the intended washback (i.e. promoting learning the Arabic language) was quite limited, and decreased over time. But in the case of the EFL oral test, teachers devoted extra time to teaching spoken language, previously marginalized, which induced the intended washback. However, the learning tasks were identical to those tested.

Furthermore, for a test to promote beneficial washback, it should be purposive, well-known to teachers and students, as well as reflecting the course objectives upon which the test content is supposedly based (Pearson, 1988; Shohamy, 2001; Brown and Abeywickrama, 2010) (see also Bailey, 1996; Hughes, 2003; Cheng and Curtis, 2004). In this vein Messick (1996: 241-242) states: “for optimal positive washback there should be little if any difference between activities involved in learning the language and activities involved in preparing for the test”. That is, there should be some sort of overlap between the content of teaching/learning the language and the content of practicing it on one hand, and the content of the activities/tasks used in preparing for tests on the other. So, in the case of the targeted exam, it would be desirable if the constructs evaluated overlap with the content of the teaching curriculum (curriculum alignment) as the latter is based on communicative principles of language learning (Orafi and Borg, 2009).

At the macro level—educational/societal settings—decision makers use tests to achieve the goals of teaching and learning, such as introducing new textbooks and curriculum (Shohamy, 1992; Cheng, 2005). Pan (2009) stated that: “tests are encouraged to promote the idea of lifelong learning and encourage people to learn

English” (260). Moreover, administratively, high-stakes language tests, further to their purpose as a disciplinary tool, are deemed to bring about positive washback when they save time and money. The literature shows that exams with automated scoring contribute significantly to reducing the cost of assessment (Toch, 2006; Haggie, 2008). However, since most of these tests focus heavily on discrete-point testing (i.e. multiple-choice or true-false items) and eliminating integrative testing, they seem to induce negative rather than positive effects on the educational process, as they strip teachers of the spur to teach higher-level skills, and are based on questions of rote skill, encouraging rote learning (Toch, 2006). This, in turn, will undermine the quality of the test administered (Haggie, 2008).

Given this study’s assumption that washback effect is inevitable—a common view in language education (Wall, 2005)—the speculation is that the washback effect of the exam under study (the new BECE) would likely be positive or negative, or both.

## **2.4 Mechanisms of Washback – Hypotheses and Models**

This section reviews three models through which washback from exams may operate. The three models discussed are: Alderson and Wall washback hypothesis, Hughes’ model and Bailey’s model. These models have been the foundation stones for the evolvement of other models (Burrows, 2004; Cheng, 2005; Saif, 2006; Green, 2007; Tsigari, 2009b). Since “there is little theory within language testing regarding washback mechanisms” (Xie and Andrews, 2013: 55), the researcher, by the end of this section, will attempt to triangulate the reviewed models in order to set a useful conceptual framework for the practical part of this study.

### 2.4.1 Alderson and Wall Washback Hypothesis

Alderson and Wall (1993) proposed the Washback Hypothesis in an attempt to clarify the idea of washback and to serve as a basis for further research. The authors highlight that as an important step towards investigating washback, a researcher needs to consider a set of assumptions, which they call the *Washback Hypothesis*. As a result of reviewing studies conducted in different contexts, and of their own work on O-level examination in Sri Lanka, they posit 15 hypotheses. These are:

- 1-A test will influence teaching.
- 2-A test will influence learning.
- 3-A test will influence **what** teachers teach.
- 4-A test will influence **how** teachers teach.
- 5-A test will influence **what** learners learn.
- 6-A test will influence **how** learners learn.
- 7-A test will influence the **rate** and **sequence** of teaching.
- 8-A test will influence the **rate** and **sequence** of learning.
- 9-A test will influence the **degree** and **depth** of teaching.
- 10-A test will influence the **degree** and **depth** of learning.
- 11-A test will influence attitudes to the content, method, etc. of teaching and learning.
- 12-Tests that have important consequences will have washback.
- 13-Tests that do not have important consequences will have no washback.
- 14-Tests will have washback on **all** learners and teachers.
- 15-Tests will have washback effects for **some** learners and **some** teachers, but **not** for others (Alderson and Wall, 1993: 120-121, bold in original).

Since this research concerns the influence of a revised test on teachers' classroom practices and curriculum, the study will focus on and try to confirm seven versions of the above mentioned hypotheses with slight amendments according to the aims of the study. Because they are considered unsuitable for the purpose of the research, the other eight hypotheses particularly those related to learning and learners are beyond the scope of this study. Thus the hypotheses addressed in this study include: a test

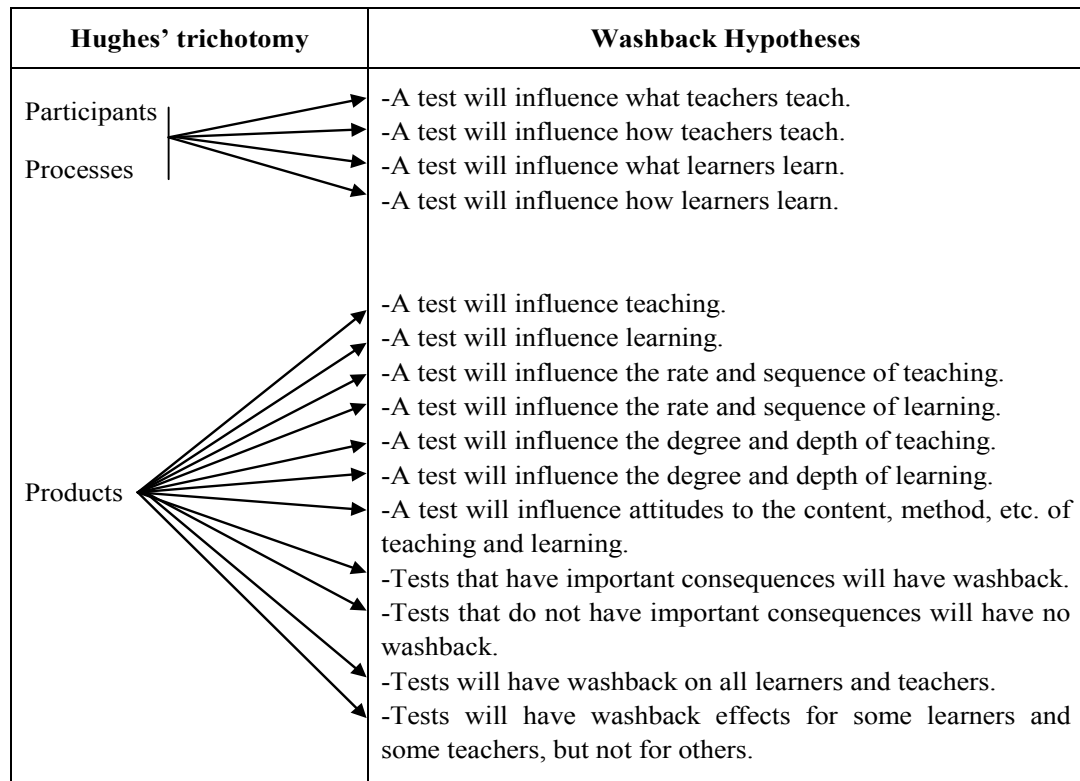
will influence teaching; a test will influence what teachers teach; a test will influence how teachers teach; tests that have important consequences will have washback; tests that do not have important consequences will have no washback; tests will have washback effects on ... teachers; and tests will have washback effects for ... some teachers, but not for others. However, the study will endeavour to test the other eight hypotheses from the standpoint of the tested hypotheses, as the study is mainly confined to the impact of the reformed exam from teachers' perspectives, not from learners'. Thus, this study will attempt to confirm the existence of this phenomenon, corroborating Alderson and Wall's (1993) answer to their rhetoric question 'Does washback exist?'

#### **2.4.2 Hughes' Model**

As cited in Bailey (1999), Hughes (1993) argued that Alderson and Wall's 15 hypotheses were quite broad, and it would be better to be precise and explicit about what constitutes exam washback. He suggests: "In order to clarify our thinking on backwash, it is helpful, I believe, to distinguish between participants, process and product in teaching and learning, recognizing that all three may be affected by the nature of a test" (cited in Bailey, 1999: 9). According to Hughes, *participants* are classroom instructors and learners, the educational system administrators, and curriculum designers, "all of whose perceptions and attitudes towards their work may be affected by a test" (cited in Bailey, 1999: 9). The constituent *process* refers to any action or interaction taken by the participants that may (or may not) contribute to learning, such as improving learning strategies, changing a teaching methodology, or introducing new materials. Finally, *product* covers what is achieved. Table ... shows

which of these three constituents relate and interrelate to Alderson and Wall's washback hypotheses.

Figure 2-1 Correspondence between Hughes' trichotomy model and Alderson and Wall's washback hypotheses

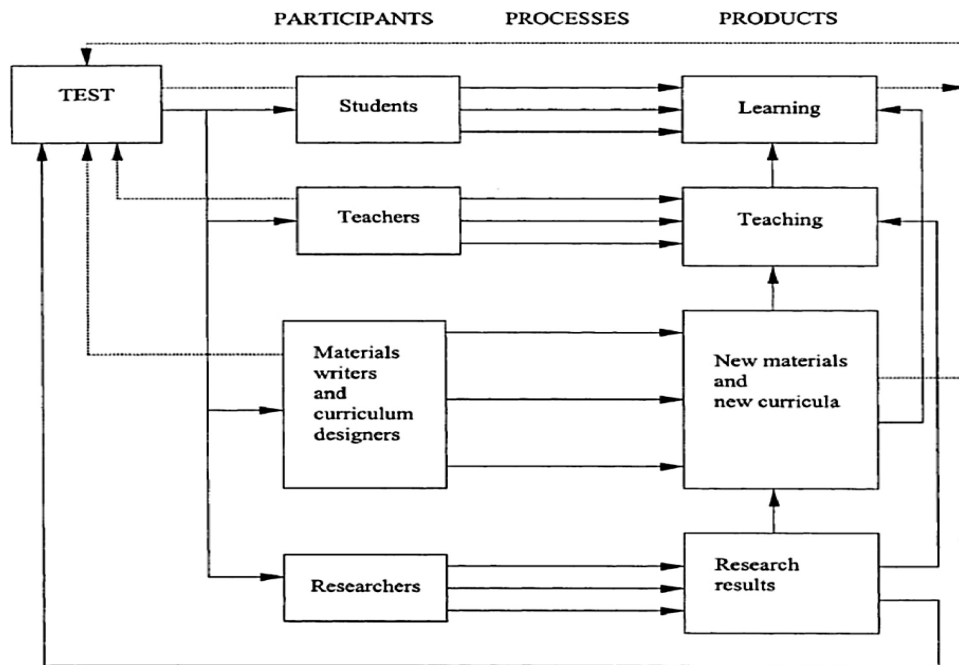


### 2.4.3 Bailey's Model

Synthesizing Alderson and Wall's Washback Hypothesis with Hughes' (1993) distinction between participants, processes and products, Bailey (1996) propounded and developed a model of washback to delineate its complicated mechanisms. (see figure 2.1 below). Bailey contends that a test not only may affect products via the involved stakeholders—participants—and their actions and interactions—processes—but that participants and processes can also in turn provide feedback, all of which can have an impact on the introduced test, i.e. reciprocal impact, as the dotted lines in Figure 2.1

represent. The model emphasizes the importance of the interaction between the different components.

Figure 2-2 A basic model of washback



Both arrows and arrows with dotted lines mean ‘influences’ (Source: Bailey, 1996: 264).

#### 2.4.4 Conceptual Framework of the Study

Reviewing Alderson and Wall's washback hypotheses and Hughes' trichotomous model (participants, processes and products), this study claims that both views serve the same purpose and tackle the same issue from different perspectives (This is highlighted in Bailey's model discussed above). The former claim deals with the theoretical orientations, i.e. conceivable results and types of influences of test impact. The latter portrays practical steps and mechanisms through which this impact is operationalized, and each constituent hinges upon the preceding one: there will likely be no processes without engagement on the part of the participants, and by extension any products or outcomes will be consequences of those processes undertaken by

participants. Additionally, Bailey (1996), as discussed above, created a basic model to elaborate on the mechanisms via which washback evolved.

As far as this study is concerned, Hughes' trichotomy—participants, processes and products—will be adopted, because it is assumed that these notions are directly interrelated and should coexist with each other in order to successfully investigate washback effects from the exam. By 'participants', I refer to teachers in relation to their attitudes and feelings towards the introduction of the revised examination in Libyan schools. As a result of their perceptions, teachers would (or would not) change their processes, such as testing/teaching practices or methods, choice of classroom activities, and teaching content, which in turn may impact upon the curriculum. The outcome effect of this behaviour will constitute the products—whether the intended consequences were achieved or unintended consequences emerged.

Having discussed the above, and with reference to other washback models, one might claim that washback as observed in language education operates in a circuitous and interactive way. In this study the researcher will triangulate the three models in order to set a framework for this study. However, additionally, this study is not limited to the theoretical implications of the above mentioned hypotheses and models, but, will explore and understand what occurs in classrooms as a result of public examination change in the context of the study. However, the models discussed above have informed the formulation of the theoretical framework of the current study.

## **2.5 Washback Impact of Public Examination Change on Aspects of Classroom-related Areas**

Washback studies have demonstrated that exam impact occurs to different degrees in relation to different aspects of teaching and learning. The relationship between examinations and the educational process can be described, as in Barnes *et al.* (2000: 632), with the metaphor “the ripple effect”. The authors reported two studies that documented the connections between the mandated assessment—a high-stakes test—and instruction, suggested the effect of tests on teaching throughout the schools to which the exam was administered. This section, thus, reviews previous washback studies regarding the effects of tests on some aspects of the classroom: teachers' feelings and attitudes towards exams (perception washback), instructional practices (methodology washback), and teaching materials and syllabus (curriculum washback).

### **2.5.1 Washback Effects on Teachers' Feelings and Attitudes Towards Exams (Perception Washback)**

Odo (2012) states that:

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

As a consequence, however, it is contended that an inevitable initial step causing the washback effect is that the nature of a test might first influence teachers' feelings, attitudes and perceptions—evaluative reactions. These feelings and attitudes, being likely as a response to learners' expectations, might in turn affect how teachers carry



out their work (Hughes, 1993 cited in Bailey, 1999). Moreover, the literature has indicated that examinations, especially those newly-introduced or revised, influence stakeholders, especially teachers' attitudes and feelings towards their instructional behaviour in the sense that they "increase teachers' stress and lower their morale" (Abu-Alhija, 2007: 57), or, in contrast, motivate teachers to work harder and adopt innovative methods and techniques "more in line with communicative and, to some extent, humanistic teaching" (Prodromou, 1995:15).

In her HKCEE study, Cheng (2005) noted that teachers were anxious about how their students, especially the less outspoken ones, would pass the revised exam. One of the teacher interviewees in Cheng's study declared she would feel guilty if she did not familiarize her students with the new exam formats and content. This is echoed by Tsagari (2009b), who investigated exam influence on participants' perceptions and material design in Greece. Interviews with teachers revealed that they feel anxious and stressed by trying hard to cover all the materials on the syllabus. In her preliminary study in the same context using diaries to gather her research data, Tsagari (2009a) also reported that: "Evidence of more intensive washback was recorded in the diaries as the date of the exam drew closer. This reached a peak in the weeks prior to its administration and was accompanied by intense physical reactions such as upset stomach, headache, and sickness"(7).

Shohamy (1993) and Shohamy *et al.* (1996) found that teachers still had negative feelings towards the Arabic test (ASL), complaining that the test was of no importance. Teachers, in contrast, approved of the EFL oral test: they viewed the test as important since it encourages students to improve their English oral skills.

However, this exam is said to “create an atmosphere of high anxiety and fear of test results among teachers and students” (Shohamy *et al.*, 1996: 309-10) (see also Wall and Alderson, 1993; Alderson and Hamp-Lyons, 1996; Ferman, 2004).

Shohamy (2007) stresses that there is a need to examine the ramifications of tests on stakeholders in relation to tests’ uses, misuses, fairness, biases, and discrimination. In their recent washback study, Cheng *et al.* (2011) investigated stakeholders’ (students and parents) perceptions of a recently introduced high-stakes nation-wide exam in Hong Kong. The study investigated students’ perceptions of the impact of the introduced exam in relation to their language learning, and parents’ perceptions of their role in the process. Through students’ questionnaires, the authors found straightforward connections between students’ perceptions of exam-related learning activities and their perceptions of their levels in English (similar results were drawn from Cheng, 1998). Moreover, based on questionnaires disseminated to students’ parents, the study demonstrated that parents’ perceptions of the exam and their opportunities to know about it significantly and positively predicted their role in supporting their children with an utmost aim to boost their children’s scores on the exam. The study concludes that parents’ perceptions about the introduced exam are directly and significantly related to their children’s perceptions about the exam (Cheng *et al.*, 2011: 221), which in turn would likely directly or indirectly affect the teacher in class.

Thus, teachers are affected by the expectations of other stakeholders. Particularly, they are pressured, for instance, by school administrators, students and their parents (*systemic stressors* in Gregory and Burg, 2006) to direct their methodologies

(Wall, 2000). Consequently, this may present teachers with a dilemma leading to what Spratt (2005: 24) calls “a tension between pedagogical and ethical decisions”: either to practice what they would like to teach stemming from their own philosophy of what real learning is, or to be enslaved by teaching to the test to enable their students to pass exams, especially when those exams are of low quality—poorly constructed—in terms of the constructs they measure. Thus the dictates of high-stakes tests may 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 (Madaus, 1988; Smith, 1991a; Gipps, 2011).

Furthermore, the literature shows that the pressure associated with high-stakes tests leads teachers/educators to cheat, sometimes in subtle ways. Amrein-Beardsley *et al.* (2010), for instance, categorized three degrees of educators’ cheating in terms of severity in response to high-stakes tests: the most serious degree, is unforgivable and inexcusable behaviour as willful and premeditated, which some teachers do with malice and forethought; the second degree may be caused by a teacher’s casual lack of concern; and the third degree might be caused by indifference or carelessness, but not premeditated or intended (Amrein-Beardsley *et al.*, 2010). The authors also noted similar behaviour among students when sitting their exams. They concluded by noting that although educators’ cheating is frowned upon by the vast majority of people, “the conditions under which educators now work, and the high-stakes testing policies by which they are now controlled, have pressured more educators into doing so [i.e. cheating]” (27).

As one aim to avoid such negative behaviour from the part of the teacher as well as the learner, some countries' education systems have witnessed the adoption of new technologies in their assessment programmes, such as automating scoring of exam papers and mechanically disseminating the results of exams to improve the marking quality and the way tests are controlled and monitored (Newton and Meadows, 2011) (for more details, see *Assessment in Education: Principles, Policy & Practices*, the special issue on marking quality within test and examination systems (18, 3), and also *Language Testing* where an entire issue (27, 3) was devoted to the development of automated scoring and feedback systems, the editorial written by Xi, 2010b). As mentioned in the previous chapter (section 1.3.2.3), the examination system in the context of this current study has undergone changes in the format and content of school exams. Among the aims, according to exam constructors, are eliminating the risk of cheating in the final exams, and disseminating the final results quickly, adequately and as transparently as possible. So, in this study it is worth exploring whether or not these aims (and the others as well) have been achieved via introducing the new examination.

Despite their deleterious effects, however, pressure on teachers and their anxiety associated with the introduction of high-stakes tests can also have positive effects on the educational process. It may make teachers more accountable. Conclusions drawn from the studies of Shohamy *et al.* (1996), Andrews *et al.* (2002) and Ferman (2004) demonstrated that teachers focused on the oral skills of the language, as these were tested in the introduced exams, encouraging teachers to spend more time on more meaningful learning tasks. In this regard, Gregory and Burg (2006) underscore that

whilst tests induce negative consequences, they may have some positive effects on instruction:

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

In a similar vein, Wall's (2005) observations and interviews with teachers revealed that they had mixed but mainly positive reactions towards the exam. Further, Amengual-Pizarro (2009) reported that the majority of teachers in her study seemed to have positive attitudes towards the introduced test. She concluded that, according to teachers, the test "was thought to be useful and necessary" as well as "reliable"(592) (see also Watanabe, 2000: 44).

To conclude, high-stakes tests do exert washback effects on teachers' feelings and attitudes, ranging from positive to negative, but what is not clear is how and to what extent these effects generate effective teaching. So, research on test anxiety and its effects on teachers during their teaching are worth investigating in relation to washback. However, this study endeavours to cast light on teachers' feelings and perceptions towards the new examination, and the extent to which these perceptions have an influence on their instruction and curriculum use.

### **2.5.2 Washback Effects on Instructional Practices (Methodology Washback)**

This study views washback effect from the standpoint of teachers' classroom practices and curriculum use. Since teachers are the key factor in the implementation of reform (Brindley, 2008) and "the 'front-line' conduits for the washback process related to instruction" (Bailey, 1999: 17), this section expounds on how teachers'

instructional practices may change through public examination change. Teaching practices, in this study, are referred to as the teaching methods, techniques and activities teachers use or adhere to in their instruction; along with, their classroom assessment practices—on-going tests—these all constitute “methodology washback”.

The literature has shown heterogeneous findings on the extent to which language tests induce changes in teachers’ classroom instruction. The findings in this area indicate that washback intensity registered various levels, from heavy washback to zero washback. Using teachers’ questionnaires, Stecher *et al.* (2004), investigating the influence of WASL tests on methods teachers used in teaching writing, found that teachers changed their methods to reflect test requirements. Likewise, in her washback study, Amengual-Pizarro (2009) examined the influence of the ET test included in the SUEE on various aspects of teaching. Her survey study revealed that virtually all teachers reported that their methods of teaching were considerably influenced by the test. For instance, she found that practicing students' oral production was dramatically reduced, as it was not assessed in the ET test. She concludes:

Contrary to previous studies that found no straightforward connection between the test and teachers’ methodology [...], the results of this study also appear to indicate that the ET affects the methodology teachers employ in actual class teaching adapting it to the purpose of the test" (594).

Shohamy’s (1993) study on three language tests (the EFL oral test, ASL test and L1 reading test) in Israel revealed that the methodology teachers adhered to became more “test-like” as the exam approached. Shohamy *et al.* (1996) revisited the washback effects of the same examinations to probe washback over time. Regarding the EFL oral test, students’ questionnaires and interviews with teachers and

inspectors revealed a similar washback effect (i.e. to a large extent positive), but in the case of the Arabic test and L1 reading test, the washback on teachers' teaching methods was severely limited, or even nonexistent. This is ascribed to the purpose for which test scores are used, the stakes (high or low) and the status of the language being tested (Shohamy, 2001) (for similar results see Yarbrough, 1999, Barksdale-Ladd and Thomas, 2000 and Jones and Johnston, 2002, all cited in Jones *et al.*, 2003:37-38).

However, one might be cautious about the above reported findings and conclusions as they largely rely upon quantitative (i.e. indirect) methods of data collection—questionnaires—which are deemed to provide insufficient evidence for elucidating a vivid picture of washback effects from examinations because “we need to look closely at classroom events [ by using direct research methods for data collection] in particular, in order to see whether what teachers and learners say they do is reflected in their behaviour” (Alderson and Wall, 1993: 127), as “teachers may hold beliefs that are not compatible with the practices” (Muñoz and Álvarez, 2010: 46) (see also Cheng, 2001; Watanabe, 2004; Cheng, 2005; Wall, 2005).

Nonetheless, some studies using direct research methods, such as classroom observation, found changes in teachers' teaching methodologies by means of public examinations. An example is Watanabe's (1996) study in Japan. Based on classroom observation of two teachers, he found that they were influenced by the test in terms of how they teach, but the degree of this influence contrasted from one teacher to another, dependent upon other factors (see section 2.6). Benefiting from the use of video records of oral tests, grading of oral tests and discourse analysis in

investigating the influence of the addition of an oral component to the use of English (UE) exam on students' oral performance in Hong Kong, Andrews *et al.* (2002) found similar results, as did Burrows (2004). They reported that: "the oral examination has had some sort of effect on students' spoken output" (220), an influence increased over time which was attributed to teachers' familiarities with the exam and its requirements, thereby (in addition to students' expectations) teachers changed their teaching methods and techniques toward the exam.

On the other hand, several studies conclude that tests have little to no effect on teachers' teaching methodologies. Wall and Alderson's (1993) Sri Lankan study, which examined the effects of introducing a new English examination that was intended to encourage a more communicative approach to teaching, revealed that the examination "has had virtually no impact on the way that teachers teach English" (127). This finding is similar to that of Qi (2004) and Cheng (2005). For instance, Cheng (2005), undertook a longitudinal washback study to investigate the influence of a newly revised public examination, HKCEE, on aspects of teaching and learning. Although the exam was designed to engineer positive washback by exerting more task-based approaches to teaching and learning, the data elicited from teacher and student questionnaires, interviews and classroom observations indicated that washback effects of the examination on teaching were limited. She comments: "the way the teachers carried out their teaching remained more or less the same, whether the testing syllabus was the old one or the new one" (Cheng, 2005: 246). She thereby suggests that teachers' teaching methods in classrooms may remain unchanged; however, their teaching activities may change as a result of exam change; for instance, reading aloud was replaced by role plays and discussions, but both were



instructed through drilling. Conversely, Watanabe (2000), who investigated washback effects of a pre-college English section examination on instruction in Japan through classroom observations and teacher interviews, reported that teachers in his study “claimed that they deliberately avoided referring to test taking techniques, since they believed that actual English skills would lead to students’ passing the exam” (45).

In conclusion, given that classroom observation is an effective method for detecting exam influence on teaching and learning, washback effect on teaching methods seems not to be inevitable, despite, in many cases, the studied exams are designed to require a modified methodology. The reported findings seem to be in line with one version of Wall and Alderson’s Washback Hypothesis: “Tests will have washback effects for some learners and some teachers, but not for others (121). However, they are rather more equivocal with regard to the other version: “A test will influence how teachers teach”(120). Thus, as far as thoroughly empirical investigations of classroom teaching methodologies as a result of exam change are concerned, a gap seems to emerge in the literature of washback effects, which is an issue worth scrutinizing in the present study.

Additionally, it appears that teachers can play an essential role in fostering different forms of washback, and thus, as noted by Spratt (2005), they play a significant role in determining the types and degrees of washback impact, promoting (or inhibiting) positive washback. Keeping this in mind, teachers (and inspectors—expert teachers) were targeted as the main population of the study.

### **2.5.3 Washback Effects on Teaching Materials and Curriculum (Curriculum Washback)**

In this study, teaching materials and curriculum refer to the content of teaching and textbooks used to deliver instruction in the context of study. Wall (2012) points out that washback impact on curriculum and teaching materials can materialize when teachers and students “pay more attention to certain parts of the teaching syllabus at the expense of other parts because they believe these will be emphasised on the test” (79). Wall's statement was based on findings from previous washback studies, which have explicitly shown that teachers, for instance, design their teaching materials and content around tests, called *curriculum alignment* (Madaus, 1988) (Smith, 1991b; Wall and Alderson, 1993; Stecher, 2002; Cheng, 2005; Abu-Alhija, 2007, Choi, 2008). Wall and Alderson (1993), for example, observe: “the new exam has had a demonstrable effect on the content of language lessons” (126-27). Nevertheless, Shohamy *et al.*'s (1996) study of the impact of the ASL test and the EFL test shows a slightly different picture. They report that the Arabic test induced little effect on teaching content whilst the EFL oral test did. Shohamy *et al.*'s results may indicate that the stakes of the test is influential in determining the amount of washback; that is the higher the stakes, the greater the effect, and vice versa (see section 2.6.2 for elaboration).

In their New Zealand IELTS study, Read and Hayes (2003) noted that in 90 percent of cases, exam preparation materials were exploited (see Alderson and Hamp-Lyons, 1996 for similar results). Cheng's (2005) findings are once again similar. She reported that the new revised examination has had a tremendous impact on teaching

materials, in that new textbooks were published and distributed to schools before the exam was administered. Cheng (1997: 37) ascribes this to “the highly adaptable and commercial nature of Hong Kong society”, and also notes that: “textbook publishers in Hong Kong not only provide teaching materials but also detailed teaching and learning activities with suggested methods” (Cheng, 2005: 130). However, the quality of such new materials has not been investigated by Cheng to determine the extent to which these materials generate positive or negative washback.

Moreover, teachers may be influenced by examinations in the way that they create their own materials and resort to past or mock exams. For example, Andrews (1995) and Lam (1995), who independently investigated content and textbook washback of UEE in Hong Kong, found washback effect on teaching materials, that “positive washback is evidenced by teachers creating more authentic materials from the mass media, [and] producing meaningful learning activities” (Lam, 1995: 95). Similarly, Watanabe (2000), like Tsagari (2009a), noted that teachers “tried to innovate during exam preparation classes ... using a variety of self-made material” (44). However, Watanabe’s (1996) study, again, showed incompatible results. Utilising classroom observation to investigate the influence of the university entrance examination on EFL classrooms in Japan, with special reference to the relationship between the exam and the use of the grammar-translation method of teaching, the author noted that teachers were not necessarily teaching skills such as writing and listening, although these were tested.

Additionally, some washback studies have mentioned the time factor in relation to the use of exam materials. As the exam approaches, there is a heavier use of mock

exams and more exam-related materials, with more time spent on these materials. For example, Andrews (1995) speaks of the role played by published materials, reporting that teachers spent an “estimated two-thirds” of classroom [instructional] time on exam-related published materials, which, as the author advocates, “represent a limiting of focus for teachers and students rather than a broadening of horizon” (80).

Washback impact on curriculum and teaching materials was also documented in studies investigating curriculum innovation and implementation uptake in EFL/ESL classrooms. The studies highlighted that high-stakes public examinations do affect the content of teaching and curriculum, and the way teachers deliver instruction. For instance, Gorsuch (1999 and 2000) investigated teachers’ practices in relation to EFL curriculum reform in Japanese high schools. She noted that although the reform urged the use of the all four language skills equally, the examination administered to the same schools emphasized testing the knowledge of vocabulary, grammar and language usage. Comparable results were also reported by Agrawal (2004), Orafi (2008) and Orafi and Borg (2009).

Agrawal (2004), for instance, who investigated the implementation of ESL curriculum innovation in some Indian secondary schools, reported that while the teaching syllabus focused on developing oral skills, instructors tended to marginalize these skills on their teaching agenda because they were not included in the exams written by the authority of education. Similarly, Orafi (2008) and Orafi and Borg (2009), using classroom observations and interviews, examined how five secondary EFL teachers implemented the English language curriculum, which was

communicatively oriented, in Libyan schools, the same context of this study. The authors noted that teachers failed to implement their instruction as recommended by the new curriculum. They identified three main factors that had an impact on the way teachers interpreted and implemented the curriculum: teachers' beliefs about teaching and learning; teachers' teaching experience; and contextual factors, the most influential of which was the examination system at the time as there was a gap between the orientation of the curriculum and that of the examination system (see chapter one, sections 1.3.1.2 and 1.3.2.2 for elaboration).

In summary, contrary to washback effect on teaching practices discussed in the previous section, it is evident that, in many cases, high-stakes tests have a considerable effect on teaching materials and teaching content. These results thereby are consistent with one of Alderson and Wall's 15 versions of the Washback Hypothesis: "a test will influence what teachers teach"(120). Thus, one of the aims of this study was to test this claim, to explore whether the introduction of the revised public examination in the context of the study will bring about washback on the curriculum, to assess the extent to which the findings of this study would be consistent (or inconsistent) with the findings of previous washback studies reported above.

To conclude, it seems that the findings of washback research on exam influence on diverse aspects of language teaching and learning are contradictory. That is, *washback intensity* of a test varies in degree from one area of language teaching to another and from one teacher to another. One might wonder whether it is the exam

*per se* that is the cause of these contradictions, or other mediating factors. The following section tackles some of these issues.

## **2.6 Factors Influencing Washback**

Although it has not been widely empirically investigated, there seems to be a general consensus among researchers that factors other than the exam itself might intervene in determining the direction and the degree of washback (Alderson and Wall, 1993; Shohamy *et al.*, 1996; Spratt, 2005; Cheng, 2008; Choi, 2008; Taylor and Weir, 2009; Dunkley, 2010; Wall, 2012; Fulcher and Davidson, 2012). For example, Wall (2012) pointed out “the difficulty of separating out the influence of tests from the effects of other variables at work in the educational contexts” (83-84). This section briefly discusses some of these influencing factors in relation to the context of the study.

### **2.6.1 Teacher-related Factors**

Teachers’ characteristics are repeatedly mentioned in the literature as key factors in determining when and how washback occurs. These characteristics include: teachers’ beliefs and perceptions, experience, academic qualifications and educational backgrounds and training. Teacher- gender-related influences have been documented in the area of language education (i.e. the broad literature), particularly regarding classroom instructional change as a result of curriculum change (Brophy, 1985; Bandura, 1997; Hopf and Hatzichristou, 1999; Shazadi *et al.*, 2011; Atta *et al.*, 2012). Most of the studies indicated that teacher gender is one of the influential variables that impacts upon the type of instruction teachers implement. Nevertheless, according to previous washback studies, teacher gender has very rarely been a

research focus. The gap in research regarding the relationship that may exist between teacher gender and the washback impact of public examination change is of interest in this study, in addition to the other mentioned variables.

In his review of various empirical studies of washback from external exams conducted in the field of language education, Spratt (2005: 5) points out “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”. However, this role may be contingent upon some variables relating to teachers’ experience, expectations, beliefs and academic qualifications, etc. In a similar vein, Cheng (2008: 352) states that: “teacher factors, including personal beliefs, past education, and academic background, seemed to be more important in determining the teaching methodology a teacher employs”. Likewise, Watanabe’s (1996) classroom observations in her washback study suggest that: “teacher factors, such as educational background, personal beliefs and teaching experience may outweigh the possible effect of the entrance examinations [a high-stakes public examination]” (318). The following paragraphs, thus, will briefly review pertinent texts in the literature on the effects of teachers’ characteristics on the kind of responses and reactions teachers generate to exams. The characteristics to be addressed are: teachers’ beliefs and attitudes towards exams; teaching experience; teachers’ training on and awareness of an examination reform; and teachers’ academic qualification.

#### **2.6.1.1 Teachers Beliefs and Attitudes**

Teachers’ beliefs and attitudes towards the exam (as discussed in 2.5.1) are deemed important in mediating the impact of testing washback. These perceptions and beliefs

can play a major role in promoting or impeding washback, and affect whether and how teachers plan and implement their classroom teaching and the content of lessons. As expressed by Chapman and Snyder (2000: 462): “it is not the examination itself that influences teachers’ behavior, but teachers’ beliefs about changes”. Language teachers’ beliefs, thus, could be a trigger for their reactions to examination change, and hence, filter the way they perceive their teaching and influence their practice (Mark, 2011).

The relationship between teachers’ beliefs and practices is not only documented in washback literature, it is also documented in literature pertinent to classroom situations, and generally understood to be a complex one. In his interpretive review of a set of studies into the correspondence between language teachers’ stated beliefs and practices, Basturkmen (2012) reports that: “context factors and teaching experience appeared to mediate this relationship, and that correspondences between stated beliefs and practices were reported mainly in situations involving experienced teachers and planned aspects of teaching” (282). The author, in light of his findings, suggested some areas for further research (for details, see Basturkmen, 2012: 291).

#### **2.6.1.2 Teaching Experience**

Much research that investigates the relationships between public examinations and EFL instruction has reported that teachers’ teaching experience is one of the key factors, which helps explain the reasons on why washback may influence some teachers, but not others (Fish, 1988; Lam, 1995; Shohamy, *et al.*, 1996; Watanabe, 1996; Cheng, 2005). Speaking of teaching experience, we characteristically refer to novice and veteran teachers. For the current study, the term “novice teachers” refers



to newly qualified teachers, who have completed their education programme, either graduating from university or a teacher education institution, specializing in English language, and have commenced teaching English in one of state schools. Although there is no full consensus in the literature on delineating the time novice teachers have been teaching (Farrell, 2012), here it can be as little as one year to as many as six years. Experienced or veteran teachers, on the other hand, in this study are classified as: those who have spent 7 – 15 years teaching are considered “experienced teachers”; and those who have had more than 15 years of experience are referred to as “highly experienced teachers”.

Fish (1988, cited in Pan, 2009: 260), who investigated teachers’ responses to mandated standardized testing, reported that novice teachers felt a greater degree of pressure and anxiety for accountability than did veteran teachers. A significant finding from examining experienced and inexperienced teachers by Shohamy *et al.* (1996) showed that experienced teachers were more perceptive and thoughtful to public examinations and thus were more likely to adhere to the test requirement and use it as their key source of guidance for instructional practices. Similarly, Lam (1995) reported some differences between veteran and novice teachers in terms of positive and negative washback:

The more experienced UE teachers are likely less [negatively] affected by syllabus innovation because they are more set in their ways and more confident of themselves as a result of more years of experience and the fact that they are more realistic in assessing what is functional in their working situations... (95).

Cheng (2005) points out that teachers may not change their way of teaching as a result of a testing system change because they lack two important characteristics: the ability and skill to change. As these two characteristics are likely associated with the

length of teaching experience, in this case it is suggested that veteran teachers would be more affected than novice teachers when an examination system undergoes reforms. Indeed, teachers with more years of teaching would have the ability to alter and modify their teaching methods and techniques in response to the demands of the exam introduced, an issue of worth scrutiny in this study.

### **2.6.1.3 Teacher Training on and Awareness of an Examination Reform**

The literature indicates that not only is the teaching experience teachers have gained pivotal to enabling them to innovate, the educational and methodological training they receive is also important in making teachers more creative in their instructional practices. For example, findings drawn from Cheng's HKCEE study revealed that washback on teachers' teaching methods was very limited, owing to the "the constraints imposed upon teaching and teachers" (Cheng, 1997:38). However, in Cheng's case, it appears that teachers themselves have not changed for the new curriculum. This claim gains support from Falvey (1996): "the majority of English teachers in Hong Kong are unprepared either for recent changes to the curriculum or for pedagogical changes" (cited in Cheng, 2005:16).

Similarly, Wall and Alderson (1993), as discussed in 2.5.2, found that there was no evidence of washback on methodology, because, in their view, teachers lack training in teaching towards specific exams and how to use exam-specific textbooks. They concluded that: "the exam can have no impact on methodology unless the teachers understand correctly what it is the exam is testing" (217). This description is strikingly similar to that offered by Wall (2005) and Chapman and Snyder (2000). Wall, for instance, points out that: "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” (283). Based on these findings, it can be contended that the kind of washback effect induced by the studied exam (the new BECE) would be partially attributable to teachers’ understanding and awareness of the exam demands.

#### **2.6.1.4 Academic Qualification**

Another important teacher-related factor that takes some account of how and why washback occurs or does not occur is teachers’ academic qualifications and educational backgrounds. Researchers in language education suggest that would-be teachers should have to major properly in the subject they intend to teach and receive improved pre-service education followed by developed in-service instruction programmes, to increase their conceptual and practical grasp of their subject areas. For example, Watanabe (1996) concludes that teachers’ educational backgrounds contribute to the kind of instruction they adhere to as a result of introducing exams. Thus, it is to suggest that teachers who majored in theoretical linguistics at postgraduate level might teach differently from those who obtained a BA degree from general university. The latter may deliver instruction differently from those who hold BE from teacher-colleges, or those who obtained a DT in education.

So, Richards (1990) states:

In second language teaching, teacher education programs typically include a knowledgebase drawn from linguistics and language learning theory, and a practical component based on language teaching methodology and opportunity for practice teaching. In principle, knowledge and information from such disciplines as linguistics and second language acquisitions provide the theoretical basis for the practical components of teacher education programs (3).

This independent/mediator variable—academic qualification—thus seems to have an influence on washback intensity. It is important to recapitulate here that teacher participants of this study come from three different educational backgrounds, i.e. hold three different academic qualifications: BA, BE and DT (see Chapter One section 1.3.3). At this stage of the study, one may contend that this variance in qualifications among teachers may have an impact on their degrees of reaction to the implemented exam, an issue that merits investigation in this study.

In summary, it seems that *test retrofit*, upgrading or changing (Fulcher and Davidson, 2009) on its own, while influential, is not the solitary reason in bringing about changes in teaching and learning. Teachers' factors appear to be inextricably crucial components in mediating the degree and kind of washback. "The challenge is to change the teaching culture: to open teachers' eyes to the possibilities of exploiting the exam to achieve positive and worthwhile educational goals" (Lam, 1995: 96). Moreover, "constant guidance and support over time are essential in order to help teachers use the system appropriately and therefore create positive washback" (Muñoz and Álvarez, 2010:33).

### **2.6.2 Exam-related Factors**

The exam introduced in an education system seems not of less importance than teachers' characteristics in determining washback discussed above. Studies on exam influence, specifically from external tests, mention factors pertinent to the exam itself. These include: a) exam proximity—the closer the test time is, the more its stakeholders are affected by it; b) the level of stakes of an exam—the higher the stakes of a test, the more likely the teachers' perceptions of instruction planning and

curriculum use will be affected by such a test; c) the status of the language the exam tests (e.g. English language in our case), the purposes and uses of the exam scores and the results on which important decisions may be made; and d) its familiarity to teachers and students. All are deemed important in determining the occurrence, degree and quality of washback (Shohamy *et al.*, 1996; Spratt, 2005; Wall, 2012). These factors likely exhort teachers and students to exert themselves, devoting substantial instructional time towards teaching and learning for the exam for one goal: passing the exam, and hence, mediating testing washback.

Shohamy *et al.* (1996) highlight that the higher the stakes involved with a test, the more likely it will generate washback. As a low-stakes test, the ASL test had a very limited washback, whereas the EFL oral test was a high-stakes test. It was also revealed that the L1 reading test had zero washback effect as the uses to which the test scores were put were not central to decision-making, and therefore the exam had become obsolete (Shohamy, 2001). Thus, it is evident that the exam introduced in an education system is pivotal in detecting washback impact.

Furthermore, the intensive use of high-stakes standardized tests, i.e. external testing, in many cases, has resulted in increasing teachers' workloads and gradually moving from applying meaningful creative methods of teaching to rigid traditional ways of teaching which can lead to rote learning rather than reasoning and meaningful learning. So, it is argued here that external tests in which teachers are not involved in their design and construction can also influence teachers' teaching and assessment practices and shift them from focusing mainly on improving learners' academic progress to focusing on passing external tests for accountability purposes. This issue

is empirically investigated in this study to see the extent to which the revised exam administered has influenced teachers' instructional practices, including assessment practices in classrooms.

### 2.6.3 Context-related Factors

Watanabe (2004) highlights the importance of context factors in mediating the process of washback. He divided these factors into two categories: *micro-context factors* (e.g. the school or classroom setting in which test preparation is being carried out); and *macro-context factors* (the society in which a test is used) (22). In relation to macro-level or societal factors such as parents or media, it is asserted that external pressure on teachers can explain why the impact of public examinations on instruction differs, especially when students' results are deemed a gauge for rating teachers' professional success (Shohamy *et al.*, 1996; Gregory and Burg, 2006; Cheng *et al.*, 2011), or "when awards or sanctions are attached to the test scores rewarding teachers of high achievers while teachers of low achieving students are punished" (Abu-Alhija, 2007: 56). Thus, as insisted by Gorsuch (2000) "the influence of students' expectations [and their parents] on teachers' instruction is potentially powerful" (685).

Regarding the micro-context factors—classrooms—Wall (2012) states that "some of the early references to washback in the language testing field assumed a direct cause and effect relationship between a test and the effects it would have in the classroom" (84). One of these effects that may interact with the introduced exam to determine its influence on teaching and learning is class size, as this may indirectly affect teaching (Watanabe, 1996). It is assumed that the bigger the class, the more likely teachers

would practice exam-related activities to save time and effort. Class size effect was also mentioned in language education literature pertinent to EFL curriculum innovations, specifically where curriculum urges teachers to take up, for instance, communicative teaching methodologies (Gorsuch, 2000; Agrawal, 2004; Waters and Vilches, 2008; Orafi and Borg, 2009). The literature indicates that teachers, in most cases, failed to implement such innovations due to some influential factors that may cause disparity between innovative curriculum and its enactment. Among others, class size is deemed the most influential—the bigger the class, the fewer communicative activities are performed by teachers; and hence, a minimum amount of such activities are practiced by learners.

Another major factor to consider here is that the grade that teachers teach in a school where the high-stake test is administered can also have a role in determining washback effects. For example, Shohamy *et al.* (1996) found that teachers who were teaching the upper-level students (the grade to which the exam was given) concentrated instruction more exclusively on the aspects tested in the exam than lower-level teachers, as upper-level students were closer to taking the exam. Similarly, Alderson and Wall (1993) found teachers in the upper grade tended to adhere to a more exam-oriented instruction to meet the exam requirements. Thus, grade levels that teachers teach may further explain why the washback effect differently impacts on language teaching and learning. In this study, this variable, as others, is considered in investigating the washback effect from the revised exam.

To conclude, according to the literature, washback cannot be treated as a simple cause-effect systematic reaction to exams, but there are other factors that interrelate

and interact with the exam that either inhibit or promote washback. The above review shows that teacher factors, among others, are more likely to influence teachers' perceptions of the impact of public examinations on their curriculum planning, implementation and practices. Thus, although it has not been the overarching goal, this study endeavors to investigate the extent to which teacher and context-related factors (i.e. gender, teaching experience, academic qualification, class size and the grades teachers teach) have had an impact on teacher participants' instructional practices and curriculum use vis-à-vis the implementation of the new revised examination (the new BECE). The motive is to explore the extent to which findings of previous studies in this area can or cannot be substantiated through the findings of this study, and to fill the gap in the literature concerning these issues, specifically that possibly associated with the role of gender, in mediating test washback.

## **2.7 Research Questions**

This section outlines in detail the research questions (RQ) mentioned in the previous chapter, section 1.2. The research method(s) to be utilized to address each research question are also stated.

**Question 1:** What is the nature and scope of the washback effect on grade 9 teachers' perceptions of aspects of teaching in relation to the new examination?

Teachers' perceptions are used to describe teachers' attitudes about and understanding of classroom practices with regard to the implementation of the revised exam. The aspects to be investigated include:



- Teachers' perceptions of the reasons for implementing the new exam;
- Teachers' attitudes towards the new exam and its quality; and
- Teachers' perceived classroom practices in response to the new exam.

The above questions will be answered mainly by conducting an attitudinal survey with teachers, supported by interviews.

**Question 2:** What is the nature and scope of the washback effect on teachers' classroom instructional behaviours as a result of implementing the new examination?

Teachers' classroom behaviours are operationally defined as teachers' actions and interactions inside the classroom. Teachers' classroom behaviours that need to be studied will include:

- The teaching methods and techniques teachers use in the classroom;
- Teachers' medium of instruction, teacher talk, teaching activities; and
- Teachers' formative assessment to their students and its relationship with the content of the revised exam.

The above aspects will be studied via interviews with teachers and inspectors of English, supplemented by documentary analysis. Also pilot observations will contribute to address these issues.

**Question 3:** What is the nature and scope of the washback effect on the content of the curriculum of English, as the change has been only on the testing syllabus?

The English curriculum is the school teaching materials of English prescribed by the ME in the context of the study (see section 1.3.1.2). Aspects to be investigated will include:

- The teaching materials teachers used in teaching (teachers' choice of textbooks, and teaching and learning resources);
- The content of the English curriculum and its objectives, and how teachers deal with vis-à-vis the exam (i.e. teachers' selection of teaching materials);
- The content of the introduced examination, and its objectives.

These issues will be addressed chiefly through documentary analysis of: the prescribed textbooks, the materials teachers might use and a sample of the revised examination. Additionally, this will be complemented by interviews with teachers and inspectors, as well as data generated from questionnaires.

**Question 4:** To what extent do the teacher and exam context related factors have an influence on the kind and degree of washback that materialized?

Teacher and context-related factors investigated in this study include: gender, teaching experience, academic qualification, class size and the grade(s) teachers teach.

This question will be answered primarily through analysing the quantitative data elicited from questionnaires completed by participant teachers.

The answers to the above four questions will contribute towards answering two further questions, question five and question six.

**Question 5:** To what extent did the new revised exam exert the intended washback envisaged by its constructors, and were there any other unintended consequences?

This question will be addressed in light of the answers generated from the above research questions, as well as the interviews and document studies.

**Question 6:** What quality of washback effect—negative or positive—did the new revised exam induce on EFL classrooms in Libyan schools?

Thus, each research question tackles one issue: RQ1 perception washback; RQ2 methodology (teaching and testing) washback; RQ3 curriculum washback; RQ4 washback intensity; and RQ5 and RQ6 overall washback.

## **2.8 Summary of the Chapter**

This chapter raises a number of themes. Empirical washback studies on high-stakes public examinations arrived at conclusions that washback does occur and can directly or indirectly impact upon classroom teaching and learning. Further, various findings of these studies indicated that washback from examinations is inextricably associated with the social and educational contexts in which an examination is administered; and hence each administered test needs a tailored-made study to investigate its washback. High-stakes public examinations have impacted some aspects of teaching and learning, varied in form and intensity (Cheng, 2005), affecting stakeholders differently. Whereas in many instances, exams do influence what teachers teach, there is a paucity of empirical data on how teachers teach, alluding to a claim that they may not always have exclusive power over what happens in language classrooms.

Thus, the studies reviewed indicate that *test retrofit* processes for one purpose or another do not always fulfill the desired consequences for an educational system as envisaged by test constructors. However, unintended/bad consequences might emerge. The literature, however, suggests that exam influence is not always direct but can be mediated through a variety of intervening variables, most importantly teacher characteristics. It appears that washback is a comparative phenomenon, in that test impact on classroom teaching is a dynamic and interactive process between a number of variables at the micro and macro levels. Thus, we may not be able to assume what factors predict washback and how it operates in one particular context. Washback therefore may be considered a phenomenon that is uncontrollable whose direction and intensity can be facilitated or hindered not only by the exam in use, but also by other factors such as contexts' and stakeholders' characteristics, issues worthy of investigating in this study.

The present study therefore is characterized not only by the investigation of the existence of washback, but also by the attempt to elaborate on reasons for the existence or non-existence of this phenomenon, and why it has taken a particular form. Importantly, the study may attempt to provide an answer to the question posited recently by Fulcher and Davidson (2012) with reference to Alderson and Wall's (1993) paper, they ask: "Maybe the question today should be: ... is washback a theory?" (4).

Furthermore, the studies reviewed in this chapter not only indicate the variety of contexts and topics of washback research, but also indicate the variety of methods used to probe the existence of washback, proving that washback is a complex,

multidimensional phenomenon, requiring more research in different contexts. The studies reviewed have used and led the researcher of this study to deploy a wider range of quantitative and qualitative research methods (i.e. triangulation) than earlier research (Wall, 2005). The methods used are essentially threefold: questionnaires given to students and teachers; classroom observation accompanied occasionally by interviews; and documentary analysis. Appendix 3 shows an overview of the most recently undertaken empirical washback studies of high-stakes public examinations addressed in this study in terms of the researcher(s), the study and the studied exam, the context, the main issue(s) addressed, and the methodological tool(s) used.

The methodologies adopted in the reviewed studies are often used differently according to the nature of inquiry, the researcher and the context (see Cheng, 2001, 2005, Wall, 2005, and Watanabe, 2004, for further reflections on methodology in washback research) (see also section 3.3.1.3). Thus, this review has informed the empirical stage of the present study in terms of adopting an appropriate research design and methodology, discussed in the following chapter.

## Chapter 3 Research Design and Methodology

---

### **3.1 Introduction**

This study is exploratory in nature, driven by research questions relating to: a) teachers' beliefs and attitudes towards a recently introduced exam in relation to their classroom instruction, perception washback; b) the influence of this exam on teachers' actual classroom instructional practices, methodology washback; c) the effects of this exam on the teaching curriculum and materials teachers use in classrooms, curriculum washback; d) investigating how teachers' and context's characteristics may have an influence on the degree of effect induced by the exam, intensity washback; f) whether the exam has achieved the intended consequences—intended washback—as assigned by its constructors, and whether there are other

consequences which have emerged; and f) whether the intended or unintended consequences, when observed, are deleterious or beneficial to EFL teaching in the context of the study. To address these issues, the researcher believed that a meticulously appropriate research design is essential.

Thus, this chapter presents the research design and strategy that steered the methodological processes for this study. First, it introduces the research design; this includes the key philosophical paradigms and ways of thinking in terms of ontology, epistemology and methodology, and taking the appropriate philosophical stance for the study. Second, it discusses the methodological processes, encompassing the approaches and strategy used for this research, including the research methods deployed in the process of collecting and analysing research data. Importantly, this chapter accounts for issues of validity and reliability; besides, the accessibility of the research site and its participants, shedding light on difficulties and challenges encountered during data collection procedures.

### **3.2 Research Design**

An explicit research design has been important as it paves the way for carrying out the study, informs the researcher where to go, what to do and how to get there, and hence ensures “that the evidence obtained enables us to answer the initial question as unambiguously as possible” (De Vaus, 2001: 9). Furthermore, as outlined by Creswell (2009: 5), research design involves three main components: a) the intersection of philosophy (i.e. research paradigm); b) strategies of inquiry (i.e. methodology); and c) specific methods for data collection. These issues are dealt with in the following sections.

### **3.2.1 Research Philosophy and Paradigm**

The research philosophy you adopt contains important assumptions about the way in which you view the world. These assumptions will underpin your research strategy and the methods you choose as part of that strategy (Saunders *et al.*, 2009: 108).

One key element of methodology is to adopt either a qualitative or quantitative strategy, or both in tandem. Equally important is the philosophical stance and epistemological and ontological perspectives (Guba and Lincoln, 1994) a researcher adheres to in his/her research study. These perspectives have a fundamental role in underpinning philosophical perspectives and the chosen methodology and, therefore, deciding methods of data collection (Cohen *et al.*, 2007). So, choosing a research paradigm is not considered optional, but essential, and “questions of method are secondary to questions of paradigms” (Guba and Lincoln, 1994: 105).

#### **3.2.1.1 Paradigm**

A paradigm is “a basic set of beliefs that guide actions” (Guba, 1990: 17). Other authors have used different terms to refer to the same notion: epistemologies and ontologies (Crotty, 1998); broadly conceived research methodologies (Neuman, 2000), and worldview (Creswell, 2009). Whatever term is used, all concur that research is undertaken within a research paradigm, “a set of assumptions about how the issue of concern to the research should be studied” (Henn *et al.*, 2006: 10).

Any paradigm comprises three guiding principles: ontology; epistemology; methodology (Guba and Lincoln, 1994). Ontological beliefs concern reality (i.e. what is the nature of the phenomenon under study?); epistemology relates to knowledge about the world, and the relationship between the knower and the known;



and methodology concerns how we gain knowledge about the world (Matsuda and Silva, 2005). Ontology, epistemology and methodology are considered paramount in educational or social research. Discussions of these concepts are provided in the next sub-sections to provide useful background on the research traditions/paradigms presented thereafter.

### **3.2.1.2 Ontology**

Ontology is concerned with the very nature of the phenomenon being studied. It relates to “the study of being, that is, the nature of existence” (Gray, 2004: 16). Creswell (1994) points out that ontology relates to the propositions we make about the existence of a phenomenon, and the nature of reality. Thus, ontological considerations of a particular phenomenon would raise queries about whether something exists as we believe it to, or independently of our beliefs.

Ontologically, there are two opposing extremes: realist and relativist perspectives (Bilgrami, 2002). The former posits that phenomenon can exist independently of human beliefs; the latter contends that beliefs about reality are subjective and personal, and different individuals interpret a phenomenon in different ways (Guba and Lincoln, 1994).

### **3.2.1.3 Epistemology**

Like ontology, the concept of epistemology is considered to be at the heart of philosophical thinking. Whilst ontology relates to the consideration of what is, epistemology concerns understanding the nature and scope of knowledge, its assumptions and bases about the status of knowledge (Schwandt, 2003), about the

knower and the known. Different paradigms (for example positivism and interpretivism) hold contrasting epistemologies as well as ontologies (Henn *et al.*, 2006) (see section 3. 2. 2).

Since the nature of any research is to gain knowledge about a phenomenon, epistemological issues must be addressed in the research design. In relation to this study, ontologically, the phenomenon studied is deemed present, but epistemologically—its scope and nature is still questioned. This entailed adopting an approach to the methodology, so knowledge could be obtained about how teachers' perceptions and classroom practices are influenced as a result of exam changes. Concerning the concept of methodology, the next sub-section briefly discusses this concept in terms of meaning and value.

#### **3.2.1.4 Methodology**

'Methodology', meaning 'the science of method' (Ernest, 1994), refers to "choices we make about cases to study, methods of data gathering, forms of data analysis etc. in planning and executing a research study" (Silverman, 2005: 99). Henn *et al.* (2006) add that research methodology also considers ethics, consequences, and the accessibility of the field in which research is conducted.

The word 'methodology' is often confused with the 'method(s)'. Henn *et al.* (2006: 9) note that "Method refers to the range of techniques that are available to us to collect evidence about the social world". Methodology, however, concerns research strategy as a whole, including, as Seale (1998: 3) notes, "the political, theoretical and philosophical implications of making particular choices of method when doing a research project". Silverman (2005) underlines that neither the

methodology of a study nor its methods, in themselves, are to be disproved – only found more or less useful.

Procedurally, the following section will discuss key philosophical paradigms/perspectives considered in social and educational research literature, adopting the most appropriate research paradigm for this research.

### **3.2.2 Key Philosophical Paradigms (Research Paradigm)**

Henn *et al.* (2006) argue that “there are two broadly divergent views of knowledge [...] which we can group as: positivist paradigm [...] and interpretive paradigm” (10). This section explicates these research paradigms, setting up the appropriate paradigm for this study, which helped identify the most appropriate methodological design. Justifications for the chosen paradigm will be discussed thereafter.

#### **3.2.2.1 Positivism**

In positivist ontology, reality and universal 'truths' are externally observable (Guba and Lincoln, 1994). That is, the “truth is out there and it is the job of the researcher to use objective research methods to uncover that truth” (Muijs, 2004: 4). It is also assumed that there is an objective reality which can be understood through an objectively investigative approach, unaffected by the researcher’s own values.

Epistemologically, it is assumed that researchers opting for the positivist paradigm use scientific methods to uncover knowledge that is highly objective and empirically verifiable. The inquirer is considered external to inquiry, meaning that the knower and the known are separable, and thus the data pooled is value-free. This is

applicable to this research, particularly in phase II where a quantitative survey questionnaire was conducted. As the research must be impartial (Creswell, 2007), the "researcher needs to be as detached from the research as possible" (Muijs, 2004: 4). To positivists, the knower reaches the known via an initially-established hypotheses which later can be received as 'facts or laws' (Guba and Lincoln, 1994), and "researchers reject or fail to reject hypotheses" (Creswell, 2007: 24). So, positivist research generally seeks to confirm (or disconfirm) hypotheses, endeavouring to increase the predictive understanding of phenomena (Drew *et al.*, 1996), and is a "theory-then-research approach" (Henn *et al.*, 2006:13).

Methodologically, research process is deductive, i.e. "researchers test an a priori theory" (Creswell and Clark, 2007: 26). Additionally, positivism is often associated with purely quantitative methods (Tashakkori and Teddlie, 1998). Nevertheless, qualitative methods can be used in positivist research, and vice versa. As such, "both quantitative and qualitative researchers use empirical observations [also interviews] to address research questions" (Johnson and Onwuegbuzie, 2004: 15). For positivists, the facts are principally elicited by deploying quantitative and objective instruments of data collection, such as survey questionnaires, tests and experiments with big samples aimed at collecting large amounts of data (which are statistically analysed and numerically synthesized), to enable generalizability and representativeness; this has been one methodological aim of this study, making this paradigm fit the purpose.

### **3.2.2.2 Interpretivism**

On the other hand, the interpretivist paradigm commences with the premise that access to reality is value-bound, dependent largely on the experiences of people

involved, and on interpretations, meanings and understandings of a phenomenon in real-life contexts. In contrast with positivism, in interpretivist ontology the social world is considered personal, internal and subjective. In addition, reality/truth or knowledge is interpreted and constructed by the inquirer, and is susceptible to changes; besides, the research seeks multiple realities (Creswell and Clark, 2007). So, reality is often an outcome of an in-depth investigation of the phenomenon under study with participants, and this is a key principle adopted for this study, particularly in phase three, with interviews and document analysis.

The interpretivist research “will tend to be small-scale and intensive. It will also usually be flexible and relatively unstructured, and based on detailed description (words rather than numbers) of what is seen and heard” (Henn *et al.*, 2006: 15). Interpretivists also prioritize to provide extra understanding of unexplored or little-explored phenomena, rather than illustrating why something happens. Thus, with interpretivism, to better understand a social phenomenon, researchers are recommended to deal with it from the standpoint of the participants involved in the research, rather than relying on establishing causal relationships of human behaviour.

The interpretive paradigm is "characterized by its concern for the individual" and "efforts are made to get inside the person and to understand from within" (Cohen and Manion, 1994: 36): knower and known are inseparable. That is to say, the relationship between the researcher and that being researched is close; for example, the researcher arranges constant visits to participants at their sites to pool up the required data (Creswell and Clark, 2007). This was applied to my research as I investigated the topic from the teachers' perspective in the local context, with the

primary participants – and the key factor in the implementation of reform, and for the washback mechanisms, related to instruction (Brindley, 2008; Bailey, 1999). Unlike positivism, the interpretivist approach to research design is inductive, in that “researchers start with participants’ views and build “up” to patterns, theories, and generalizations” (Creswell and Clark, 2007: 24). So, interpretivist methodology is associated with qualitative methods of data collection such as interviews and observations, generally used with the aim of understanding identifiable phenomena within a particular social context (Morris, 2006; Cohen *et al.* 2007). The table below succinctly compares these paradigms in terms of ontological, epistemological and methodological perspectives, as well as their approaches and axiology (the details included are drawn from pertinent literature reviewed in the previous sections. However, of course, the table is not exhaustive:

Table 3-1 Comparison of the basic elements of positivism and interpretivism

Paradigm	Positivism	Interpretivism
Ontology (reality)	Singular, external, objective, realist and pre-determined	multiple, specific, internal, relativist and changeable
Epistemology (knowledge)	based on a phenomena that are directly observable, i.e. objective: the investigator and the investigated are separable	based on meanings and interpretations that are not directly observable, i.e. subjective: the knower and the known are inseparable
Methodology	Quantitative methods of data collection are chiefly employed	Qualitative methods of data collection are often used
Approach	Deductive (Theory verification) theory-then-data approach (numerical & statistical data). Research is often conformity in nature	Inductive (Theory generation) data-then-theory approach (verbal & analytic data). Research is often exploratory in nature
Axiology	Inquiry is value-free, and researchers are unbiased	Inquiry is value-bound, and researchers are biased

As “it is indeed possible to have two paradigms, or two worldviews, mixed throughout a single research project” (Teddlie and Tashakkori, 2003: 11), the present researcher utilised the best from both paradigms (i.e. positivism and interpretivism) by using appropriate methods, in what is known as an eclectic approach (eclecticism). This has been a key principle for the research design in this study. The next section discusses the reasons behind this choice of mixed research paradigm.

### **3.2.3 Justification for Adopting a Mixed Approach to Research Paradigm**

The notion of merging two contrasting schools of thought (positivism and interpretivism in this case) in the same research study (known as the compatibility thesis i.e. accommodation between two paradigms) has recently become a feasible strategy for some research studies. This strategy has sparked off considerable debate in the literature. Smith (1983) speaks of the incompatibility of the positivist and interpretivist approaches. He argues that:

One approach takes a subject-object position on the relationship to subject matter; the other takes a subject-subject position. One separates facts and values, while the other sees them inextricably mixed. One searches for laws, and the other seeks understanding. These positions do not seem to be compatible. (12).

Other advocates of the incompatibility thesis (e.g. Guba and Lincoln, 1994), who reject the integration of two opposing paradigms, including their associated methods, in a single research study, underscores the contrast between positivist and interpretivist paradigms and contend that the two preclude each other.

Regarding the compatibility thesis, to which the current researcher adheres, Denzin and Lincoln (2003) point out that "within the past decade, the borders and boundary

lines separating these paradigms and perspectives have begun to blur" (246).

Likewise, Lincoln and Guba (2003) note that:

various paradigms are beginning to 'interbreed' such that two theories previously thought to be in irreconcilable conflict may now appear, under a different theoretical rubric [eclecticism in this case], to be informing one another's arguments (254).

Moreover, Howe (1998) and Johnson (1994) note the compatibility of positivism and interpretivism, in one model, where researchers make use of both quantitative and qualitative methods. This means that a researcher may employ a quantitatively data-based questionnaire at one phase of research and use qualitatively data-based interviews and/or observations at another phase of the same study, or use the former and latter interchangeably, i.e. triangulation; this method is applied in the current study.

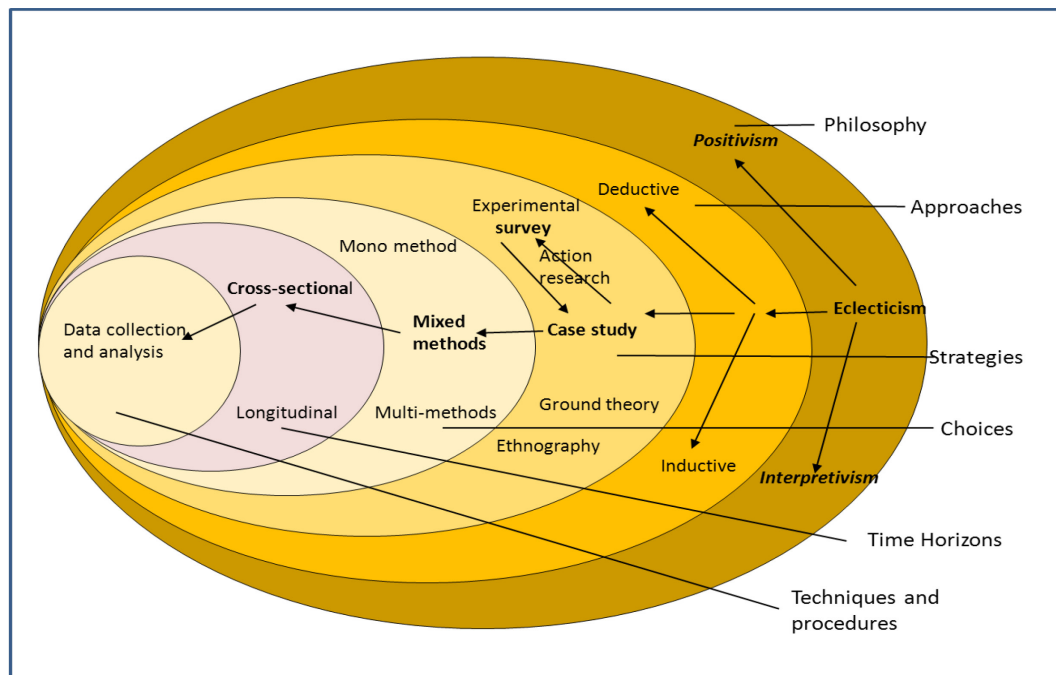
Miles and Huberman (1994) uphold this view: "we have to face the fact that numbers and words are both needed if we are to understand the world" (40). Tashakkori and Teddlie (1998: 5) recommend researchers to use "whatever philosophical and/or methodological approach that will work for the particular research problem under study", as in some cases adopting a single paradigm would not suffice.

This study was a confluence of the two paradigms; it is neither radically positivist nor purely interpretivist. The rationale behind this strategy was that while one perspective enables the researcher to gain generalizable data, the other helps collect deep and rich data (Johnson and Onwuegbuzie, 2004). Based on this choice, assuming that washback is often deemed complex and multidimensional (Alderson and Wall, 1993) the current researcher decided on a mixed-methods approach of research design as a key element in designing the study. The overall methodological



process adopted for this research passes through particular stages of the research “onion” model developed by Saunders *et al.* (2009) (see figure 3.1): research philosophy; approach; strategy; choices; time horizons; data collection and data analysis. According to this study, the first layer shows the research philosophy adopted: positivism and interpretivism; the second layer refers to the research approach that comes out of this chosen philosophy (deductive and inductive), constituting the research design. While the third layer concerns the strategy of the research, adopting a case study as an overall strategy complemented by survey strategy, the fourth shows the proposed approaches/methods (i.e. mixed methods and triangulation). Both third and fourth layers constitute the overall methodology. The fifth layer refers to the time horizon, adopting cross-sectional research. Lastly is the process of data collection and data analysis techniques. The arrows drawn in the figure depict the directions of the methodological steps of the research study.

Figure 3-1 The research "onion" model

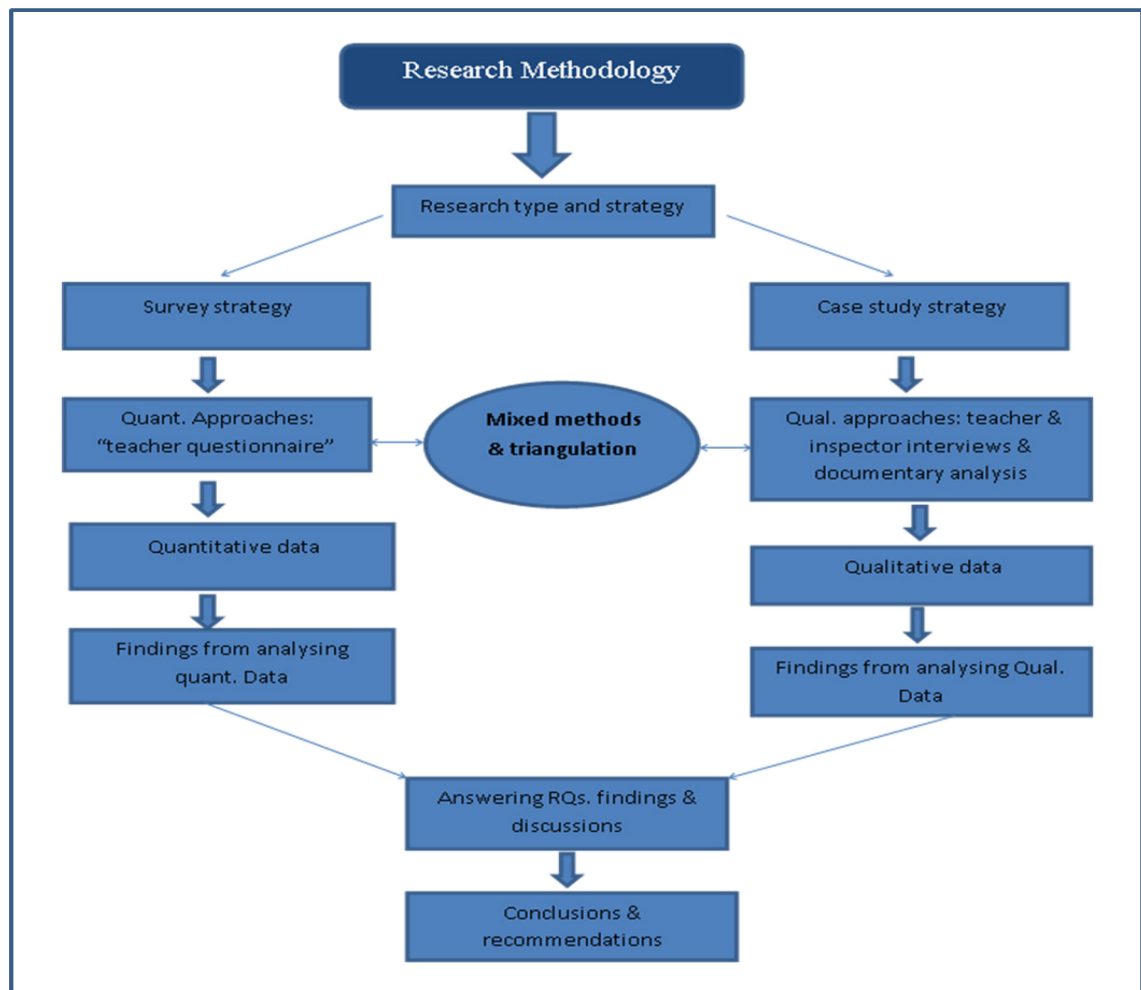


(Adopted from Saunders *et al.*, 2009: 108)

### 3.3 Research Methodology

Research methodology is described as the research strategy including the type of research, choice of research approaches and methods (Henn *et al.*, 2006). This section introduces the type of research that characterizes this study, and presents the choice of approaches and methods the current researcher made use of, and the selection of the research sample. It also portrays the research phases. This section also sheds light on the pilot study conducted, and its implications for the main study, considering issues of validity, reliability and participants' accessibility. Figure 3.2 depicts the road map of the methodology adopted in this research.

Figure 3-2 A road map of the research: overall methodology.



### **3.3.1 Research Type and Strategy**

Creswell (2009) identifies two strategies of inquiry: quantitative strategies that include survey research or experimental research; and qualitative strategies that may include one or more of the following: ethnography, grounded theory, case studies, phenomenological research and narrative research. Since the study adopted a mixed approach, discussion in this section is devoted to one type of each category identified by Creswell: the "survey", i.e. teacher questionnaires, and "case study" i.e. interviews with teachers and inspectors, as they were employed in the study.

The purpose of choosing a survey questionnaire and case study for the present study is associated with the assumption that the former will provide ample data, elicited quantitatively via questionnaires (phase II), aimed at procuring generalizable data, which in turn will form the basis for the data that will be pooled through qualitative methods, teacher and inspector interviews, and classroom observations (the latter of which were planned but not executed due to the uprising in Libya during the data collection stage; see sections 3.7 and 3.8), complemented by documentary analysis. Both were the focus of the case study (phase III). Sections 3.3.2.2 and 3.3.2.3 discuss in detail the employment of the two strategies used in this study.

#### **3.3.1.1 Qualitative and Quantitative Approaches**

The primary rationale for combining quantitative and qualitative research methods was to adequately address the research questions, endeavouring to overcome problems that might arise, such as the lack of validity and representativeness of quantitative data, and the generalizability and reliability of qualitative data

(Hammersley, 1992; Silverman, 2005). This combination was worth deploying in this research as it can yield desirable outcomes:

- Corroboration: the same results are derived from both quantitative and qualitative methods.
- Elaboration: the qualitative data analysis exemplifies how the quantitative findings apply in particular cases.
- Complementarity: the qualitative and quantitative results [may or may not] differ but together they generate insights.  
(Brannen, 2005: 176).

Considering these points and admitting that merging quantitative and qualitative approaches in the same study is effective for producing rigorous data, the simultaneous use of qualitative and quantitative, qualitative followed by quantitative, or quantitative followed by qualitative data are possible combinations (Morrison, 2007; Teddlie and Tashakkori, 2009). This research made use of both, sequentially (quantitative followed by qualitative) “because going back and forth between qualitative interpretations and quantitative analysis is explicitly seen as yielding important insights concerning the phenomena [washback in this case] under study” (Rocco *et al.*, 2003: 596-97). However, the use of qualitative methods outweighed that of quantitative methods because direct questionnaires (i.e. those containing close-ended items) preclude participants to show their real attitudes and perceptions in response to an inquiry while qualitative data elicited via interviews, observations and content analysis could circumvent this (Baker, 1992).

Integrating quantitative and qualitative approaches in the same research study entails a convergence of research methods. The following section discusses this in relation to the research undertaken.

### 3.3.1.2 Mixed Methods and Triangulation

Mixed methods research design is defined as "the collection or analysis of both quantitative and qualitative data in a single study in which the data are collected concurrently or sequentially" (Creswell *et al.*, 2003: 212). Given the nature of the research questions posed, quantitative data was first collected via teacher questionnaires, followed by the collection of qualitative data through interviews with teachers and inspectors, and document study, i.e. reviewing textbooks and analysing exam papers. This design was adopted because one method by itself may not suffice (Cohen *et al.*, 2011), as "each method has its own strengths and weakness" and "one's weaknesses are often the other's strengths" (Henn *et al.*, 2006:128).

Another feature that characterised the design of this research is triangulation, "the combination of methodologies in the study of the same phenomenon" (Denzin, 1978: 291). Two approaches of triangulation were employed, namely: data triangulation, resorting to more than one source of data to answer a question (from teachers, inspectors and documents); and methodological, i.e. between methods triangulation (Denzin, 2009), using more than one tool to collect data (questionnaires, interviews with teachers and inspectors, and document study) (see figure 3.3 below, section 3.3.2.1).

Thus, the current researcher adopted a mixed-methods approach of research design as a key element in designing the study. Prior to explicating and justifying the chosen research methods used to collect the research data, the researcher saw it as essential to shed light on research methods conventionally adopted by researchers and documented in the washback literature.

### **3.3.1.3 Methodological Considerations from Washback Research**

The washback literature reviewed in Chapter Two and outlined in appendix 3 showed that a wide range of methods have been used to assemble data to provide concrete evidence for detecting washback effects on aspects of teaching and learning. The studies tended to involve questionnaires, interviews, classroom observations, diaries and testing measures. These methodologies are often used differently – separately or in tandem – according to the nature of inquiry, the researcher and the context.

The washback literature underlines the complexity of detecting washback effects from examinations (Tsagari, 2009b; Wall and Alderson, 1993; Wall, 2005). This makes researchers use various research methods and different sources of data in separate stages of the research process. Wall and Alderson (1993), for instance, underscore the importance of mixing research methods to address questions such as "why the teachers do what they do, what they understand about the [...] examination and what they believe to be effective means of teaching and learning" (62). To tackle these questions, the authors state that "observations on their own cannot give a full account of what is happening in classrooms". Therefore, they concluded that "it was important for us to complement the classroom observations with teacher interviews, questionnaires to teachers and teacher advisors, and analyses of materials (especially tests) teachers had prepared for classes" (63). Similarly, Cheng (2001: 20) points out that "survey data alone are useful but insufficient for understanding washback ... we need to ask teachers and watch them teaching".

The methods used in washback studies can be classified as indirect methods such as questionnaires, tests and dairies, and direct methods such as interviews, observations and documents (Alderson and Wall, 1993). Some studies of washback have used only indirect methods (e.g. Amengual-Pizarro, 2009; Andrews, 1995; Lam, 1995; Cheng 1998; Stecher *et al.*, 2004; Choi, 2008), which are questionable as they provide data based on self-reported anecdotes; while other studies (such as the present one) use indirect and direct methods in tandem (e.g. Burrows, 2004; Ferman, 2004; Cheng, 1997, 1999; El-Ebyary, 2009; Munoz and Alvarez, 2010; Saif, 2006; Watanabe, 1996, 2000; Wall and Alderson, 1993), the results of which reveal facts about phenomena that might not be revealed using only indirect methods.

The present researcher has employed two approaches of triangulation: data and methodological triangulation. This was to allow a fuller understanding of the phenomenon studied, because the researcher believed that using only one method would provide insufficient evidence, regardless of its usefulness (Wall and Alderson, 1993; Bailey, 1999; Watanabe, 2004; Cheng, 2005).

In summary, for the purpose of this research, survey questionnaires (phase II of the study; see 3.3.2.2) were used to give a general understanding of how teachers perceive the impact of the revised exam on curricular planning and instruction. Taking into account the limitations of using a questionnaire (i.e. limited scope and flexibility of the data collected (Cohen and Manion, 1994; Baker, 1992), the present researcher looked for other methods that could help collect data and provide insights into the influence of the exam. Therefore, teacher, inspector and Examination Board director interviews, as well as document analysis—qualitative data (phase III)—were

conducted to provide rich and detailed information about whether what happens tallies with what the teachers report in questionnaires; and whether further aspects of washback on teaching can be unveiled. This is in order to allow an exploration of “the relationship between testing and classroom teaching activities” (Cheng, 2001: 24). Last but not least, use of a multi-faceted approach, as Hyland (2002: 194) notes, “helps overcome the deficiencies of any particular method as well as offering a way of cross-checking data”. The methods used, which are discussed in the sections that follow the next section, complement each other in this research.

### **3.3.2 Sample and Methods of Data Collection**

#### **3.3.2.1 Sampling Frame and the Selection of Participants of the Study**

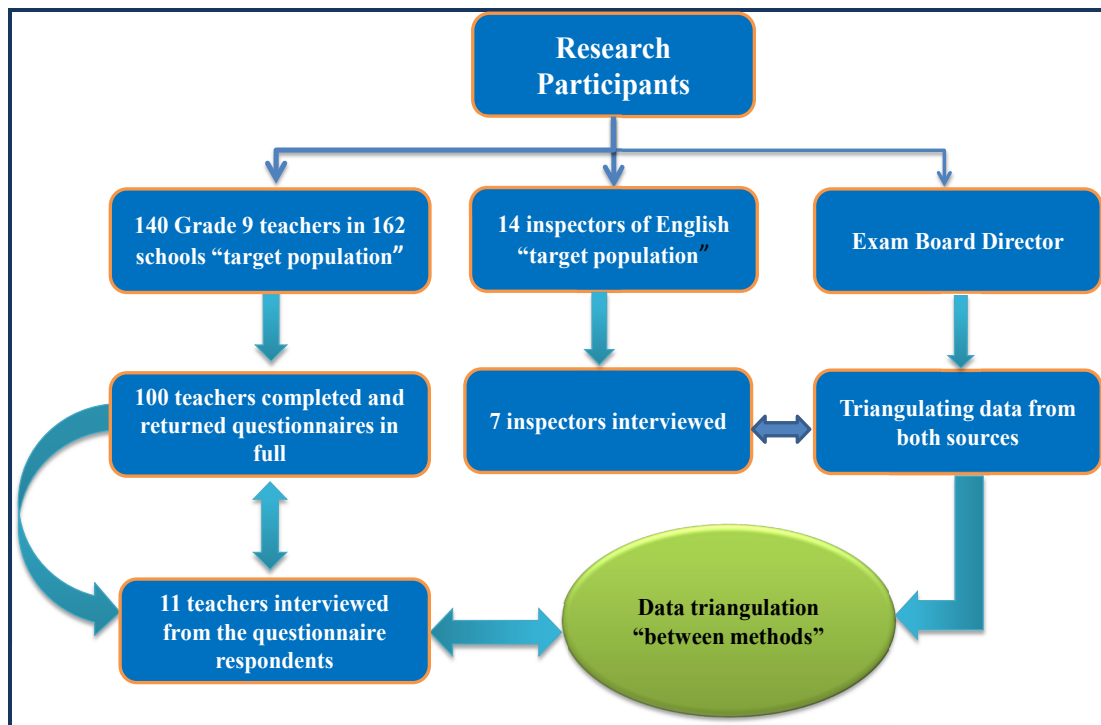
“[T]he listing of all units [i.e. people] in the population from which the sample will be selected” is referred to as a sampling frame (Bryman, 2008: 168). Two kinds of population, wider and target are often dealt with when conducting educational research. The former refers to the larger group of people/units who share the phenomenon under study; the latter is a part of the wider population, and refers to the sample upon which the results might be generalized (Larson-Hall, 2010). The wider population in this study refers to all Libyan teachers of English who teach grade 9 school students (to whom the revised exam under study is administered). The target population refers to a subset of the wider population, one-hundred-and-sixty-two EFL teachers teaching grade 9 classes in the schools of the chosen city from whom the sample involved in this study has been randomly selected (see figure 3.3).



As the target population of teachers in the chosen city is largely ascertainable in the context of the study, and for generalizability purposes, the selection method for participants was simple random sampling (probability sampling), aiming to: make statistical inferences (Robson, 2002; Creswell, 2005), give each teacher an equal chance of being selected (Fink, 2003), and hence keep the sampling errors to a minimum (Bryman, 2008). The sampling for teacher questionnaires was carried out following Bryman's (2008: 172) method of selection. The steps followed were:

- Defining the population. This was all grade 9 teachers in Basic Education schools in Misrata (N=162 in 162 schools).
- Selecting a comprehensive sampling frame taken from the records of the inspectorate office in the authority of education.
- Deciding on the sampling size (n): 140 teachers in about 140 schools.
- All schools stated in the list of schools were assigned consecutive numbers from 1 to (N). In my case, this ranged from 1 to 162.
- Random numbers were generated using a computer programme, through the internet (Research Randomizer: <http://www.randomizer.org/form.htm>), to select 140 different random numbers between 1 and N (162), sorted from lowest to highest.
- The schools/teachers, to which the n (140) random numbers refer, constituted the sample of teacher questionnaires.

Figure 3-3 Selection of participants and data triangulation



As Robson (2002) argues, it is not possible to deal with the whole of a population (i.e. all Libyan teachers of English teaching grade 9) in a survey, so the present researcher has chosen schools situated in a north-western Libyan city, Misrata, the third largest city in the country in population and size. The schools of this city were selected because of their convenience in terms of distance and time, i.e. the researcher could access them easily, as they are linked by a good trunk road; so logistical factors such as proximity and accessibility had an impact on the choice of city. Also it is the researcher's hometown and he worked in its schools as an English teacher for several years. This, in turn, enabled the researcher to know the locations of several schools, and to access them easily; besides, the researcher's twelve years of teaching (1995-2006) in Misrata's schools has enabled the researcher to build a good relationship with teachers, some of whom in later years have become inspectors of English. As a result, this had a good impact on obtaining high return rates for

teacher questionnaires, and positively influenced the willingness of teachers and inspectors to take part in the interviews cooperatively, especially regarding inspectors who had a substantial role in distributing questionnaires to teachers in a short period of time (see 3.6 and 3.8 for elaboration).

For the main study, 140 questionnaires were distributed to 140 teachers in 140 state schools. Of them, 118 questionnaires were returned to the researcher, 100 of which were completed in full, with a final return rate of 71.4%, which is deemed a high response rate in educational research (Cohen *et al.*, 2011). As the educational system in the context of the study is centralised, all Libyan public schools have the same educational facilities and resourcing, and most English teachers in these schools are from the same socio-economic and cultural backgrounds. The chosen sample, thus, not only indicates that this sample will not have a negative effect on the representativeness of the national population, but also that data generated from the questionnaires can be generalized to the wider population (Fink, 2003; Cohen *et al.*, 2007; Dörnyei and Taguchi, 2010).

As discussed in Chapter One section 1.3.3, that EFL teachers in Libyan schools are either graduates from colleges of arts or from colleges of education. Also, a few teachers hold a Teacher's Diploma. Table 3.2 below presents the descriptive data related to the surveyed teachers' characteristics. As teaching is considered a popular profession among women in the context of this study, it was not surprising that most (74%) of the sample teachers were female, and 26% of teachers were males. The sample consisted of 70 teachers (70%) who hold a Bachelor of Arts, 19 teachers (19%) who hold Bachelor of Education and 11 teachers (11%) who hold a Teacher's

Diploma. The table shows that most teachers (74%) their teaching experience is ranging from seven years to over fifteen years. The table also shows that grade 9 teachers also teach other grades. Of the 100 teachers, 29 teaching grade 5, 15 teaching grade 6, 24 teaching grade 7 and 32 teachers teaching grade 8. Also, 85% of teachers had 31 – 40 students in class and 15% had 20 – 30 students in their class.

Table 3-2 Demographics of surveyed teachers

<b>Characteristics</b>	<b>Number</b>	<b>Percentage</b>
<b>Gender</b>		
Female	74	74.0
Male	26	26.0
Total	100	100.0
<b>Academic qualification</b>		
Bachelor of Arts (BA)	70	70.0
Bachelor of Education (BE)	19	19.0
Teaching Diploma	11	11.0
Total	100	100.0
<b>Years of teaching experience</b>		
0 – 2	0	0.0
3 – 6	26	26.0
7 – 10	21	21.0
11 – 15	34	34.0
Above 15	19	19.0
Total	100	100.0
<b>Grades teachers teach other than grade 9</b>		
Grade 5	29	29.0
Grade 6	15	15.0
Grade 7	24	24.0
Grade 8	32	32.0
Total	100	100.0
<b>Class Size (number of students)</b>		
20 – 30	15	15.0
31 – 40	85	85.0
Above 40	0	0.0
Total	100	100.0

It should be noted here that the differences in the curriculum studied by student teachers of the three certificates (see section 1.3.3) could have an influence on their perceptions and classroom instructional practices, as a result of administering the revised exam. Therefore, the teacher's qualification was treated as an independent variable in the data analysis procedure. Furthermore, teachers' classroom

instructional practices and conceptions are often formed in relation to their time teaching English in schools. Therefore, practical experience of teaching was considered as another independent variable to be examined in relation to the exam influence. In addition, the direction and magnitude of the relationship between participants' gender and the exam impact will be examined. In other words, the degree of effect of teacher's gender, academic qualification, teaching experience and the grade(s) teachers teach are addressed during the questionnaires data analysis, and then reported in the findings in the forthcoming chapter, section 4.2.4.

As interviews and document analysis (in addition to classroom observation in the pilot study) constitute the case study strategy/design adopted in the third phase of the present study, in which researchers typically use a small sample, and "the persons to be included within the group must be distinguished from those who are outside" (Yin, 2003: 24), non-probability purposive sampling was used to select 11 teachers for interviews, from the questionnaire respondents (i.e. using data triangulation). Figure 3.3 below shows participant selection and data triangulation as applied in this study. The reason for deploying purposive sampling for the interviews was to enable the researcher to select "individuals who can provide rich and varied insights into the phenomenon under investigation so as to maximize what we can learn" (Dörnyei, 2007: 126). Table 3.3 portrays the characteristics of the interviewed teachers relating to their gender, academic qualification, teaching experience and grades they teach other than grade 9. It should be noted here that due to the conflict in Libya, it was not possible for the researcher to select more teachers and inspectors for the interviews. Also, due to religious/Islamic considerations (males and females should not meet alone unless they are in groups or with their close relatives) only three female

teachers were available for interviews as schools, which were the proper place to conduct such interviews, were closed (the three female teachers were interviewed in their houses in the presence of their close relatives, fathers and husbands).

Table 3-3 Characteristics of interviewed teachers

<b>Characteristics</b>	<b>Number</b>
<b>Gender</b>	
Female	3
Male	8
Total	11
<b>Academic qualification</b>	
Bachelor of Arts (BA)	5
Bachelor of Education (BE)	4
Teaching Diploma	2
Total	11
<b>Years of teaching experience</b>	
7 – 10	2
11 – 15	5
Above 15	4
Total	11
<b>Grades teachers teach other than grade 9</b>	
Grade 5	4
Grade 6	3
Grade 7	2
Grade 8	2
Total	11

The secondary participants for the study, deemed a useful source of data, were inspectors of English. As noted in chapter one, inspectors of English usually visit the same teachers in their schools, who have participated in completing the questionnaires and been interviewed. Due to the small number of inspectors compared with the number of teachers, the convenience sampling method was used as the sampling was restricted only to inspectors who visit and supervise grade 9 teachers in the schools of the chosen city. The target population were 14 male inspectors working in the inspectorate office at the time of the empirical study (no females were assigned as inspectors in the inspectorate office). As noted by

Oppenheim (2000) that sampling requires compromises between theoretical sampling and practical constraints such as time, and accuracy is more important than sample size, of the 14 inspectors, 7 were available for interviews. Also, the Examination Board direct was also interviewed in order to gain important information that contribute to address the research questions (see section 3.3.2.3.1.1).

The discussion now focuses on the choice of methods used to collect data. To recapitulate, the researcher has opted for two main strategies, survey strategy and case study strategy. While the former employs teacher questionnaires, the latter uses teacher and inspector interviews as well as documentary analysis.

### **3.3.2.2 Survey Strategy**

Survey research is a strategy for gathering information about beliefs, attitudes, perceptions and behaviour in relation to a studied issue (Creswell, 2009). For this study, cross-sectional survey questionnaire design was adopted as the main instrument of data collection in the second phase of the study, because this type of design is seen practical for doctoral research studies (Nardi, 2006; Wiersma and Jurs, 2009; Cohen *et al.*, 2011). It is also a useful design for “comparing the different values on key variables possessed by groups of cases; rather than possessed by any particular case” (Aldridge and Levine, 2001: 31).

#### **3.3.2.2.1 Teacher Questionnaire**

A questionnaire was one of the major instruments used in this research study. The teacher questionnaire is discussed below in terms of rationale and strategy, design structure and content, and validation procedures.

### ***Rationale and Strategy***

Questionnaires are used in order to generalise from a sample to a population, and to find out and make inferences about people's attributes, perceptions and behaviour (Punch, 2005; Bryman, 2008; Creswell, 2009). Moreover, data elicited by questionnaires enables standardised, objective comparisons to be made systematically (Fraenkel and Wallen, 2005).

Teacher questionnaires provided a sampling frame for the interview study (Brannen, 2005), and thus provided exploratory data important as a basis for data that was thereafter qualitatively pooled, specifically building an idea of which questions might be included in interviews. Moreover, survey questionnaires are not only economical temporally, compared with other methods, they also take into account assurance and confidentiality and require respondents "to reveal information about feelings, to express values, to weigh up alternatives etc., in a way that calls for a judgment about things rather than the mere reporting of facts" (Denscombe, 2003: 146).

Nevertheless, questionnaires have drawbacks, especially when used in isolation; they can only provide a 'thin description' of the target phenomenon and may not provide insights into teachers' real practices (Borg, 2006; Dörnyei, 2003). Another disadvantage of using questionnaires, especially self-administered, is that they are demanding in terms of preparation and administration, and "offer little opportunity for the researcher to check the truthfulness of the answers" (Denscombe, 2003: 160).

Although they require more time and effort for handing in and collecting, self-administered questionnaires were adopted for this study as a mode of administration. The main reason to opt for this strategy is to increase return rates (Brown, 2001;



Dörnyei and Taguchi, 2010). The strong social ties among the members of Libyan society helped increase the response rate of questionnaires. It was not possible to apply other forms of questionnaires because of difficulties encountered, i.e. the high cost of telephone calls and the poor internet facilities which currently exist in Libya; the majority of Libyan teachers of English do not use the internet or exchange emails due to socially and culturally embedded factors (Twati, 2008: 2). However, in the pilot study, the mode of administering teacher questionnaires was via emails, as the piloting was conducted in the United Kingdom, which has very good broadband services, and all targeted teachers have become familiar with using the internet.

### ***Design and Content***

The current researcher considered many factors at the time of constructing the questionnaire. The first was to clarify the reason for asking specific question(s), and how the questions/items were related to the main research questions. More specifically, a focus was made on the research questions posed before thinking about questions and items to write in the questionnaire in order to assure that all research questions were covered. Moreover, the questionnaire questions and items tended to be arranged in the order in which the research questions were ordered, so that certain themes could easily be transcribed and categorised for data analysis.

Closed-response question type covered almost all questionnaire questions, to which teachers will respond with Likert-style answers, allowing for quantitative data in the form of descriptive statistics. They impose restrictions on respondents' answers (This drawback was mitigated by the interviews where informants had options to

voice their opinions in detail). However, they assist in producing quick results which lead to quick analysis and faster feedback (Cohen *et al.*, 2007; Sarantakos, 2005).

Additionally, as the types of questions used in questionnaire design can be classified into a number of categories, as suggested by Patton (1987), encompassing demographic questions and questions about perceptions and practices, the teacher questionnaire (TQ) (see appendix 4) for the main study was designed in English, comprising three parts. Part One (TQ 1-10) comprised ten questions to elicit general information about participants, covering demographic details such as gender, age, academic qualification, years of experience in teaching, and current teaching situations, such as grades they teach, teaching load per week, class size and whether they have been involved in workshops related to the revised exam or undertaken training on language testing and evaluation.

Part Two entails six questions comprising 35 items (TQ 11 – 16). All items were designed using a five-point Likert scale; from strongly disagree to strongly agree. After the first three questions respondents were given room to add relevant information. This part of the teacher questionnaire was designed to elicit attitudinal data, addressing the first research question, and to understand teachers' perceptions of their classroom practices in relation to the implementation of the revised exam. Part Three of the questionnaire was designed to gather behavioural data. This part consisted of 14 questions and comprised 77 items altogether (TQ 17 – 30), split into two sections. Section one consisted of eleven questions (TQ 17 – 27) and comprised 65 items. This section mainly dealt with the second research question, relating to instructional, teaching and testing, that may be affected by exam change. The second

section consisted of three questions and comprised twelve items (TQ 28 – 30). This section was designed to gather data that provides insights about the impact of the revised exam on the content of teaching and curriculum, addressing the third research question. Part-three questions were designed using both a multiple-choice format and five-point scale of frequency.

The questionnaire allowed anonymity, in that respondents were informed that all data will be strictly confidential, to encourage frankness (Robson, 2002). As a purposive sample of the teachers who participated in answering the questionnaire were involved in answering the interview questions, the questionnaires given were not totally anonymized but contained codes linked to teachers' names and schools. This enabled the researcher afterwards to contact the teachers for interviews.

### ***Validation Procedures***

To ensure the validity of the teacher questionnaire two stages of design were undertaken: careful qualitative input, to ensure content validity; and piloting, to ensure construct validity. The questionnaire qualitative input design is largely based on theoretical considerations from relevant literature reviewed in Chapter Two. Thus, the questionnaire model for the study is an adaptation of Cheng's (2005: 256-261) teacher questionnaire used in her Hong Kong washback study. There are two main reasons to adopt this model: it has been successfully used in a previous washback study where teachers were anticipated to change their teaching as a result of exam change. Secondly, the exam that Cheng studied is similar to the current case in terms of type and importance (both are high-stake public examinations administered as school-leaving exams at the national level, albeit in different contexts). Thus,

Cheng's questionnaire was adapted to accord with the research aims and questions, and the research context.

Most questions/items in Cheng's questionnaire (see appendix 34) related to learners or learning and policy-making issues were dismissed because they were beyond the scope of this research. For instance, questions 7, 8, and 9 in part two of Cheng's questionnaire were removed because they were designed to elicit information about students' language learning and learning motivation. Other questions were amended to meet the context of the research. For example, the idea of TQ11 (also TQ16) for the questionnaire of this study was derived from Cheng's (i.e. part two, TQ1), but the items of this question were rephrased based on the announced intentions of the exam constructors when designing the test. Additionally, most of the questions in Cheng's questionnaire (e.g. part two questions 2, 3, 4, 5, 6 ... etc.) were used verbatim in this study's questionnaire (e.g. 12, 14, 17, 18, 19, 20, 28 and 29) as they fit the purpose of this study. However, some questions (e.g. 13, 15, 25, 26, 27 and 30) of the teacher questionnaire were designed by the researcher based on theoretical considerations from relevant washback studies reviewed in Chapter Two.

The second method of validation was piloting. Section 3.3.5.2 outlines how piloting the teacher questionnaire contributed to its validation.

### **3.3.2.3 Case Study Strategy**

The second strategy adopted for the present study was the 'case study strategy'. A case study is flexible in dealing with complex issues (Denscombe, 2003), and "benefits from the prior development of theoretical propositions to guide data collection and analysis" (Yin, 2009: 18).

Creswell (2009) defines a case study as a research strategy in which the undertaker investigates intensively an issue or activity related to one or more individuals. “The case(s) are bounded by time and activity, and researchers collect detailed information using a variety of data collection procedures over a sustained period of time” (227). Additionally, Yin (2009: 18) describes a case study as “an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident”.

There are several kinds of case studies (see De Vaus, 2001: 221-24; Creswell, 2007: 74; Yin, 2009). Yin (2009) identifies three: exploratory, descriptive and explanatory. When the “how” and “why” questions are posed in investigation, the explanatory case study strategy is most suitable. On the contrary, an exploratory case study is usually favoured when the study inquiry focuses on the “how” questions. Descriptive case studies seek to provide a detailed narrative description of the actions and interactions related to the inquiry; thus, it focuses on the “what” question. As the boundaries between these strategies are not always sharp, and “there are large areas of overlap” (Yin, 2009: 9), the present study tended to fall in the scope of the exploratory type, and took some features of the descriptive one as it endeavoured to investigate how teachers act and react to syllabus testing changes.

Case studies have five different applications (see Yin, 2009: 19-20 for more discussion); “the most important is to explain the presumed causal links in real-life interventions that are too complex for the survey or experimental strategies” (Yin, 2009: 19). Moreover, case studies can be applied to “describe an intervention and the real-life context in which it occurs” (Yin, 2009: 20). Another reason for

adopting this kind of research strategy is that the focus was on one particular context dealing with specific issues, with an aim to “illuminate the general by looking at the particular” (Denscombe, 2003: 31), a process which case studies can deal with, unlike some other strategies (Denscombe, 2003). In the present study, eleven grade nine EFL teachers and seven English inspectors represented the case, through interviews. This was complemented by documentary analysis. It was planned to complement this strategy by conducting classroom observations in the main study but this was not possible due to the constraints discussed in section 3.8.

#### **3.3.2.3.1 Interviews**

An interview is “a purposeful interaction between two or more people focused on one person trying to get information from the other person” (Gay and Airasian, 2003: 209). Berg (2009: 101) succinctly defined interviewing as a “conversation with a purpose”. In this research two sources of data were used during the case study stage, namely teachers and inspectors, and the method used to collect data was interview. As discussed above in section 3.3.2.1, eleven EFL teachers and seven English inspectors were selected for interviews. This section discusses how these interviews were conducted, highlighting rationales and procedures.

#### ***Rationale***

The purpose of employing interviews as a source of information hinges upon the “the richness of interview data [which] leads to more possibilities in terms of exploring the issues involved” (Brown, 2001: 78). Interviews were principally employed because their flexibility “allows the interviewer to explore new avenues of opinions in ways that a questionnaire does not”, and their personal nature “may encourage

interviewees to be more open to express tentative or exploratory opinions, ideas, and speculation that would not come out on a questionnaire” (Brown, 2001:78).

Another advantage of using interviews in the present study was to mitigate the drawbacks of questionnaires, such as imposing restrictions on respondents’ answers, and also, as Wall and Horak (2006: 26) note, “to gather data that would offer deeper insights than questionnaires could provide, even if it meant working with fewer participants”. Thus, interviewing was more likely to foster spontaneous answers, and increase response rates.

Methodologically, interviews were used in the present research as Denscombe (2003:166) suggests:

- Follow-up a questionnaire. Where the questionnaire might have thrown up some interesting lines of enquiry, researchers can use interviews to pursue these in great detail and depth. The interview data complements the questionnaire data.
- Triangulation with other methods. Rather than interviews being regarded as competing with other methods, they can be combined in order to corroborate facts using a different approach.

The first purpose identified by Denscombe is directly related to the rationale for which teachers were interviewed, while the second is associated with inspector interviews. However, both interviews lent themselves to triangulation as a way of supplementing data (Denscombe, 2003).

### ***Procedure***

For the purpose of the inquiry of this research, semi-structured interviews were conducted individually with all participants (teachers, inspectors and the

Examination Board director). The motive was to “probe far beyond the answers to their prepared standardized questions” (Berg, 2009: 107). Another advantage of this type of interview is related to the interviewer’s ability to control the interview; it is also possible for the interviewee to have some freedom for developing the interview (Bryman, 2008; Berg, 2009).

As with teacher questionnaire, the interview questions were designed and organised in a similar order in which the research questions were designed, so that the data could be thematically transcribed and categorised, and then easily interpreted. The interview schedule of questions was given to the participants two or three days before the interviews were conducted, to prepare them for the interviews and give them a general idea of what kind of questions they were going to answer so that the interviews could run smoothly. They could choose to converse in English or Arabic. The researcher wanted to make sure that there was no language barrier for the participants and no misunderstanding between the interviewer and interviewees. Additionally, the interviewees could express their feelings freely. Of the eleven teachers interviewed, six interviewees spoke in English, four in Arabic and one in both English and Arabic. Whilst most of the inspector interviews were conducted in English, the Examination Board director interview was conducted in Arabic only. All interviews that performed in Arabic were translated into English.

Interviews were conducted with seven inspectors, one of whom was a senior English inspector with experience of more than thirty years who had recently worked as the subject expert in the inspectorate office, English language division. Also, the Examination Board director in the authority of education within the chosen city was



interviewed to complement and confirm some data elicited from teacher and inspector interviews. All interviews were audio-recorded, underpinned by written field notes, both of which offered the opportunity to check and listen to the interviewees' utterances many times.

#### **3.3.2.3.1.1 Examination Board Director Interview**

Two unstructured interviews were conducted with the Examination Board director, before and after interviewing teachers and inspectors: the first meeting was informal held in January, 2010 in order to obtain some information about the new examination policy, such as reasons of change and the way the examination was constructed and administered; the second interview was conducted during data collection for the main study in June 2011 to probe into some issues raised during teachers and inspectors' interviews regarding the intended and unintended consequences of implementing the new examination, and the extent to which the new exam has achieved its desired aims.

#### **3.3.2.3.1.2 Teacher Interviews**

Although teacher interview questions were principally based on teachers' responses to the questionnaire, given the nature of the research questions and based on the review of the washback literature (e.g. Wall and Alderson, 1993; Qi, 2004; Burrows, 2004), the researcher has prepared an interview schedule of questions for the interviews (and also a different schedule for inspector interviews) (see appendices 6 and 7). The interview questions were open-ended aiming to probe deeply into participants' ideas and perceptions (Cohen *et al*, 2000: 275), and were based on three of the 15 washback hypotheses posited by Alderson and Wall (1993): "A test will

influence what teachers teach”; “A test will influence how teachers teach”; and “Tests will have washback effects for ... some teachers, but not for others” (120-21). Moreover, the questions were designed to elicit information confirming the data gathered through the questionnaire, and to explain newly-emerging issues.

#### **3.3.2.3.1.3 Inspector Interviews**

The seven interviewed inspectors were supervising all grade 9 teachers in the chosen city who comprised the target population of this study. This means that data they provided about these teachers’ instructional practices was not confined to their observation of only participant teachers or their schools. Further, it related to their observation of the target population, i.e. all grade 9 EFL teachers teaching English in the 162 basic education schools in the city, which may increase generalizability.

Interview questions (see appendix 7) were designed on the basis of both teacher questionnaire and teacher interview schedule, with reference to the research questions. Inspectors’ interviews, thus, helped to verify and support results obtained from teacher questionnaires and interviews. They mainly aided in answering research questions two and three when they report on teachers’ classroom practices and behaviour before and after the implementation of the revised exam, compensating to some extent for not implementing the planned classroom observations, and the lack of baseline data.

Both teacher and inspector interviews were validated by various channels: some senior doctoral colleagues at the School of Education, University of Leicester, were asked to voice opinions about the interview strategy and structure. Moreover, both teacher and inspector interviews were piloted (See section 3.3.5.3 for elaboration.)

### 3.3.2.3.2 Documentary Study

According to Creswell (2005), documents are used for gathering data in mixed-methods research. Although “documents can be treated as a source of data in their own right” (Denscombe, 2003: 212), they were used in this study as a supplementary instrument to help find accurate answers to the research questions. As they “do not speak for themselves but require careful analysis and interpretation” (Cohen *et al.*, 2011: 253), some samples of teachers’ on-going tests are scrutinised in order to assess relationships between their content and the revised exam content. Should consistency be found between the two types of test, this could be analysed in terms of washback in this particular aspect of classroom instructional practices. This contributes to answering the second research question, regarding the impact of the revised exam on instructional practices, teaching and testing.

Importantly, a sample of the new exam is analysed critically question by question. Particularly, the exam is studied in terms of content to see the extent to which its objectives, as envisaged by constructors, correlate with both the content of the test and of the textbooks (teaching curriculum) for grade 9 students. Thus, it was essential to explore the current textbooks in terms of content and objectives, as textbooks produced for teaching in schools are deemed a significant source of research evidence in educational research (Cohen *et al.* 2011).

Content analysis was chosen in this instance for exploring textbooks because it is a tool for gathering and analysing data from written materials. It was also used because it was very difficult to contact textbook authors due to the absence of contact information and time limitations. Also, it is highlighted that content analysis of any

materials used in addition to those textbooks as a result of exam changes would be a valuable source of documentary data. These processes contribute to answering the third research question, about the influence of the revised exam on the content of the curriculum. Thus, the researcher decided to use content analysis to pool data from the following documents: sample(s) of the revised exam, samples of the teachers' classroom tests, and the existing textbooks or other teaching materials teachers use in instruction, besides some archival documents, in order to elicit information on the technical quality of the test, test specifications, and links to curriculum objectives.

Further a number of other applied linguists were purposively selected and asked to look at samples of the revised exam, to add their critique. In order to elicit valuable information, bearing in mind the time limitations, some of the candidates were linguists who are closer to the educational system of the country and have direct or indirect contact with the related English curriculum in Libya schools (Some of those linguists are PhD holders, and others are PhD students at UK). Content analysis to the written reports, obtained from those persons, was of a valuable source of documentary data as far as the quality of the test and its impact on EFL education in the context are concerned.

### **3.3.3 Methods of Data Analysis**

Quantitative and qualitative data were analysed and interpreted by different means which are explained below.

### **3.3.3.1 Quantitative Data**

The questionnaire quantitative data was stored and coded, and then analysed using SPSS software, as this allows for a range of statistical techniques producing concise results. Frequency distributions were first calculated for all questionnaire items, considering Brown's (2001) advice on presenting results in percentage terms. Likert-style answers enabled quantitative data in the form of descriptive statistics, and so primarily to address the first research question. Thus, central tendency statistics such as mean and standard deviation were used to present an overall picture of the teachers' perceptions of the revised exam in relation to their instructional behaviour and use of teaching materials and curriculum. Additionally, comparing means and frequencies, and examining correlations, allowed making inferences from previous descriptive-data-based analyses. In this data I reported the mean. I considered whether or not the data was interval or ordinal. Whilst it is possible to make a case, the questionnaire data uses this item type is ordinal data. If the data is normally distributed, it does not make a difference whether reporting means or mediums, and all questionnaire data uses the mean as a matter of convenience and accuracy. For instance, in the standard data analysis text in this field, Brown (2001) recommends the use of the mean because that provides more power in the analysis of the data. Furthermore, it is a standard practice to do this in the applied linguistics literature, and it is very rarely questioned. For example see the study by Fulcher (2012) that reports means for similar questionnaire data as it is used in this study.

In addition, to explore the degree of effect (washback intensity) of main independent variables on each of the dependent variables, inferential statistics were used to

address particularly research question four. Whilst the one-way ANOVA was used to find out the difference among more than two groups followed by conducting post-hoc tests, the independent sample t-test was used to find out the difference between two groups. For this, the key aspects of the statistical output: the mean, standard deviation (SD), F and P values, and eta squared (effect size), were presented and interpreted. Differences in questionnaire findings were tested for statistical significance, and a probability of less than 0.05 regarded as statistically significant (for elaboration see, section 4.2.4).

### **3.3.3.2 Qualitative Data**

The data generated from the interviews and document study was dealt with qualitatively following a content analysis method of data analysis, as this was described as the best procedure for reducing qualitative data and for coding and analysing open-ended questions (Cohen *et al.*, 2007). Following Miles and Huberman (1994), three concurrent flows of activity were conducted to the interview data: data reduction, data display and conclusion drawing and verification. Qualitative data was compiled by identifying categories and themes that match the purpose of the interview(s) question(s), which were within the scope of the research questions, then transcribed (Silverman, 2010). For practical purposes, the researcher tried to order the interview questions in the order the research questions occur. Although they contributed to address all research questions, interview transcripts and documentary analysis were used for answering primarily research questions two and three.

All interview data were transcribed verbatim for each participant, and then presented in both versions, Arabic and English. “As the meanings have to be tested for their plausibility, their sturdiness, their ‘confirmability’ - that is their validity” (Miles and Huberman, 1994: 11), the interview transcripts were sent to participants to read and verify whether they are compatible with what they had said during the interviews. All participants confirmed the consistency between the transcripts and their original ideas. Also, a colleague at the school of education was asked to interpret some interview data in order to compare his interpretations with those of the researcher. A high degree of concordance was noticed between the two kinds of interpretations.

Data was transcribed separately for each identified category, relying on recordings and field-notes taken during interviews. I started colour coding the interview transcripts by categorising texts according to research aims and questions. This type of coding procedure is called selective coding (Charmaz, 2006). The transcription was read through and specific parts of the text were bolded or highlighted in different coloured fonts. Each colour represents a specific category to which the statements belong.

For example, for the teachers interviews, I used red for the answers of the question(s) relating to teachers’ beliefs and attitudes towards the exam introduced in relation to their classroom instruction, blue for the answers of the question(s) regarding the influence of this exam on teachers’ classroom teaching and testing practices, green for the answers of the question(s) addressing the effects of this exam on the teaching curriculum and materials teachers use in classrooms, and yellow for the answers of the question(s) tackling the influence of exam effects, negatively or positively, on

EFL teaching and learning in Libyan schools. These statements were later grouped under specific themes that arose as a result of the analysis. So, there were four main different groups of selected texts from the interview transcripts (see appendix 33). This procedure was followed for every transcript, teachers', inspectors' and the Examination Board director.

### **3.3.4 Research Phases and Timeline**

Regarding the research design and considering the time constraints of doctoral study, the empirical parts of this research were carried out in three phases (see figure 3.4).

Phase I was the preliminary stage, in which the researcher conducted his pilot study (see below). The timescale for this phase was October – November 2010. The primary aim of the pilot study was to test and refine the proposed research instruments (Marshall and Rossman, 2010). Three major research methods were used in this phase: teacher questionnaires, interviews with teachers and inspectors, and classroom observations. Thus, this phase addressed all initial research questions.

Phase II, the perception stage, lasted from February – March 2011. This phase aimed to elicit teachers' reactions to and perceptions of implementing the revised exam, and also to provide contextual information about teachers, enabling a general understanding of how teachers perceive themselves to be affected by the implementation of the revised exam in relation to classroom instructional practices. This phase dealt with the first research question. In this phase, a teacher survey questionnaire was the major research method.



Phase III, the implementation stage, looked at the influence of the revised exam after its implementation. It was scheduled to last from March – June 2011, overlapping with phase II. However, this phase was interrupted by the unrest and uprising in Libya in 2011 (see 3.7). The data collection procedure in this phase involved a sub-sample drawn from the survey conducted in phase II, addressing the remainder of the research questions. Interviews with teachers and inspectors and documentary study were conducted in order to probe in-depth data and to gain hard evidence of washback impacts from exam change. It was planned to conduct classroom observations at this stage, to reveal what happens inside the ‘blackbox’ (Long, 1980) as far as the introduction of the revised exam is concerned. But, unfortunately, due to constraints discussed in section 3.8, classroom observation was not possible.

Figure 3-4 Research stages and timeline

<b>Phase I</b>
Stage: Pilot study, the preliminary stage Timescale: October 2010 – November 2010 Methods: Teacher questionnaire, interviews with teacher and inspectors, classroom observation
<b>Phase II</b>
Stage: Perception stage (surveying teachers' perception of the new exam) Timescale: February 2011 – March 2011 Method: Attitudinal survey questionnaire administered to teachers to elicit their attitudes and perceptions of implementing the revised exam vis-a-vis various aspects of classroom teaching and learning.
<b>Phase III</b>
Stage: the implementation stage (case stud) Timescale: March 2011 – June 2011 Methods: <ul style="list-style-type: none"><li>• Follow-up interviews with teachers who participated in the questionnaire;</li><li>• Interviews with inspectors, and Exam Board director;</li><li>• Document study of textbooks, teachers' on-going tests and a sample of the new examination.</li></ul>

The three phases were conducted sequentially. Thus, the study was initially conceptualised in quantitative terms for investigation and then in qualitative terms for verification, reflecting the paradigmatic approach chosen earlier to fit the study. Different research methods were incorporated during the third phase rather than in the second phase. Findings were synthesized to provide answers to the research questions.

### **3.3.5 Pilot Study and Implications for the Main Study**

As Gorard (2001: 102) states, “all research designs need to be piloted or pre-tested”, the research methods discussed above have undergone trials. This section discusses piloting procedures, assessing the potential of these methods for the study, and the implications of piloting. The pilot study was conducted in October and November 2010. The data collection methods used in the pilot were the same as those used in the main study except the classroom observation which was utilised in the pilot study but not in the main study, due to the difficulties discussed in section 3.8.

#### **3.3.5.1 Participant Selection and Sampling Frame**

It was stated in Chapter One that the research site is Libyan schools, at the Basic Education Stage. However, due to travel costs and time constraints (Brown, 2001), the researcher decided to conduct the pilot in Libyan schools in the UK, with similar teachers to those intended for the main study. All participants in the pilot were members of the target population but not part of the final sample, keeping in mind De Vans’s (2002) advice: “the closer the match between the pilot sample and the final sample the better” (17). Further justification for piloting the research tools in the UK is that the researcher was principally trialling research instruments, and therefore the researcher’s aim was to test data collection tools in order to find out if teachers understand the instruments, and whether the researcher gets the kind of information envisaged from these instruments. It was hoped that piloting the instruments on this sample would not impact on the main study, as the main study would be conducted in Libya.

According to the cultural affairs department in the Libyan Embassy in London, there are 14 Libyan schools across UK in which Libyan students study in a manner similar to that applied in Libyan schools in Libya in terms of teaching and testing syllabi. These schools are run and supervised by the Libyan Authority of Education, through the cultural affairs department. This means that grade 9 students in these schools take the same exam as students in Libya. Furthermore, all the piloting teachers were Libyans most of whom have come from Libya to the UK to pursue Master's or doctoral degrees in education. Others were dependents on their spouses who also were Master or doctoral candidates at a UK University. Importantly, all participants have worked under the conditions of the revised exam (i.e. they have taught grade 9 classes at the Basic Education Stage schools in Libya). Ten teachers were asked to participate in the questionnaire piloting. They agreed, and questionnaires were emailed to them, seeking participation in the interviews and observations.

### **3.3.5.2 Piloting Teacher Questionnaires**

Piloting is essential for success and reliability (Cohen *et al.*, 2007). Thus, 'trailing' was the second method used to validate teacher questionnaires, along with the qualitative input discussed in section 3.3.2.2.1.

The primary aim of piloting teacher questionnaires (TQ) (appendix 5) was to ensure the construct validity of the questionnaire and reduce problems in administering it. The purpose was cross-checking –to see whether teachers understood each question as intended by the researcher, and whether the items were understood the same way by each teacher (Dörnyei and Taguchi, 2010). This allowed the surveyor to amend, refine or exclude some questions due to wording, relevance and need, besides

gaining valuable feedback on the questionnaire's readability. Furthermore, obtaining information regarding the time needed to answer questions was of particular interest. Hence, in addition to completing the questionnaire, participants were asked to give feedback on clarity, coherence, and language used (i.e. if there was any ambiguous wording, as they completed the questionnaire in their second language), and how long it took to complete.

Questionnaires were emailed to ten teachers, all of whom answered the questionnaires in full. However, some teachers articulated doubts about some terms and statements. For example, one said she did not know the meanings of "scepticism" (TQ13) and "mock exams" (TQ 20) (see appendix 5). Another teacher doubted the interpretation of the phrase "scheme of work" (TQ 25), while another said that item 5 TQ18 was unclear. Another teacher declared that TQ20 item 7 was ambiguous, particularly the word "consult". Accordingly, speculating this might be the case with the targeted teachers in the main study, all unclear terms or items were revised to avoid misinterpretations on the part of respondents.

Furthermore, although the researcher estimated that it would take twenty minutes to complete the questionnaire; most respondents felt it took longer: thirty minutes or more. Hence, the ancillary questions, not specifically related to the purpose of the inquiry, were eliminated, such as TQ 17 and TQ22. Also TQ31 and TQ33 were reworded in one question—TQ29 in the main study (see appendix 4). Other extraneous items and terms were excised such as the term 'mock exams' in TQ20 for the pilot study was replaced by 'previous exam papers' in TQ19 for the main study. TQ16 was reworded, besides; TQ28 and TQ29 were amended as they appear

respectively in TQ26 and TQ27 in teacher questionnaire for main study. The primary motive was to improve response rate and, to minimise the risk of getting uncompleted questionnaires in the main study. Thus, some clarifying modifications were made to the wording of some questions. Also, some room was added to some questions to encourage respondents to provide ideas the surveyor may not have considered. As a result, the 34 questions of teacher questionnaire for the pilot study were reduced to 30 questions in teacher questionnaire for the main study. Apart from these changes, the pilot study did not throw up any problems. Thus, a revised version of the TQ was prepared for use in the main study (see appendix 4).

### **3.3.5.3 Piloting Interviews**

This section discusses two pilot interview schemes, one conducted with two teachers and the other with one inspector.

#### **3.3.5.3.1 Teacher Interview**

Of the ten teachers who participated in the questionnaires, two volunteered to participate in the interview (referred to as teach A and teacher B). Upon the interviewees' request, interviews were conducted in Arabic, individually on a face-to-face basis, and were thirty minutes long each. Both interviews were audio-recorded (with teachers' permission), and supplemented by written field notes in order to trial the data collection procedure and equipment. Interviews were semi-structured, and interviewees were given the interview schedule beforehand – emailed one day before the interviews took place (Denscombe, 2003) – giving them time to voice their opinions about the exam and its impact on instruction, including teaching

materials. This made the interviews more successful in terms of eliciting the range of answers required.

Each interviewee was asked to give feedback on any problems they encountered when answering questions, to determine the extent to which questions were understood as intended, assessing what was and was not working. This procedure served as a means of testing question wordings and gathering information for setting further questions which had not previously occurred to the researcher. The interviewees articulated no particular problems in answering any of the questions. Hence, the interview list of questions was used for the main study with only minor wording corrections. One important implication of the piloting was that the two teachers' interviews indicated the interview data to be elicited from teachers in the main study would be rich enough to address the relevant research question(s). Briefly speaking, although the number of interviewees was small, the data revealed that to some extent the exam has an impact on both teachers' instructional behaviour. Thus, it appeared worth utilising teacher interviews as a data collection method in the main study.

Additionally, trialling the interviews helped identify variables and categories and their relationships (Gray, 2009), and allowed logical reordering of questions to facilitate the process of transcribing data for the analysis stage, as the interview piloting has indicated that the research would generate considerable data in the main study.

### **3.3.5.3.2 Inspector Interview**

As there were no Libyan inspectors of English available for an interview at the time in the UK, and due to logistical and financial reasons, the inspector interview was piloted on one Libyan inspector via a Yahoo messenger call. The interviewee was initially contacted by telephone to obtain agreement and to assign a scheduled time for the interview. The interview schedule was emailed to the interviewee beforehand to allow him to prepare.

Since the interviewee had little knowledge about the research topic and the issue investigated, it was useful to follow Gray's (2009) advice: "explain the purpose of the interview, who the information is for, how the information is going to be collected ... and how it will be used" (379). As with the teacher interviews, the inspector was asked about the clarity of questions and whether he had difficulty answering any of them. The interviewee declared that all questions were clear. Hence, the version of the inspector interview used in piloting was employed verbatim in the main study, except for the reordering of some questions.

Nevertheless, two issues raised during the inspector interview may have implications for data collection in the main study. First, regarding Q6, the inspector answered: "he has never written such an exam". He was then asked what strategy he would use, or whether he would be obliged to follow a certain procedure. Interestingly, his answer was: "If I was chosen to write such exam, I will write it as they want me to do it," by which he meant the Examination Board, the Authority of Education.

Secondly, when the inspector was asked about the main reasons for the Authority of Education to change the exam syllabus (i.e. Q2 and Q3), he answered: "I do not have



a direct and clear answer to this question”. The interviewer was unsure whether it was only this inspector who was unaware of the answer (or he was eschewing answering the question in favour of responsibility), or whether this could be the case for all inspectors asked in the main study. So, the researcher believed that for the purpose of the main study – and in order to gain specific answers to these two important questions and to hopefully gain information related to the implementation of the exam and its specifications – it would be useful to involve more participants, and also the Examination Board director. Thus, piloting teacher and inspector interviews had a positive impact on the implementation of interviews during the main study.

#### **3.3.5.4 Piloting Classroom Observation**

To recapitulate, classroom observation was used in the pilot study but not in the main study because of the constraints and difficulties encountered by the researcher. The following sub-sections discuss the proposed plan to implement classroom observation in the main and pilot study of this research.

##### **3.3.5.4.1 Rationale**

Patton (1990: 25) suggests that “to understand fully the complexities of many situations ... observation of the phenomenon of interest may be the best research method”. Scholars like Alderson and Wall (1993), Cheng (2005) and Watanabe (2004) have highlighted the importance of employing classroom observation in washback research. For instance, Alderson and Wall (1993) emphasised that classroom observation is conducted “to see whether what teachers [...] say they do is reflected in their behaviour” (127). Thus, in order to inquire into unclear issues

arising during the interviews and to confirm the accuracy of questionnaire and interview data, the present researcher decided to opt for classroom observation as another method of data collection.

Data based on observational evidence, which triangulates with other data pooled via questionnaires and interviews, will unveil what happens in classroom teaching, and hence can deepen our understanding of the effect of this exam on aspects of teaching and learning, providing useful insights about the washback phenomenon in general.

#### **3.3.5.4.2 Procedure**

For the purpose of this study, non-participant observation was planned, in that the researcher was supposed to be a “complete observer, one who merely stands back and 'eavesdrops' on the classroom proceedings” (Taylor and Bogdan 1984: 15). Considering the researcher’s role in the observation as an outsider observer, i.e. not to interrupt participants during the observation sessions, and as implicated in the pilot study (see below), follow-up contact would also have been made with teachers and inspectors, either in person or by phone during or after data collection. The purpose is to invite them to clarify ambiguous points, and this was what happened during the piloting of classroom observation.

It was intended to observe four teachers of grade 9 classes for an extended period. Those teachers teach other grades in the same school (i.e. grades 8, 7 or 6) to which the revised exam is not introduced. The current researcher intended to adopt Wall and Horak’s (2006) classroom observation strategy used in their longitudinal study of the impact of changes in the TOEFL test on teaching and learning in Western Europe. The researchers observed the same teachers teaching TOEFL and non-

TOEFL classes. Likewise, observation sessions were planned to be undertaken with the same teachers teaching different classes in the same schools; however, this strategy was deployed in the pilot study, to include a comparative element as this is a common feature in washback studies (e.g. Wall and Alderson, 1993; Alderson and Hamp-Lyons, 1996; Hayes and Read, 2004; Shohamy *et al.*, 1996). The reason for this decision was two-fold: “The first was to see whether there were any similarities or differences in the two classes. The second was to try to tease out the teacher variable from the test impact variable” (Wall and Horak, 2006: 28).

Keeping in mind the objectives of the revised exam outlined in see section 1.3.2.3 and the objectives of the current English curriculum (see section 1.3.1.2), the purpose of this observation was to compare teacher practices in both cases. In this case, the purpose of the comparison was methodological: to identify matches and mismatches between the two practices (Hayes and Read, 2004). Should differences be found, this could be analysed in terms of washback associated with the revised exam (Burrows, 2004; Qi, 2004; Wall and Horak, 2006).

#### **3.3.5.4.3 Instruments**

“Observations in natural settings can be rendered as descriptions either through open-ended narrative or through the use of published checklists or field guides” (Rossman and Rallis, 1998: 117 in Angrosino, 2008: 162). Thus, in order to make teachers’ behaviour in class more detectable, it was intended for this research to utilise a classroom observation scheme, as used in the pilot study (see appendix 8). The scheme is adopted from the Communicative Orientation of Language Teaching Observation Scheme (COLT) (Spada and Frohlich 1995), because it “describes

classroom instruction in terms of the types of activities that take place” (Froehlich *et al.*, 1985: 29), and it has been usefully utilised in some washback research studies (e.g. Hayes and Read, 2004; Burrows, 2004; Wall and Horak, 2006; Hsu, 2009). Part A seems particularly appropriate as it provides a useful framework in which teachers’ classroom practices in relation to the revised exam can be recorded.

Additionally, as other activities may be observed which are not specifically covered by the COLT scheme, such as exam-related activities, test-taking strategies or giving students information about the exam, the researcher has also designed an observation note-taking sheet to record such categories (see appendix 9). Essentially, all observation sessions were intended to be taped for verification purposes at the data analysis stage. The observation scheme as well as the note-taking sheet were planned (and executed in the pilot) to be revised immediately after each observation session, to ensure that all relevant data would be coded for analysis (Teddlie and Tashakkori, 2009).

Like the teacher questionnaires and interviews, instruments used for classroom observation had undergone validation procedures via piloting.

#### **3.3.5.4.4 Classroom Observation in the Pilot Study**

The main purpose of piloting classroom observation instruments was that they help the researcher test these instruments in order to determine their usefulness for the main study, to envisage the kind of data to be collected via this tool, and to elicit data that teaches the researcher more about the effect of the revised exam on classroom instruction and curriculum.

During the interview sessions, the two interviewed teachers (teacher A and teacher B) agreed to participate in classroom observation. Teacher A was observed for eight sessions of 40 minutes each: five teaching grade 9 classes, and three teaching grade 7 classes, whereas teacher B was observed for a shorter period (only five sessions: three teaching grade 9, and two teaching grade 6 classes) as she withdrew from being observed due to personal circumstances. The purpose of observing different classes was to determine any matches or mismatches between the teacher's practices in those different contexts. As mentioned above, the COLT observation scheme and the observation note-taking sheet were used to detect and record teachers' classroom instructional behaviour. Additionally, upon participant agreement, the observation sessions were tape-recorded for retrieval of data and verification procedures.

Although there were only two participants, with few sessions, the researcher observed some impacts on teaching and learning in the classroom, especially observations conducted in grade 9 classes. For example, the researcher noticed that the two teachers focused on exam-related activities, occasionally leaving aside the textbook designed for teaching, giving students extra grammatical exercises and asking them general information questions (sometimes in a written format) that reflected the reading passages in the textbook on which some items of the exam studied are based (The researcher was well aware of the content of the teaching syllabus prescribed in Libyan schools as he had taught this syllabus when working in Libyan schools from 1995 – 2006). Thus, the behaviour observed indicates that the exam has an influence on some aspects of teaching and learning, specifically when the exam is soon, and the impact observed could be limited to classes in which the revised exam is administered as a school leaving exam. However, this impact cannot

be taken for granted because of the small sample and the short time during which the piloting was conducted (Alderson and Wall, 1993). In contrast, during the main study, if classroom observation was applied, this behaviour could have been interpreted as a direct effect of the revised exam. Notwithstanding, as it was thought that the data elicited from piloting classroom observations was quite substantial, such data was analysed and utilised along with the data collected for the main study, presented in the following chapter.

To conclude, conducting classroom observations has given useful insights and informed the implementation of the study's main observation sessions. For instance, after one observed session, the teacher was asked questions probing issues noticed during the observation. This procedure indicated the need for conducting follow-up post-observation interviews in the main study, or as a recommendation for prospective washback studies in the context of this study.

### **3.4 Validity and Reliability Issues**

As the study is characterised by a combination of positivist and interpretivist perspectives, credibility/validity refers to the extent to which the findings of the study and the reality are congruent (Merriam, 2002), and dependability/reliability refers to the replicability of findings. Thus, this research, particularly in Phase Two (i.e. quantitative questionnaire), attempted to make generalisations to the whole population and improve validity through careful sampling and appropriate methods of data collection and analysis. To interpretivists, "the understanding of reality is really the researcher's interpretation of participants' interpretation or understanding of the phenomenon of interest" (Merriam, 2002:25). This interpretation will be

evolved through the qualitative data obtained via interviews with teachers and inspectors, and content analysis of some relevant documents, in phase 3.

As the notion of triangulation bridges issues of reliability and validity (Cohen *et al.*, 2011: 196), the triangulation and mixed methods research design adopted for this study, driven by the aim of the study and research questions, strengthened the internal validity/credibility (Cohen and Manion, 1985). The researcher gathered data from participants through their words and actions, both of which were recorded, and drew on his own experience and knowledge gained from the literature. To enhance credibility, the methods chosen for data collection have undergone piloting, as discussed in the previous sections. Moreover, transcripts were typed and given back to participants for review, and participants signed consent forms prior to participation. Also, some interview data was generated in the participants' first language, Arabic; the translated transcriptions were checked by a qualified colleague. This enabled a thorough interpretation of data and enhanced credibility, boosting the trustworthiness of the research.

### **3.5 Ethical Considerations**

As educational research usually involves using human beings as subjects and institutions as places to conduct research, it was important for this research to protect and respect individuals (Cohen *et al* 2007) as well as sites.

Research ethics concerns, protecting persons involved in research from psychological, emotional and physical harm (Cohen *et al.*, 2011; Silverman, 2010). BERA (2011) underscore that educational research should be undertaken with respect for all people involved. To ensure the ethical integrity of this research, and

taking into account BERA's (2011) ethical guidelines for educational research, the following steps were taken.

It was important to gain informed consent (Cohen, *et al.*, 2007). Permission was sought from the Libyan authority of education to access teachers in their schools and inspectors in their offices. Also, participants in interviews – teachers and inspectors – signed consent forms (see appendices 18 and 19) before taking part. The consent forms, as highlighted by De Vaus (2002), included the researcher's and his sponsor's names, information on the general purpose of the study, the participant's right to decline at any time, and information about the confidentiality and anonymity of data, to encourage frankness (Robson, 2002).

Obtrusiveness had also to be considered. As participants were all working professionals with busy lives, and case study research requires intensive engagement, the researcher endeavoured to be cautious about the possible impact (e.g. unnecessary distress) of participation on their lives, and in relation to the kind of data they would provide. The researcher therefore scheduled all interview sessions in advance, for times deemed convenient by participants. Also, permission to record conversations was sought from interviewees before the interviews took place.

Thirdly, as the researcher was intensively studying participants' perceptions and behaviour, confidentiality was deemed paramount (Cohen *et al.*, 2007). All participants were informed that the data they provided would be kept strictly confidential: neither revealed to anyone, nor shared with other participants. Procedures used to assure confidentiality were specified on teacher questionnaires adopted for the second phase of the study, and in consent forms used for phase III.



Finally, the participants' 'anonymity' was also respected. This was maintained by concealing participants' identities in written reports (Erickson, 1985) and teacher questionnaires. All questionnaire respondents were guaranteed anonymity, except for 11 teachers who were asked to indicate their approval to be interviewed; besides, questionnaires contained codes linked to teachers' names to enable the researcher to follow up those teachers in their schools to conduct interviews (Brannen, 2005). Pseudonyms were adopted in transcriptions from interviews with teachers, inspectors and the Exam Board director.

Based on the above steps and considerations, a research ethics approval review form was submitted to the administration office of the School of Education, University of Leicester. As a result, ethical approval to conduct all phases of the study was granted by the Research Ethics Committee at the University (see appendix 10).

### **3.6 Accessibility of Research Site and Participants**

Accounting for accessibility was an influential factor for selecting the site of this study. This decision of selection enabled the researcher to access schools, teachers and inspectors with reduced effort and cost. What made the data collection procedures successful, despite challenges and difficulties encountered because of the Libyan conflict at the time of some stages of the fieldwork, is that Libyan society is bound by strong social ties, in that most people in a region cooperate as they are known to each other. For example, in the context of this study, as the researcher is one of this community, it was not difficult for him to access teachers and inspectors, asking for their participation, making the research site more accessible. Also, the kindness of the research participants, and their empathy with the researcher's need to

successfully complete the data collection despite the challenges, contributed to increasing the response rates of questionnaires and interviews.

Furthermore, as the current researcher worked as an EFL teacher in basic education and secondary schools in the chosen city for more than 12 years and as a member of university teaching staff for 3 years, it was expected that, at the time of data collection, many of his colleagues and friends would be available to provide support. Fortunately, some former colleagues were working as head-teachers in these schools and some as inspectors of English in the region. It was also possible to meet some teachers taught by the researcher during his work in schools and as a lecturer in the department of English in the college of education at the University of Misrata (some of those students have since become teachers teaching grade 9 classes in the targeted schools). Nonetheless, formal procedures had to be followed to legalise this process. Hence, the current research commenced the process in the UK where ethical approval to conduct the empirical part (i.e. the thesis stage) of the study was granted by the Research Ethics Committee at the University of Leicester. Official letters were also gained from the authority of education in the region where the study was conducted to access schools and meet teachers, inspectors and the Examination Board Director.

Having discussed the ways through which the research site and participants were accessed, the discussion now explores how the research methods adopted to collect the data to address the research questions were used, and what challenges were encountered. But, as the process of the data collection, especially in phase three of this study, faced obstacles which caused interruptions to the fieldwork, the researcher

sees it as essential to acquaint the reader briefly with a background about the situation in the context of the study. The section that follows will describe the data collection procedures followed, and the challenges and difficulties encountered as a consequence of the conflict.

### **3.7 Turmoil and Unrest in the Context of the Study**

The present study was carried out in Libya, Misrata from January to June, 2011. The data collection procedure, mainly in phase three, was affected by the security situation in Libya as a consequence of the conflict that took place in most parts of the country, February 2011 Revolution. The following paragraphs, in addition to the next section, acquaint the reader with what was happening during the crisis and its repercussions on the study—the researcher was a witness to what went on.

The new generation in the Arab World is seen to be tired of ruling regimes which they did not elect. People who are in power in most of these countries have ruled for more than 20 years, in many cases exhausting these countries' economic resources and threatening the stability of the lives of many people who wish to live in freedom and dignity. Their philosophy, their way of thinking, their understanding and their search for identity has recently emerged, and this is called the Arab Spring Revolution. This began with rallies, peaceful demonstrations and protests across the Arab World, e.g. in Tunisia, Egypt, Libya, Yamen, Bahrain and Syria, against the authoritarian regimes, asking for reforms and promoting identity, dignity, democracy and freedom.

In Libya, one of the main motives for revolt against the ruling regime was the austerity people have suffered for a long time, lacking good health care, good

education and good infrastructure, even though Libya is the world's 15th largest exporter of crude oil, accounting for 7 percent of global daily output, and has the largest oil reserves in Africa (OPEC, 2011: 17). Consequently, demonstrations against the regime of Muammer Al-Gaddafi, who ruled the country for more than 41 years and has been described as a long-time dictator, began in February 17th in the eastern city of Benghazi, the second largest city after the capital Tripoli. Unfortunately, the rebellion in Libya did not go as smoothly as those in the neighbouring countries, Tunisia and Egypt, where the rulers of both countries were toppled with fewer casualties.

The regime's security forces maltreated the demonstrators, resulted in many people being killed and wounded. This triggered protests across the country against Al-Gaddafi, who was then seen by the majority of Libyans as a tyrant. Mounting protests spread to the overwhelming majority of Libyan towns and cities, including the capital Tripoli, and Misrata, the third biggest city. These demonstrations were, again, brutally crushed by security forces and militias loyal to Muammar Al-Gaddafi. The situation in those cities, especially Misrata (the present author's home town and where the data for this study was collected) escalated dramatically and tragically. The regime troops invaded and besieged most parts of Misrata and Zawyia, killing and injuring hundreds of civilians. As a consequence of the crackdown on anti-government protestors, civilians backed by some defecting army units started to defend themselves, as in the case of the city of Misrata. They were forced to carry weapons and fight the regime forces and mercenaries, though most had no experience using weapons and fighting.

Al-Gaddafi's excessive violence not only provoked outrage across the country, but also across the world. The international community, including the League of Arab States and the UN, condemned the harsh methods used by the regime forces. Many world leaders from countries such as France, UK, USA and Italy, in addition to the Security Council of the UN, announced that Muammer Al-Gaddafi had lost his legitimacy to rule Libya, calling on him to step down and make room for the National Transitional Council NTC (NTC is an interim government which had already been established by the opposition in Benghazi). However, Al-Gaddafi turned a blind eye to all international calls to stop the violence and establish constitutional reforms or to step down. He, further, vowed to fight and smash his opposition, ordering his loyalists and security militias to take to the streets against the protesters. As a result, and based on the United Nations Security Council, the resolutions 1970 and 1973, a no-fly zone (then followed by a no-drive zone) was imposed over Libya as a way of ensuring the safety of the civilian population (The Guardian, 2011).

To protect civilians and civilian population areas, on March 19 a military operation was launched. The operation was initiated by French and British airstrikes, preventing huge military convoys of army brigades of Al-Gaddafi forces from committing massacres in Benghazi and the nearby towns and cities. The operations were thereafter conducted by NATO, in addition to Qatar and UAE. The US, meanwhile, announced it would be playing a supporting role.

Despite NATO military operations being in process, pro-Al-Gaddafi troops supported by mercenaries were still threatening civilians in the streets of besieged cities

and towns (e.g. Misrata, Zawyia and Zintan), using civilians as a human shields, making it difficult for the NATO airstrikes to reach them. On-going heavy fighting on the ground between forces of the Libyan NTC – revolutionary fighters – and those loyal to the regime continued to impact the daily lives of Libyan civilians, especially in the city of Misrata. The people of Misrata suffered terribly from Al-Gaddafi's forces for more than fifty days. Bombarding and shelling the city with artillery, Grad rockets and tanks, caused many casualties among civilians and revolutionary fighters (according to medical reports, on 15 May 2011, within the first two months of the crackdown, more than 500 were killed and 2000 wounded), destroying many houses and buildings and leading to the evacuation of thousands of families. The intense fighting continued in the city streets from March to early May, when it was announced that the city was freed after defeating Gaddafi's forces, which had retreated after NATO-led air strikes, on the outskirts of the city, about 30 miles towards the east, south and west, making the city life getting back to normal (the 'liberation war', as described by many Libyans, finally ended with the capture and death of Ghaddafi and some of his sons and loyalists in October 2011). Based on these positive developments, by mid-May the researcher started to think about implementing the data collection procedures of the third phase of the study, i.e. interviews. The following section tackles these issues in detail.

### **3.8 Data Collection Procedures, Challenges and Difficulties**

Following the discussion in section 3.6, this section explicates the procedures the researcher followed in collecting research data, and the challenges and difficulties encountered.

In January 2011, the researcher travelled to Misrata to conduct fieldwork. Immediately after gaining official permission to access research participants, the distribution of teacher questionnaires started on the first of February, 2011, when teachers in Libya were just returning to schools for the second term of the school year. This time was ideal for enhancing teachers' participation as they were less busy, in that their tasks of marking students' work or examination sheets were not yet due. It is believed that this good timing was one of the influential factors that made the response rate of questionnaires high, supporting the productivity of appropriate timing as emphasised by Cohen *et al.* (2011).

Based on the sampling frame and participants' selection discussed in section 3.3.2.1, 140 questionnaires were planned to be delivered to 140 teachers teaching in about 140 schools. Due to time constraints, access to the all teachers/schools was quite difficult as the schools were scattered over a wide area, and the city's parts and institutions were not linked by good postal services (as a developing country, this is the case throughout Libya). Teachers did not generally have access to the internet, so the researcher could not email the questionnaires to them. The researcher therefore had to look for a solution.

As they usually visit teachers in their schools, and due to the good relationship the current researcher has with some of them as discussed in section 3.6, English inspectors were asked for assistance in delivering questionnaires to teachers. Fortunately, they agreed to help, and hence, they succeeded in distributing and collecting many teacher questionnaires, while the rest were distributed by the researcher. Nevertheless, the process of distributing the questionnaires to teachers

was very demanding as many schools/teachers needed to be visited at least twice, one when delivering the questionnaire and the other when collecting it. This took about three weeks, moving from and between schools distributing and collecting questionnaires.

The procedure commenced by meeting the school head-teacher first to assign an appointment with the teacher according to his/her availability based on the teachers' class timetables, though many teachers were met and handed the questionnaire at the first visit. Telephone calls were also made with school head-teachers and teachers themselves to confirm these appointments. Each questionnaire was delivered in person, in order to give teachers a general idea about the questionnaire and the study aims, and to tell each teacher what exactly was needed from them, and to boost the response rates (Cohen *et al.*, 2011). Another appointment was made with each teacher to collect the questionnaires. When collecting the questionnaires, some teachers raised issues that were not clear to them in the questionnaires; immediately these issues were sorted out and, thus, the questionnaires were fully completed by the teachers, and then collected. This played a substantial role in raising the response rates. However, a few questionnaires were not returned. Those teachers who had no doubts about the questionnaires left them with head-teachers sealed in the envelope provided for collection.

Thus, 140 questionnaires were distributed to 140 teachers covering all selected schools. 118 were returned to the researcher. Because of the uprising and unrest in Libya, specifically in Misrata, which escalated rapidly and tragically in March and April 2011, it was not possible for the researcher to collect more questionnaires as



schools were closed at the time and these schools were the easiest, if not the only, channel through which the researcher could access teachers. Of the returned questionnaires, 18 were excluded from the data analysis process as preliminary data screening showed some items had been left blank. Altogether, 100 questionnaires were completed with all questions and items answered in full. So the number of teachers who comprised the sample of the survey questionnaire (100 teachers) represents 71.4% of the target population, i.e. the return rate was about 71.4%. By the end of this procedure, the researcher had completed the data collection for the second phase of the study, the perception stage.

The next step was to conduct interviews with teachers sampled from the questionnaire respondents (see section 3.3.2.1), constituting an important part of the third phase of the study, the implementation stage. Many teachers, as indicated in the returned questionnaires, were willing to participate in the interviews and in the observations. Of them, 11 were selected for interviews according to availability and accessibility. Unfortunately, the data collection stage at this time was interrupted by the crisis in Libya discussed in the previous section. Consequently, data collection was suspended from March to May, 2011. During that time, the researcher, as an eyewitness to the situation worried that the rest of the study would no longer be achievable. However, after freeing the city in May, the researcher started preparing to conduct interviews with teachers and inspectors in order to pursue the other part of the data collection. For the reasons discussed above in section 3.6, the researcher came to believe that proceeding with data collection at this stage was feasible.

Although by mid-May it was announced that the battle for Misrata was over, and city life was returning to normal, conducting such procedures remained demanding both on the part of the researcher and of participants. So, participants' feelings and availability had to be considered even more. Since schools were still closed for security purposes, interviews with teachers occurred out of schools. Most interviews were carried out in teachers' houses. Accessing teachers was quite problematic as the researcher does not know most of the teachers' residential areas, though their schools' names and telephone numbers were stated in the returned questionnaires. It was not possible for the researcher to contact teachers via telephone as all telephone networks were cut by the loyalist forces during the crackdown. Nor were their names or places known to the researcher, as teacher anonymity was considered when designing questionnaires. As they know teachers' names, schools and places of residence, the inspectors' assistance was sought to access teachers for interviews. Two inspectors were met and gave information that helped access teachers. Also, the two inspectors contacted some teachers whom they know, seeking their participation in the interviews. However, it is important to note here that the purposive sampling of participants adopted to conduct interviews was borne in mind. That is, selecting the well-qualified and experienced teachers "to provide rich and varied insights into the phenomenon under investigation" (Dörnyei, 2007: 126).

The first step was to contact teachers, seeking their participation in the interviews. Thirteen teachers were contacted. 11 teachers agreed to take part in the interviews, whilst two refused to participate due to personal circumstances (one said that he was not psychologically in a situation to be interviewed). The next step was to meet the eleven teachers individually, giving each the interview schedule and the informed

consent forms to sign, and agreeing an appointment for the interview. The researcher scheduled the interviews dates and times in his notebook for reference. As teachers were off school, conducting interviews was not restricted to a specific limited period of time in a school day. Interviews were carried out within ten days, approximately an interview per day, from 15 – 24 May, 2011. As the interviewees were themselves the questionnaire respondents, before conducting the interviews each interviewer's questionnaire was scrutinised to check the responses to ascertain whether some areas required probing in the interviews, and to elicit extended information about the effect of the revised exam on teachers' instructional practices and curriculum.

Regarding inspectors interviews, it was less problematic to access inspectors for two reasons: the first is that there are fewer of them, compared to the interviewed teachers; secondly, the researcher, as mentioned above, has a good relationship with some inspectors, which enabled him to contact inspectors more easily. Nonetheless, as a consequence of the bad situation due to the Libyan conflict, accessing inspectors in their offices was not possible as the inspectorate office was closed during this stage of data collection. So, the procedure used to contact inspectors was the same as that used with teachers: contacting the inspector first in their house, giving them the interview schedule before the interview took place and asking for their signatures on the consent forms, then assigning an appointment for the interview according to their availability. In addition to the Examination Board director, seven inspectors were interviewed. It took one week for the researcher to conduct all these interviews, from 27 May – 2 June, 2011.

Another problem facing the researcher during the third phase of this study was the feasibility/possibility to implement the classroom observation as a method of data collection as planned in the early stages of the research design (see section 3.3.5.4). Although it was applied in the first phase of this study, i.e. pilot study and yielded valuable data, it was not possible for the researcher to carry out classroom observation in the main study because schools were closed as a consequence of the crisis. It was preferred and important, but not a pre-requisite, for the study to use classroom observation, though many doctoral studies related to exam influence on aspects of teaching and learning have been successfully conducted without using classroom observation as a method of data collection. Some studies have utilised only questionnaires (e.g. Andrews, 1995; Lam, 1995; Cheng, 1998; Stecher *et al.*, 2004; Choi, 2008; Amengual-Pizarro, 2009), others have used only interviews (e.g. Qi, 2004), some have deployed questionnaires and interviews in tandem (like the current study), in addition to documentary analysis (e.g. Shohamy *et al.*, 1996; Ferman, 2004; El-Ebyary, 2009), some have used interviews and observations (e.g. Wall and Alderson, 1993; Saif, 2006), and others have utilised a combination of questionnaires, interviews and classroom observations (e.g. Shohamy, 1993; Cheng, 1997, 1999; Watanabe, 1996, 2000; Burrows, 2004; Wall, 2005; Munoz-Alvarez, 2010) (see appendix9).

So, as a consequence of the difficulties and challenges encountered during the third phase of the study, the researcher was forced to cancel the deployment of classroom observation as a method of data collection, an issue which can be considered a limitation of this study.

The researcher believes that the data collection procedures followed and the difficulties and challenges encountered in this study during the fieldwork of data collection may have important implications for other prospective research studies, specifically in terms of boosting the response rates of participants. That is, the good timing and the purposive selection of the context where it is possible for the researcher to depend on people they know in administering questionnaires and encouraging participants to take part in interviews can significantly increase the response rates of their participants. Indeed, the researcher's good relationship with many stakeholders in the context of the study was important, specifically where the research or a part of the research data collection stage faced obstacles or was interrupted due to political or military unrest or conflict. This may empirically/practically confirm the supposition of Cohen *et al.* (2011: 264) that the researcher's status and prestige, and the respondents' personal empathy with the researcher can be important factors for persuading participants to return questionnaires, and to volunteer to participate in other parts of data collection.

### **3.9 Summary of the Chapter**

In conclusion, this chapter has explicated the research methodology and design adopted for this study to address the research inquiry. To fit the purpose, the researcher has opted for a mixed research paradigm, combining positivism and interpretivism, using a combination of quantitative and qualitative approaches. From the quantitative paradigm, the researcher has selected a survey strategy, and from the qualitative a case study strategy. As for the research tools to pool the necessary data, the researcher has employed teacher questionnaires for the former and teacher as well as inspector interviews and documentary analysis for the latter. It was planned

to conduct classroom observation in the main study, as implemented in the pilot study, but this was not viable due to the conflict at the final stage of the data collection procedures. Nevertheless, the results of classroom observation data pilot was used and presented along with the data collected via questionnaires, interviews and documents. It was hoped that the findings of the questionnaires, interviews, documentary analysis and the pilot observations will be sufficient to triangulate one another and answer the research questions, which looked at the impact of introducing a newly revised high-stakes EFL public examination on teachers' instructional practices, materials and curriculum. A detailed report and discussions on the findings will be reported in the following chapters.

## Chapter 4 Presentation of Results

---

### **4.1 Introduction**

As discussed in the foregoing chapter, the empirical part of the research was carried out in three phases. Whilst Phase I concerned the pilot study, Phases II and III comprised the data collection of the main study lasted from February to June 2011. Questionnaires, interviews and observations, complemented by document study, were utilised to collect research data in order to appropriately address the research questions.

### **4.2 Results and Analysis**

As the research data was gathered by means of quantitative and qualitative methods, quantitative data will triangulate with qualitative data, both reported in this section based on specific themes stemmed from the research aims and

questions. This section, with reference to the exam and curriculum review reported in Chapter One sections 1.3.1.2 and 1.3.2.3, reports document study which analyses the new exam and reviews the classroom on-going tests participant teachers wrote for their students. The main themes identified and reported in this chapter include:

- Teachers' perceptions of and reaction to the new exam;
- Washback on teachers' classroom instructional behaviour;
- Washback on teaching materials and curriculum;
- The degree of washback effect vis-à-vis teacher and context factors; and

#### **4.2.1 Teachers' Perceptions of and Reaction to the New Exam**

It was hypothesised that teachers' perceptions of the exam would impact on their reaction as far as their teaching practises are concerned. So, the first area studied was to explore teachers' perceptions of and reactions to the new exam, addressing the first research question. In this section, three sub-themes relating to teachers' perceptions are categorised, namely: a) perceptions of the reasons behind implementing the new exam; b) attitudes towards the new exam and its quality; and c) perceptions of the new exam in relation to their classroom practices.

##### **4.2.1.1 Teachers' Perceptions of the Reasons behind Implementing the New Exam**

In Chapter One (1.3.2.3) the aims and intents of introducing the new exam (BECE) were outlined. TQ 11 was devised in accordance with these aims in order



to elicit teachers' perceptions and opinions of the change. Table 4.1 shows the results of this, giving the mean scores on a Likert scale of agreement. Teachers were given a space to allow them to add other underlying aim(s) behind the new policy which they thought had not been announced by the exam constructors.

It can be seen that teachers, on average, disagreed that the revised exam was implemented 'to provide students with a gauge of their language learning achievement regarding the English curriculum' or 'to motivate students to use integrated skills' or 'to cover all the components of the curriculum' and/or 'to prepare students for future education'. However, they generally perceived the purpose of the test as 'to score the answer sheets mechanically', 'to eliminate the risk of cheating' and 'to disseminate the results quickly, adequately and as transparently as possible' as the main reasons behind implementing the new exam.

Table 4-1 Teachers' perceptions of the reasons for implementing the new exam.

<b>Q11. What do you see as the major reasons for the education authority to change the old English exam by administering a new revised one?</b>								
<b>Variables</b>	<b>1 str. Disagree</b>	<b>2 disagree</b>	<b>3 Undecided</b>	<b>4 agree</b>	<b>5 str. Agree</b>	<b>Mean</b>	<b>SD</b>	<b>Rank</b>
To score answer sheets mechanically	-	-	-	53%	47%	4.47	.501	1
To eliminate the risk of cheating	-	-	-	55%	45%	4.45	.500	2
To disseminate the results quickly and transparently	-	-	-	73%	27%	4.27	.446	3
To prepare students for their future education	-	66%	5%	28%	1%	2.64	.926	4
To cover all the components and content of English curriculum	32%	47%	-	20%	1%	2.11	1.10	5
To motivate student to use integrated skills	33%	67%	-	-	-	1.67	.472	6
To provide students with a gauge of their achievement regarding the curriculum	61%	38%	-	1%	-	1.41	.552	7

Questionnaire findings in this category revealed, to some extent, that there is a discrepancy between what teachers perceived to be the reasons and intentions behind the imposition of the exam, and the intended washback of the exam as envisaged by its constructors. This indicates that teachers are likely to react to the exam according to their educational beliefs, not according to exam constructors' beliefs and intentions, which raises the possibility of unintended consequences from the exam.

The above finding was reflected by those elicited from interviews conducted with a sample of teachers and inspectors in the third phase of the study. For example, one teacher said "one of the advantages of this exam is that it considerably reduced the risk of cheating. Also, the exam papers are marked mechanically, helping disseminate the results quickly and transparently" (T2, Abdullah). This finding was verified by some of the interviewed inspectors, as one said:

... the main aim of imposing the new exams, which many welcomed, is to overcome the phenomenon of cheating. By the new exam, the risk of cheating, which was common activity with the old exams, is considerably reduced. The exam has only one marker, i.e. the computer, which increased the credibility and reliability of marking and in disseminating the results ... (Insp4, Assadi).

Most of the interviewed teachers stated that the exam did not motivate students to use integrated language skills which indicates this kind of activity is ignored in the new exam. One teacher said: "one disadvantage of the new exam is that it does not consider students' upcoming years of education. Also, it discourages the use of integrated skills in class and not all the syllabus is covered" (T8, Alghoul).

As they noted in the space provided in TQ11, it was found that some surveyed teachers perceived 'economic and financial purposes' as other reasons behind

implementing the exam which were not announced by the exam policy makers. During teachers' interviews, when this issue was brought up, most of the teachers declared, upon which inspectors agreed, that one further aim of imposing the new policy was to reduce costs. The following is a representative quote:

In the previous years and under the conditions of the old examination, all students' exam papers were corrected manually by many teachers. Also, there were committees of examinations, consisted of teachers and inspectors, responsible for preparing and disseminating the results. To do this job a big number of teachers and employees were recruited by the authority of education. Huge amounts of money were paid for this work across the country, which placed a large financial burden on the authority of education budget. Therefore, the new exam, I think, was imposed accordingly, as all previously manual work is being done now by computers (T9, Sultan).

Thus, it is to suggest that the new exam policy, further to its purpose as a disciplinary tool, serves economic purposes, to switch to multiple-choice tests and eliminate essay tests to cut costs. Interestingly, this was supported by the examination board director when he was opined about the claim that the new exam has been imposed to drive down costs. He agreed, stating: "yes, it placed less financial burden on the budget of the authority of education as all work is being done now by computers recruiting far less people compared with the old exam policy" (EBD).

This particular finding supports the claim that high-stakes standardised tests can be manipulated to hold down testing costs, to save time and money, albeit their troubling consequences (for more illustration on how state-wide tests are deployed to save money see Toch, 2006: 14-16; Haggie, 2008: 4).

To explore the extent to which teachers perceived the change(s) between the new exam and the previous one, TQ12 was designed. This question was also intended

to elicit teachers' opinions of the new exam. The results are shown in Table 4.2. It can be seen that the overwhelming majority of teachers regarded 'more emphasis on reading texts stated in students' textbooks' and 'more emphasis on grammatical usage' as two main issues characterising the new exam, whereas all teachers agreed that the new exam was neither more closely related to the principles of the English curriculum nor more focussed on essay writing. This finding indicates that teachers knew what the changes were in the new exam when compared with the previous one, though none of the teachers (according to teachers' responses to TQ9) had been involved in workshops relating to the exam.

Table 4-2 Teachers' awareness of key differences between new and old exams

Q12.What are the key changes that you have perceived in the revised exam paper with comparison to the previous exam?								
Variables	1 Str. disagree	2 disagree	3 Undecided	4 Agree	5 Str. agree	Mean	SD	Rank
More emphasis on reading passages stated in the textbooks	-	2%	-	62%	36%	4.32	.584	1
More emphasis on grammatical usage	-	4%	4%	92%	-	3.88	.434	2
More related to the principles of the current English curriculum	55%	45%	-	-	-	1.45	.500	3
More emphasis on essay writing	64%	36%	-	-	-	1.36	.482	4

Similar findings were reported by some teachers during the interviews, particularly when they were asked about similarities and differences between the new and old exam, all provided similar responses, and showed awareness of the change that occurred. For example, one teacher stated:

The two exams are completely different. Under the new policy students are required to transfer their answers into an answer sheet purposively provided, while in the old exam students not. The old one tests writing while the current one overlooks it. The old exam also tests reading comprehension based on a given passage followed by comprehension questions, however, it was tested subjectively. In the new exam policy, testing students' reading comprehension is based on objective questions revolving around general or specific information taken literally from the reading texts stated in the prescribed textbooks. Testing the writing skill and testing reading comprehension properly are totally ignored and marginalised in the new exam. (T1, Tamer).

Another teacher added:

In the new exam all questions are objectively-based, such as multiple-choice, true-false and matching items, whereas most of the questions in the old exam were designed on a subjective way. In the new exam, many students can pass or do well on it even by chance. So students' achievements in the exam do not reflect their real level of language (T4, Sulaiman).

#### **4.2.1.2 Teachers' Attitudes towards the New Exam and Its Quality**

In addition to TQ12, TQ13 asked teachers about their views towards the new exam, regarding its format and quality. The findings are shown in Table 4.3. The table shows that all teachers have similar negative attitudes and opinions about the exam. The vast majority of teachers opined the exam as not 'a good tool to gauge students' performance' and 'inappropriate'. All teachers regarded the exam as not a good indicator of a student's ability in using language in real life situations. However, for the fourth value, rank 1, a high proportion of teachers (59% agreed and 37% strongly agreed) considered the exam as to some extent appropriate if it included testing other aspects such as: listening and speaking, and also testing reading and writing properly. These results may be linked to teachers' responses

to TQ15, where all teachers agreed that it is better to assess students in all four language skills and their components (this is elaborated in Table 4.5 below).

Table 4-3 Teachers' beliefs and attitudes towards the new exam.

Q13. According to your own beliefs, the new revised exam is								
Variables	1 Str. disagree	2 disagree	3 Undecided	4 Agree	5 Str. agree	Mean	SD	Rank
To some extent appropriate but it needs to include other aspects such as: listening and speaking, and also testing reading and writing properly	-	1%	3%	59%	37%	4.32	.584	1
Not appropriate at all	8%	61%	26%	5%	-	2.28	.683	2
A good tool to gauge students' performance	3%	89%	2%	6%	-	2.11	.530	3
A good indicator of students ability in using language in real life situation	71%	29%	-	-	-	1.29	.456	4

This finding suggests that teachers' perceptions of the changes in the revised exam are, to a large extent, incompatible with the intended washback envisaged by exam constructors, especially with regard to aspects that relate directly to language learning and teaching.

Although similar results were obtained from interviews, during the interviews, most teachers, and inspectors as well, shared the opinion that the new exam is better than the previous one in terms of style and format. Teachers also claimed that it is easier to work under the conditions of the new exam, enabling their students do well in it. Transcripts showed that students can do well in the new exam without the need to study as hard if they have practiced the way of answering the questions. This indicates that the exam's face validity is quite high,

as the possibility of the participants making haphazard responses was reduced by providing both an example and instructions which were translated into Arabic; besides, adequate time was given to students to complete the test. One teacher said: "the method and style of the exam is good and useful for students. It makes students answer the exam easily without a need to compose answers, they instead choose one appropriate answer, and that is it (T5, Almadani). Another teacher stated "I myself and many other teachers find teaching English under the conditions of the new exam easier than before. But this does not mean that the exam is a good language test" (T7, Badi). Also inspectors' interviews showed that most of the inspectors welcomed the exam in terms of design and format. One of the inspectors stated that the new design and format of this exam is good as it facilitated the way in which students can answer the questions, and the way in which the answer sheets are marked and the results are issued" (Insp.1, Ben Husain).

Despite the above positive views, notwithstanding, most teachers criticized the exam in terms of practicality and content validity. In the first instance, informants indicated that the exam lacks practicality as some teachers mentioned that last year exam contained spelling mistakes and ambiguous items, as a consequence, students did not know what was exactly needed. (in the interviews I was usually maintaining a copy of the new exam in order to raise some related issues with teachers). During one of the interviews, a teacher pointed to the exam paper saying:

Look! You can see here that this item is vague as it asks students: "Libya is in south Africa", true or false. In this case the student is quite confused, does this question test grammar or knowledge about geography? if it tests

grammar, it is true, and if it tests knowledge of some information, it is false. So, it is not clear enough what is the question/item intended to test. Here the student's answer can be wrong or might be correct, depending on the exam designer's intention, and the kind of answer that was inserted into the computer. Even if the student does not know the answer, he has 50% chance of being right (T8, Alghoul).

Interestingly, this statement is in line with the researcher's comments on the exam reviewed in section 4.2.5.

The exam was also criticised in relation to students' language level, as one teacher stated:

The exam results do not reflect the students' real levels in English. That is, we find some students who succeeded the exam with a very high mark, but he or she cannot express himself or herself in a real life situation or even write a very short well-structured paragraph (T8, Alghoul).

Another teacher stressed:

I believe that the exam encourages our students to memorise some information stated in the texts of the course-book prescribed. For example, regularly includes questions about general information, such as dates of specific events or asking about places of some capital cities. So, if students grasp such information they would be able to pass it (T6, Asghayer).

Regarding exam content, data elicited from teachers' interviews (also document studies, see 4.2.5) suggests that the exam strategy does not represent the current English curriculum by ignoring some of its important elements. For example, one teachers declared that "the exam is not comprehensive. I really did not like this exam. I think the previous one is better than this, especially when it comes to testing writing and grammar" (T3, Hisham). Another teacher added: "the exam does not cover the whole components of the textbook prescribed. What the exam, further, needs is to include questions on writing, speaking and listening" (T5, Almadani). During inspectors' interviews, the majority of inspectors, in line



with teachers' statements, criticised the exam for many shortcomings, particularly in terms of content vis-à-vis the content of the current curriculum of English. For example, one inspector reported:

This exam will never motivate students to use integrated skills and it does not cover the whole syllabus as it only tests reading improperly, and to some extent grammar, and neglects testing the communicative aspects of the language, though the existing teaching syllabus recommends them. The previous exam covers more aspects from the syllabus than the new exam. This has a detrimental impact on teaching and learning these aspects in the classroom, as teachers and students were found ignoring them..., it does not cover the all components of this curriculum, nor it would prepare students for their future education (Insp3, Mahjoub).

So, admitting that exam content is usually more important than its format, inspectors' views about the exam and its implementation were generally negative, and in accord with those of teachers reported above. These results indicate that the exam has failed to achieve some important objectives as envisaged by exam constructors; however, it succeeded in fulfilling others. Also, other unintended consequences have emerged.

Thus, teachers' interviews, triangulating with inspectors', provided invaluable information regarding test quality and impact. The results of the interviews concerning the research participants' attitudes towards the new exam confirm the findings of the responses to teacher questionnaires as shown above in Table 4.3. Particularly, teachers criticised the exam especially in terms of its content validity as most teachers declared that congruence between the new exam and the curriculum is tenuous. This indicates that the exam does not constitute a representative sample of the language skills it purports to test, and needs to be recast to cover the components and achieve the objectives of the current English curriculum. Also, the transcripts indicated that the new exam focussed on rote

learning skills which in turn discouraged understanding and classroom participation. However, evidence to the contrary also existed, as contradictions occurred throughout interviews. For example, whilst most teachers found the exam poor in terms of content, the same teachers welcomed the policy regarding exam format and style.

#### **4.2.1.3 Teachers' Perceptions of the Exam Concerning Their Classroom Practices**

Regarding teachers' questionnaire, this category was dealt with based on two sub-categories: any possible pressure or extra work the new exam has put on teachers in terms of their classroom instruction (TQ14); and the language areas in which they think students should be tested (TQ15). These perceptions are listed and reported below in Tables 4.4 and 4.5.

Firstly, as shown in Table 4.4, teachers indicated that 'feeling more responsible for the success or failure of students on the new exam', 'organising focused activities that reflect exam activities' and 'familiarising students with the content and format of the new exam' were the greatest demands on them under the conditions of the new exam. Also, most teachers (71% agreed and 1% strongly agreed) felt that the exam forced them to revise the existing teaching materials. However, the majority (92%) perceived that the exam did not encourage them to follow a new teaching syllabus, i.e. resorting to teaching resources other than the existing textbooks (see also 4.4). The findings relating to these particular issues indicate that the exam did have a direct effect on teachers' perceptions in relation to their classroom instruction (Sections 4.2.2 and 4.2.3 report the findings

regarding the impact the new exam has induced on teachers' classroom instruction and curriculum planning).

Table 4-4 Extra work or pressure involved in teaching towards the new exam

<b>Q14. What kind of extra work or pressure if any do you think the revised exam has put on you in your teaching?</b>								
Variables	1 Str. disagree	2 disagree	3 Undecided	4 Agree	5 Str. agree	Mean	SD	Rank
Feeling more responsible for the success or failure of your student on test	-	1%	1%	58%	40%	4.37	.562	1
Organizing more focused activities that reflect exam activities	-	-	-	64%	36%	4.36	.482	2
Familiarising students with the content and format of the revised exam	-	-	-	86%	14	4.14	.348	3
Employing new teaching methods	-	24%	6%	69%	1%	3.47	.869	4
Revising the existing teaching materials that you have	-	27%	1%	71%	1%	3.46	.903	5
Doing more lesson preparation	-	41%	-	51%	8%	3.26	1.088	6
Following a new syllabus	-	92%	4%	4%	-	2.12	.433	7

Secondly, when teachers were asked about the language areas in which they think students should be tested (TQ15), all teachers, as depicted in Table 4.5 below, showed disapproval on the idea of assessing students in only one aspect/skill of the language. They, instead, perceived an integration of the four language skills and their components would be a better way of assessing students' language achievement in class. These aspects include: 'oral and listening skills', 'writing and spelling', 'reading and vocabulary' and 'grammar'. Teachers were given space after this question to add their own comments. However, none added anything. The findings under this category are in accordance with the results reported above in section 4.2.2 Table 4.3, rank 1.

Table 4-5 Language areas in which teachers think it is better to assess students

<b>Q15. According to your educational beliefs, it is better to assess students in the following area(s) of language:</b>								
Variables	1 Str. disagree	2 disagree	3 undecided	4 Agree	5 Str. agree	Mean	SD	Rank
An integration of all aspects mentioned above	-	-	-	13%	87%	4.87	.3380	1
Only oral and listening skills	100%	-	-	-	-	1.00	1.000	2
Only writing and spelling	100%	-	-	-	-	1.00	1.000	2
Only reading and vocabulary	100%	-	-	-	-	1.00	1.000	2
Only grammar	100%	-	-	-	-	1.00	1.000	2

#### 4.2.2 Washback on Teachers' Instructional Practices

This section reports teachers' responses to questionnaire and interview questions relating to the influence of implementing the new exam on teachers' everyday classroom instructional practices. Pilot observations were also important in determining the teachers' instructional behaviour in class. The results will contribute to address the second research question. For this particular purpose, TQ Part Three, Section One was designed. This included 11 questions, TQ17 – TQ27. The findings of teachers' responses are reported in this section according to two main sub-themes: a) teachers' teaching practices and b) teachers' classroom testing practices.

##### 4.2.2.1 Teachers' Teaching Practices

The questionnaire and interview questions regarding this theme were based on two washback hypotheses taken from Alderson and Wall (1993: 120-121): "A test will influence **how** teachers teach"; and "Tests will have washback effects for ...

some teachers but not for others” (bold in original). Two main sub-categories were identified: teachers’ teaching methods and techniques, and teachers’ classroom teaching activities.

#### **4.2.2.1.1 Teachers’ Teaching Methods and Techniques**

TQ17 comprises five items asking teachers about what key changes they had made to their teaching in line with the exam change. Table 4.6 below lists these changes in descending order according to their frequency, means, standard deviations and rank.

The table shows the order of alterations teachers claimed to have made as a result of administering the new exam. It shows that all teachers (in the first two variables) seemed to have tailored their teaching to meet the new exam demands and content, particularly when the exam is imminent. During teachers’ interviews, the likely effect of ‘proximity’ to the exam on teachers’ practices was also reported as most teachers declared that the closer the final exams, the more their classroom instructional practices become exam-like. As one teacher stated:

I usually try to make my lessons exciting and interesting, and students active by encouraging them to practice pair and group work activities .... But because of the demands of the final exam, as the exam approaches—approximately one or two months before the exam date—I usually change my instructional behaviour to reflect exam requirements, during which I train my students on the way how they can deal with the new exam (T5, Almadani).

The table also shows that most teachers (73%) have adopted new teaching methods. This result is highly compatible with the teachers’ responses to TQ14 (see table 4.4 above) when they were asked about their perceptions of the kind of pressure the exam has put on them in terms of ‘employing new teaching methods’.

However, the mean is lower than two when it came to ‘using a more communicative approach in teaching’ and ‘emphasising listening and speaking skills’. This may be associated with teachers’ perceptions of the key changes they perceived in the revised exam vis-à-vis the old one (see Tables 4.1 and 4.2 above). This finding reflects teachers’ awareness of the underlying aims behind the new exam and its content, thus suggesting that teachers understood the need for students to be prepared in the areas focused on in the test (see also Tables 4.10 and 4.15, TQ19). It was expected that teacher behaviour in this area would have negative repercussions regarding their selection and arrangement of the teaching materials and curriculum (see 4.1.3.2, Table 4.16 that reports teachers' responses to TQ30).

Table 4-6 Alterations to teaching methods and techniques due to exam change

Q17. What are the most significant changes you have made in your teaching in the context of the new revised exam content?								
Variables	1 Str. Disagree	2 disagree	3 Undecided	4 Agree	5 Str. agree	Mean	SD	Rank
Practicing activities similar to the final exam increased frequently in my classroom as the exam approached	-	1%	-	66%	33%	4.31	.526	1
Teaching according to the test content	-	-	-	72%	28%	4.28	.451	2
Adopting new teaching methods	-	26%	1%	71%	2%	3.49	.904	3
Emphasizing listening and speaking skills	48%	52%	-	-	-	1.52	.502	4
Using a more communicative approach in teaching	56%	44%	-	-	-	1.44	.499	5

Teachers’ interviews not only supported the findings tabulated in the above table, but they also provided in-depth information about the influence of the exam on methodologies and techniques teachers use relating grammar teaching, teaching of

vocabulary, reading and writing and use of communicative activities. It appeared that students' expectations were deemed a priority for teachers. All interviewed teachers admitted that they have followed certain teaching practices as a result of the exam change in order to meet students' needs. This was reflected on a teacher's reaction to one of his students' demand, as he narrated:

During one of my lessons, I asked students to do an exercise stated in their textbook which was about writing composition; some students surprised me by saying: "no teacher! no teacher, please! We do not want to take this because it is not included in the exam. Please teacher, just give us things that would appear in the exam". ... As a consequence of students' behaviour, my behaviour in class has changed to meet my students' expectations as their primary concern has been passing the exam. So, I should show my students how the final exam would be and how to answer it in order to do well on it (T8, Alghoul).

Regarding teacher talk in class, as shown in Table 4.7 below, the mode 'talk to the whole class' took up the highest percentage of the class time among teachers. This result reflects the findings reported in the following section (see Table 4-9 TQ21), which showed that teachers generally did not organise group work or discussion, and/or integrated language tasks, during which teachers are likely to be encouraged to carry out such activities as far as the principles of the current curriculum are concerned.

Table 4-7Teacher talk

Q22. How often do you do the following in class?								
Variables	1 never	2 seldom	3 some- times	4 often	5 always	Mean	SD	Rank
Keep silent	98%	2%	-	-	-	4.30	.46057	1
Talk to the whole class	-	-	-	70%	30%	2.00	.56533	2
Talk to groups of students	13%	68%	19%	-	-	1.90	.62765	3
Talk to individual students	25%	60%	15%	-	-	1.02	.14071	4

Similarly, during teachers' interviews, most teachers stated that they are generally the dominant speakers in class, indicating that they adhere to a teacher-centred approach in their instruction; however, the curriculum recommends teachers to apply communicative techniques and activities (see section 1.3.1.2). Some teachers spend much of the instructional time talking, around half of which is spent talking about exam-related issues.

As one said:

I speak too much in class; I mostly spend 80% or so from the class instructional time talking to the whole class, and rarely to individuals or groups, spending about 50% preparing and practicing students on the most frequently and possibly occurring questions in the exam. The thirty one-year experience I have as an English teacher helped me to expect the questions that might be in the exam (T4, Sulaiman).

Some teachers admitted they use a lot of Arabic during their lessons to explain grammar rules and demonstrate meanings of new words and patterns, to enable their students to grasp as much information as possible from the texts in the course-book. The motive behind this, according to teachers, was to meet students' expectations. As one teacher said:

My way of teaching has changed or has been different from what a teacher should teach regarding the principles of the prescribed curriculum. I usually focus on teaching language rules directly using mostly Arabic more than English. This is because of the final exam and because of three other reasons: students' low levels in English, big classes and the insufficient time allotted for teaching English. For example, I give students the new words and their meanings using the blackboard encouraging students to memorise them. After that I explain the grammar rule that stated in the unit, and then explain the text provided. This is in order for students to comprehend the content of the exam. I am doing this as there is a lot of pressure on me from my students in classrooms, as they always ask me which questions and information that would possibly be in the final exam (T4, Sulaiman).



During the pilot observations of two teachers conducted in Phase One of the study, it was noted that the observed teachers were teaching to the test, reflecting questionnaires responses and interviews statements reported above. One of the observed teachers commented during a post-observation interview saying:

I speak too much in class; I mostly spend 80% or so from the class instructional time talking to all students, and very seldom to individuals or groups, spending about half of the time practicing students on the most frequently and possibly occurring questions in the final exam, particularly, aspects related to grammar and reading (Ta1, Makhoulf).

This statement was very similar to those elicited from teachers' and inspectors' interviews in the main study. For instance, one inspector stated:

Teachers to some extent practiced some activities related to the content of the final exam. Indeed, most of them spoke Arabic more than English in class especially when discussing the grammar sections, which they devote lengthy of time. Also they tended to use Arabic when explaining the meaning of some words and patterns from the reading texts, encouraging students to grasp every bit of information the texts contain. In many cases teachers adhered to traditional methods of teaching in which teachers acted as senders and students as negative receivers. (Insp.1, Ben Husain).

The other observed teachers further explained that there are other reasons, in addition to the exam, behind the teacher-centered approach some teachers adhered to, as he commented:

If I abide by the instructions and the objectives of the curriculum, keeping in mind the insufficient time for teaching English per week and the students' low level of English, I would not be able to finish or cover the syllabus, and consequently I will be blamed on that by some stakeholders such as the head-teacher, students and their parents. And this is the case with many teachers in our schools(Ta2, Mustafa).

The above results reveal that grade 9 teachers are negatively influenced in terms of everyday classroom teaching practices due to the new exam. Teachers seemed to use the Grammar Translation Method when explaining grammatical aspects and

providing word meanings, and to a certain extent they adhere to the Reading Method when teaching the reading texts in the course-book. Due to the new exam and other reasons such as time constraints and students' low levels in English, most teachers adhered to teacher-centered approaches. This indicates that classroom teaching practices are not only influenced by the exam, but also by other intervening variables, such as insufficient time allotted to teach English, crowded classes, and low levels of student proficiency. Such behaviour contradicts the strategy of the current learner-centred English curriculum, which revolves around enhancing students' communicative skills and their reading and writing abilities (Orafi and Borg, 2009).

Although most teachers taught to the test, two interviewed teachers gave somewhat different responses. One said:

Honestly speaking, in the beginning of the school year, mainly in the first three or four months, I teach my students in a normal way without reference to the new exam format or content. I usually try to make my lessons exciting and interesting, and students active by encouraging them to practice pair and group work activities .... But as the exam approaches I tend to teach to the test ... (T5, Almadani).

This result reflects teachers' responses to TQ27 which explored the extent to which teachers' Grade 9 classroom teaching was similar or dissimilar to their teaching with other grades. It is important to recapitulate here that all grade 9 participant teachers of English teach other grades (5, 6, 7, and/or 8) as well. Although, the responses, as shown in Table 4.8 below, indicate that teachers' teaching policies, generally, are similar in both contexts, the percentages indicate that there is no general consensus among grade 9 teachers that their teaching

practices are identical as 29% of teachers responded that their teaching was ‘to some extent similar’ in both contexts.

Table 4-8 Degree of similarity between teachers’ teaching in grade 9 and in other grades

<b>Q27. To what extent is your English teaching in grade 9 classes and other grades/classes similar?</b>	
To some extent similar	29%
Similar	44%
Very similar	27%

Interestingly, referential statistics (see section 4.5.3 below) showed that most of teachers whose responses were ‘to some extent similar’ were teaching lower-level classes, i.e. grade 5 and grade 6. This in turn would suggest that the washback effect in this particular issue was, to some extent, limited to grade 9 classes. This particular finding was supported by interview data where most interviewed teachers (also noted by inspectors) who taught grades 5 and 6, in addition to grade 9, proclaimed that they, to a large extent, abide by the requirements of the current curriculum when teaching these classes. However, they encounter difficulties relating to time constraints and class size teachers. One teacher stated:

As the final exam is administered only to grade 9 students ..., I often try to make my lessons in grade 5 classes more interesting and the students more active by applying communicative tasks .... But the main problem facing me in this process is the limited time allocated for teaching English, and the big numbers of students I have in some classes (T9, Sultan).

This was confirmed by some of the interviewed inspectors as one of them said:

Not all teachers use traditional methods in their teaching, I observed some teachers, especially in lower-level grades, such as grades 5 and 6, use integrated skills and try to perform communicative activities in their

classroom teaching, and they were using English as a medium of instruction supplemented with occasional Arabic explanations (Insp.4, Assadi).

Although classroom observations were not undertaken in the mains study due to the difficulties and challenges encountered (see Chapter Three, 3.8), the findings reported above were found consistent with classroom observations undertaken in the pilot study, where teachers were noted more curriculum oriented than exam oriented in lower-level classes. That is, the two observed teachers adhered to more communicative instruction in grade 5 and 6 classes than in grade 9 lessons. This was further confirmed by a post-observation interview with one of the two teachers, who stated that: “as I am the only assessor of students’ performance in the whole year and because of self-satisfaction in teaching I try to abide by the requirements of the current curriculum and students’ needs in learning the language properly” (T1a, Makhloof).

#### **4.2.2.1.2 Teachers’ Classroom Teaching Activities**

The second sub-theme related to washback on teachers’ instructional practices was mainly covered via teachers’ responses to TQ21, where teachers were asked to range a set of classroom activities in terms of frequency. As shown below in Table 4.9, the most frequent activity carried out by teachers was to ‘use previous years’ tests to review the drills and format of the revised exam’. This particular finding is further elaborated and thus confirmed by during teachers’ and inspectors’ interviews reported in the following section. It can also be noted from Table 4.9 that teachers paid less attention to the other activities ranked 3, 4, 5, and 6. The table, however, unveils (at the very bottom) that teachers were very rarely practicing activities of integrated language skills in their teaching, indicating that

teachers adhered to less communicative methods of language teaching. This matches the findings shown in Table 4.6 (TQ17) reported in the previous section, where teachers overwhelmingly disagreed that the exam made them use a more communicative approach in teaching or made them emphasise listening and speaking skills, contradicting the principles of the current English curriculum, but being more oriented towards the exam demands. This also ties in with the findings reported above in the previous section as shown in Table 4.7.

Table 4-9 Teachers' classroom teaching activities

<b>Q21. How often do you do the following activities in class?</b>								
<b>Variables</b>	<b>1 never</b>	<b>2 seldom</b>	<b>3 some- times</b>	<b>4 often</b>	<b>5 always</b>	<b>Mean</b>	<b>SD</b>	<b>Rank</b>
Use previous years' tests to review the drills and format of the revised exam	-	-	8%	91%	1%	3.93	.293	1
Explain specific language items such as words or sentences	-	-	39%	60%	1%	3.62	.508	2
Demonstrate how to do particular language activities	-	18%	52%	30%	-	3.12	.685	3
Tell the students the aims of each lesson	-	36%	43%	21%	-	2.85	.743	4
Correct and explain homework	-	68%	32%	-	-	2.32	.468	5
Make revisions to previous lessons	-	72%	28%	-	-	2.28	.451	6
Organise group work or discussion	17%	81%	1%	1%	-	1.86	.449	7
Organise integrated language tasks	34%	57%	8%	1%	-	1.76	.637	8

It was elicited from teachers' interviews, and noted during classroom observations (in the pilot study) that whilst teachers' main classroom activities focussed more on teaching reading tasks and grammar points and teaching them to students deductively, teachers were less much more oriented towards conducting other activities related to the communicative aspects of language as the curriculum

recommends. Interestingly, this finding corresponds with that reported in section 4.2.3 TQ14, both suggested that teachers' perceptions of what they may do in response to examination change as they report in the questionnaires tallies with what they actually do in classroom real practices. This contradicts with the view that ... see chapter 2, the literature.

#### **4.2.2.2 Teachers' Classroom Testing Practices**

This sub-section concerns the extent to which teachers classroom testing practices were influenced by the administration of the new exam. Data pertinent to this aspect was elicited via questionnaires and interviews, besides; document study as discussed in section 4.2.5 below. All contributed to address the second research question. Two main sub-themes were identified: teachers' use of mock exams and the on-going tests teachers write for students during the school year. To recapitulate, the new exam was designed and administered by the Authority of Education as a school-leaving exam, and teachers in their schools are accountable for other forms of assessments, such as on-going tests, including monthly and periodical tests during the school year (see chapter one, section 1.3.2.1 for elaboration).

Firstly, regarding teachers' use of mock exams, when teachers were asked about what kind of learning strategies they recommend to their students as far as the content of the new revised exam is concerned (TQ19), the teachers, as shown in Table 4.10 below, overwhelmingly agreed that the strategy 'to train on previous exam papers similar to the revised exam' as the most frequently recommended by them.

Table 4-10 Learning strategies in the context of the new exam

Q19. What are the learning strategies you recommend to your students in the context of the new exam? I recommend students...							
Variables	1 Str. Disagree	2 Disagree	3 Undecided	4 Agree	5 Str. Agree	Mean	SD
To train on previous exam papers similar to the revised exam	-	5%	-	94%	1%	3.91	.455

Moreover, reviewing Table 4.9 above (TQ21), it can be seen that the teaching activity ‘use previous years’ tests to review the drills and format of the revised exam’ had the highest rate of frequency among teachers, as 91% of teachers often did this, 8% sometimes and 1% always, with a high mean score of 3.39. This finding is a reflection to teachers' responses to TQ20, when all of them (16% agreed and 84% strongly agreed) considered ‘external exams’ as the most influential factor in their teaching.

Most teachers’ statements (also inspectors’) at the semi-structured interviews supported teachers’ responses to questionnaires. Teachers reported that they usually make their students practice exam-related activities and train them on mock exams similar to the new exam. They do so in order to prepare students for the new exam. One teacher said:

The exam has made my students feel worried about what and how they do to pass it .... So I became worried as well and there has been a lot of pressure on me. Therefore, I often train my students on the drills of the exam to familiarise my students with it as its format is new for them. For example, when I finish a lesson or a unit I should come up with some questions similar to the final exam questions, which I give to students in a form of homework or a quiz either by the end of the class or during the next class. This method was very fruitful and beneficial as I noticed my students have done well in the exam, and the school head-teacher as well as students’ parents were satisfied with the results (T10, Marwa).

These findings suggests that most teachers perceived the use of mock exams as an important activity to prepare students for the new exam, and perhaps as a way to provide teachers with feedback on students' expected performances in exams.

Secondly, interviews data not only revealed that teachers made use of mock exams in their daily classroom instruction, but teachers further designed their own tests akin to the final exam in terms of format and content, applying them on a regular basis (appendix 31 shows a sample of teachers' on-going tests). One interviewed teacher declared:

I coach my students on how to answer the exam by giving them practices on some mock exams and past years' exams. I usually write and design my own exams similar to the final/new exams. I administer these tests to students as weekly, monthly or periodically tests (T7, Badi).

Document study, i.e. comparing a sample of the new exam (see appendix 2a) with the on-going tests that grade 9 teachers write for their students during the school year (appendix 31 contains a sample of these tests), high consistence between the two types of exam can be noted in terms of design and content (for details see section 4.2.5).

To probe further into this issue, during inspectors' interviews, inspectors were asked whether they observed teachers adhering to exam-related activities. All inspectors mentioned that they noted the majority of grade 9 teachers using mock exams and previous years' exams in classrooms. They also stated that teachers designed their own exams that were similar to the new final exam, using them as on-going tests during the school year, in order to prepare their students for the new exam format and content. The reason behind this, according to inspectors, was that teachers felt accountable for their students' performance in the exam. It



seems there is concordance between what teachers said they do under the conditions of the new exam regarding teaching and testing practices and what inspectors said what teachers do based on their observations in classrooms (this was also reflected by the classroom observation sessions undertaken during the pilot study).

Thus, interviews findings corroborate teachers' responses to the questionnaire (i.e. TQ25). Using multiple-choice items, (consistent, slightly consistent, to some extent consistent, and highly consistent), Table 4.11 showed that the vast majority of teachers (85%) found the content of their classroom on-going tests highly consistent with that of the final/new exam, and the remaining 15% of teachers said that the two extremes are to some extent consistent. Noticeably, none of the teachers showed reactions to the variables 'inconsistent' or 'slightly consistent'. This finding indicates that teachers are influenced by the new exam when they plan and write their classroom on-going tests to grade 9 classes during the school year. Interestingly, this appears in line with the finding reported in Table 4.9 TQ21 above, in which a considerable number of teachers stated that they use previous years' tests to review the drill and format of the new exam as a very common classroom activity.

Table 4-11 The consistency between the content of teachers' on-going tests and the content of the new exam

<b>Q25. To what extent is the content and format of the classroom tests you usually write to your grade 9 students consistent with the content and format of the revised exam?</b>	
To some extent consistent	15%
Highly consistent	85%

Furthermore, teachers in questionnaires, in a multiple-choice format, were asked (TQ26) to rate the consistency of the content of their grade 9 on-going tests with the content of the tests they write for grades they teach other than grade 9. Table 4.12 below shows varied responses: 22% of teachers said slightly consistent, 31% to some extent consistent, 47% highly consistent and none chose 'inconsistent'. These findings indicate that teachers' testing practises in both contexts (i.e. grade 9 and other grades) are not identical, and the consistency appeared to be lower than that in the previous category (see Table 4.11 above), suggesting that teachers, in some cases, were less affected by the content and format of the new exam.

Table 4-12 The consistency between teachers' on-going tests to grade 9 and other grades

<b>Q26. As you teach classes other than grade 9, how consistent are the classroom tests you usually write for grade 9 classes with those you write for other grades/classes?</b>	
Slightly consistent	22%
To some extent consistent	31%
Highly consistent	47%

Interestingly, descriptive statistics for this variable compared with the independent variable (i.e. classes/grade teachers teach other than grade 9; TQ5), revealed that most teachers whose responses were 'slightly consistent' and 'to some extent consistent' were teaching grades 5 and 6. This finding matches the finding reported above (see Table 4.8 TQ27), where teachers were asked about the extent to which their grade 9 classroom teaching is similar to their teaching for other grades, indicating that washback in these two particular settings was, to a large extent, restricted to grade 9 and 8, and to some extent grade 7, but not to grades 5 and 6. This suggests that washback effect from the exam is less detectable when

teachers undertake testing practices in the lower-level classes, i.e. grade 5 or 6 than teaching higher classes specifically grade 9. Congruent results were obtained when the effects of these independent variables on the main dependant variables were tabulated (see sections 4.2.4.3.1 and 4.2.4.3.2 below). Moreover, as most of the interviewed teachers were teachers of many years of experience and holding DT or BA qualifications, the results presented in this section substantiate those elicited from inferential analyses reported below in sections 4.2.4.1 and 4.2.4.2.

To sum up, triangulating data from all sources (data and methodological triangulation), suggests that washback effects from the new exam on teachers' classroom instructional practices—teaching and testing—were observed, particularly in grade 9 classes, but registering a low degree of washback intensity with other grades. However, washback effects on testing practices were more discernible than on teaching practices.

### **4.2.3 Washback on Teaching Materials and Curriculum**

The third theme highlighted, which related to exploring the washback effect from the new exam is reported in this section. The purpose was to report teachers' responses to questionnaires and interviews related to the influence of implementing the new exam on teachers' use of teaching materials and curriculum, informing the third research question. As there were parallels between questions in the questionnaires and the interviews, the design of both methods regarding this category was based on two washback hypotheses: "A test will influence **what** teachers teach"; and "Tests will have washback effects for ... **some** teachers but not for others" (Alderson and Wall, 1993: 120-121).

The findings are reported below in two main categories: a) teachers' choice of teaching materials (i.e. learning resources) they use in their instruction (TQ23); and b) teachers' selection and arrangements of the content of teaching materials vis-à-vis the current English curriculum (TQ29 and TQ30).

#### **4.2.3.1 Teachers' Choice of Teaching Materials, Learning Resources**

TQ23 was designed using Likert items of frequency, where 1 = never and 5 = always, to find out the extent to which the new exam made teachers use extra teaching and learning resources in addition to the existing textbooks. As shown in Table 4.13, teachers in this category expressed how often they used a list of seven resources in their teaching. It should be noted here that the trinity of 'course-book', 'workbook' and 'teacher's book' are all covered by the term 'textbooks', which have been prescribed and introduced in schools by the Authority of Education since 2000 (Orafi and Borg, 2009).

It can be seen that the course-book and the blackboard were the most frequently used resources, followed by the workbook and teacher's book. On the other hand, pictures and cards were less frequently used in teaching, and none of the teachers utilised a language laboratory or television or radio programmes.

Table 4-13 Teaching and learning resources and aids

Q23. How often do you use the following teaching and learning aids in your teaching?								
Variables	1 never	2 seldom	3 sometimes	4 often	5 always	Mean	SD	Rank
Course-book	-	-	-	26%	74%	4.74	.441	1
The blackboard	-	-	-	97%	3%	4.03	.172	2
Workbook	-	-	27%	71%	2%	3.75	.479	3
Teacher's book	-	1%	79%	20%	-	3.19	.419	4
Pictures and cards	55%	43%	2%	-	-	1.47	.540	5
Language lab.	100%	-	-	-	-	1.00	.000	6
Television/radio	100%	-	-	-	-	1.00	.000	6

Teachers' excessive reliance on textbooks in their teaching is associated with their responses to TQ20, when they overwhelmingly (92% agreed and 3% strongly agreed) considered 'textbooks' as another influential factor, in addition to 'external exams', in their teaching. This finding indicates that teachers often relied solely on the textbooks in their teaching, not using other teaching and learning resources. Teachers' responses hitherto indicate that the new exam had a limited effect on teachers' choice of textbooks as teachers overwhelmingly relied on the existing textbooks and curriculum. This was confirmed by data elicited from teachers' interviews. The results showed that a heavy reliance on textbooks is ubiquitous in Libyan public schools, and textbooks are by far the most widely available teaching materials. Most interviewed teacher commented that they used only the existing textbooks in their instruction. The following statement is representative:

My classroom teaching is always confined to the content of the current English curriculum .... I just follow each lesson guideline stipulated in teacher's book. I teach based upon the content of the lessons in the textbook and sometimes workbook. The current textbook is quite enough to teach the students even under the conditions of the new exam (T3, Hisham).

This was further confirmed by the interviewed inspectors, who declared that teachers used the current course-book and workbook as a main resource in their teaching. None of the inspectors stated that any teacher used other teaching materials. Nonetheless, teachers' responses to TQ29, as shown in Table 4.14 below, suggests that 29% of the teachers planned and arranged their teaching contents according to both the current textbooks prescribed and other teaching materials they prepared.

#### **4.2.3.2 Teachers' Arrangements and Selection of Teaching Content and Materials**

Although results showed that the new exam has had a very limited washback on teachers' choice of textbooks as they did not resort to other teaching materials, responses to questionnaires (TQ29) and interviews revealed that the exam has induced a profound negative washback on the content of the current prescribed textbooks, i.e. curriculum. The question was designed using a multiple-choice format, giving three options: 'according to the current textbook and its content', 'according to the teaching material you prepare by yourself' or based on a 'mixture of both'. As shown in Table 4.14, 71% of teachers relied on the current textbook and its content in their teaching, and none of the teachers based their teaching solely on the materials they prepare by themselves.

Table 4-14 Teachers' arrangements and planning of teaching content and materials

<b>Q29. How do you arrange your teaching in the classroom?</b>	
According to the current textbook and its content	71%
According to both the current textbooks prescribed and materials you prepare	29%

This finding corresponds with the finding reported in Table 4.15 below (TQ19), when 70% of teachers stated that they would not recommend their students to use teaching materials other than their current textbooks. This finding can also be linked to teachers' responses to TQ14, reported above in Table 4.4, 92% of teachers perceived that the new exam did not encourage them to follow a new teaching syllabus. Thus, as mentioned above, washback on the choice of textbooks was negligible, as no new textbooks were published as a result of the exam.

Table 4-15 Learning strategies in the context of the new exam

<b>Q19. What are the learning strategies you recommend to your students in the context of the new exam? I recommend students...</b>							
<b>Variables</b>	<b>1 Str. Disagree</b>	<b>2 Disagree</b>	<b>3 Undecided</b>	<b>4 Agree</b>	<b>5 Str. Agree</b>	<b>Mean</b>	<b>SD</b>
To use teaching materials other than their current textbooks	19%	51%	1%	29%	-	2.65	1.140

In contrast, however, some teachers (29%) as mentioned above (Table 4.14), responded that they arrange and select the content of teaching according to both the content of the current textbook and other materials they prepare. This finding is, interestingly, associated with teachers' responses depicted in the Table 4-15 above, particularly when 29% of the teachers said they would recommend their students to use teaching materials other than the existing ones.

During teachers' interviews, it was of particular interest to verify the aforementioned results and explore what kind of teaching materials some teachers claimed to utilise in addition to the textbooks prescribed. Some of the interviews

revealed that the teaching materials teachers made use of were of two sources: teacher-made materials exclusively based on the content of the existing textbooks prepared as summaries or exercises/tests taken from some lessons and units in the prescribed course-book and workbook (appendix 32 reflects these practices); and external commercial materials based on pamphlets and booklets commercially published by retired teachers (the researcher also acquired a copy of these materials for reference). For the former, teachers said that many students use these materials, especially before the exam. Although this method, according to teachers, was very demanding, it was helpful for them to teach to the exam, and useful for students to perform well in the final exam. For the latter a minority used the external materials, but they would not recommend their students to use such materials. One teacher said:

I revised the course-book content, I summarised each unit in accordance with the exam content in a form of sentences and important points I envisage to exist in the final exam. The summary is followed by questions based on the information given. The exam made me focus on some areas, such as grammatical structures and meanings of some words and also some information stated in the texts that can be tested in a form of MCQ, true-false, or matching, and ignoring others, such as pronunciation tasks and listening or speaking activities and sections (T1, Tamer).

Another teacher, who claimed that he summarised each unit of the course-book in a form of some questions written as hand-outs and giving them to students as homework or tests, added that: "... many teachers used this method and found it helpful for them and useful for students to pass the final exam" (T4 Sulaiman). (Appendix 32 contains a sample of this material prepared by the same teacher).

Interestingly, two teachers, who stated that they used external materials other than the existing one, commented that the motive behind resorting to extra materials



was that the course-book lacked details, specifically relating to the knowledge of grammar. As one teacher said:

As the textbook prescribed does not provide enough explanation for the grammar rules, I tended to look for another relevant resource to get more information to explain it to my students in class, and sometimes make summaries with exercises of these grammar rules in sheets and give them to students as a reference to study for their exams. But the main idea should be taken or linked directly to the course-book prescribed (T5, Almadani).

With reference to the above finding, it was of interest to find out whether the content of the current textbook/curriculum was influenced by the introduction of the new exam, i.e. teachers' selection of teaching contents. In the questionnaire, teachers, based on curriculum review discussed in Chapter One section 1.3.1.2, were asked (TQ30) what aspects of language from the current English curriculum they focus on most in class, and why: grammar lessons, reading, writing, speaking and/or listening sections. It can be seen from Table 4.16, that 'reading comprehension lessons' was the most common aspect of teaching (97% often and 3% always). This was followed by 'the knowledge of grammar and vocabulary' (76% sometimes and 23% often). Writing aspects were, to some extent, marginalised on teachers' teaching agendas (9% sometimes, 72% seldom and 19% never). As the mean scores drop dramatically, it could be seen that teachers were, to a large extent, ignoring other aspects of teaching, such as the speaking and pronunciation sections that stated in the textbook. Moreover, to a very large extent, they did not pay attention to the listening tasks and activities, violating the objectives of the current English curriculum reported by Orafi and Borg (2009), organised around activities based on communicative principles (see section 1.3.1.2).

Table 4-16 Teachers' selection of the content of teaching regarding the current textbooks

<b>Q30. Which of the following language aspects from the content of the current teaching syllabus do you focus in your teaching?</b>								
<b>Variables</b>	<b>1 Never</b>	<b>2 Seldom</b>	<b>3 Sometimes</b>	<b>4 Often</b>	<b>5 Always</b>	<b>Mean</b>	<b>SD</b>	<b>Rank</b>
Reading comprehension lessons	-	-	-	97%	3%	4.03	.17145	1
Knowledge of grammar and vocabulary	1%	-	76%	23%		3.21	.47768	2
Writing aspects	19%	72%	9%	-	-	1.63	.67652	3
Speaking and pronunciation sections	68%	32%	-	-	-	1.32	.46883	4
Listening tasks and activities	99%	1%	-	-	-	1.01	.10000	5

Teachers' responses and behaviour in this category could be a reflection of the responses reported early in this chapter (see Table 4.2 TQ12), when teachers were asked about the key changes they perceived in the new exam. Also, the finding here corresponds with the finding reported in Table 4.4 TQ14, as 72% of teachers agreed that they think the new exam has put pressure on them to revise the existing teaching materials. All of this supports the finding reported in TQ11, item 3, Table 4.1, variable 5, in which the majority of teachers (79%) disagreed that the new exam aimed to cover all the components of the English curriculum, as envisaged by its constructors.

During observing the two teachers in the pilot study, it was noted that teachers tended to pay more attention to reading texts and grammar points while neglecting others. During the interviews in the main study, most teachers' statements indicated that they tailored the content of the existing curriculum to accord with the content of the new exam, focussing on aspects that repeatedly appear in

the final exam, though this, as they stressed, contradicts their educational beliefs and the principles of the curriculum. One said:

Because of the exam demands, students' expectations, I myself and many teachers I know tended to ignore sections that less frequently occur in the final exam, such as writing, listening and speaking sections, telling students these sections are not important. Meanwhile focusing on other aspects such as grammar points and reading texts and the information they contain (T6, Asghayer).

This was in line with other teachers' statements which unveiled how the reformed exam made teachers adapt the teaching materials. For example, one teacher, who was presenting a booklet he designed for students' exam preparation, commented:

I focus on some aspects such as reading texts and grammar spots, while marginalising other aspects such as the communicative activities stated in the textbook and the writing tasks as they are untested. I also devised some summaries and sheets. Based on the textbooks available, I designed a booklet in which I summarised each unit, section by section, with most important and possibly appearing elements in the exam. This booklet contained about 240 true-false items, about 150 multiple-choice questions as well as matching items. This was very beneficial for students to get prepared for the final exam, and yielded good results. And most of my students, by studying from this booklet, have done very well on the exam. Last year, my students' results reached 90% (T8, Alghoul).

During inspectors interviews, when the above issues were brought up, inspectors all mentioned that they noticed teachers are narrowing the curriculum, tailoring instruction to accord with the new exam. For example, one inspector said:

I noticed teachers focusing on some sections of the textbook while skipping others. They concentrate on the reading passages and the information they include, and the sections that deal with the knowledge of grammar, ignoring the writing exercises, and the listening or oral aspects of the language (Insp.5, Gmati).

Moreover, some inspectors reported that they noticed some teachers preparing some materials derived from the textbooks in the form of summary sheets based on the reading texts and grammar rules prescribed in the course-book, usually

written in a question-answer format. The inspectors mentioned that teachers do so to prepare students for the final exam. Following the above excerpt, another inspector added:

Teachers tend to focus on issues tested in the final exam. They also prepare summaries of some units that reflect the content of the final exam. As a part of my job, one day I asked a teacher about the reason for this behaviour, she said that she is doing so to meet her students' expectations, and because she does not have enough time to teach the syllabus as it should be taught. Frankly speaking, if I was her, I would do the same (Insp.1, Ben Husain).

Interestingly, the interviewed teachers who reported that they revised the content of the current teaching syllabus to reflect the exam content were holding the qualification of DT. This, again, corroborates the results discussed below in sections 4.2.4.1 and 4.2.4.2, where the kind of qualification and the long years teachers spent in teaching have had an impact on their responses to the exam in relation to the context of the existing textbook/curriculum. The statements reported above highlight the deleterious impact the new exam had upon the curriculum. All interviewed teachers admitted that the new revised exam greatly impacted on curriculum planning, because English education in Libyan schools at Basic and Secondary Education stages is driven by measurement, especially regarding the high-stakes public examinations administered to students annually as school/stage-leaving exams. Teachers said they had to change curriculum planning, and hence narrow the current curriculum, in order to meet the new test content. The major perceived changes of their curriculum planning were driven by and derived from the new testing policy. Although only one section in each unit of the current curriculum of the English textbooks is dedicated to developing reading skills, promoting students' contextual ability generated more concern regarding

the selection and arrangement of teaching materials. Selecting the grammar sections and teaching them deductively to students in separate lessons was also emphasised. Teaching the textbook sections that related to the skill of writing, speaking and listening were marginalised, despite being core sections in the prescribed textbooks.

In summation, regarding the extent to which the new exam has induced washback on teaching materials and curriculum, results from various resources clearly revealed the considerable negative effect the new exam had upon the content of teaching—selection and arrangements of teaching materials—as far as the content and principles of the current English curriculum are concerned, however washback on the choice of textbooks was not, to a large extent, materialised. So, exam impact was detected in textbook selection rather than textbook choice.

#### **4.2.4 The Degree of Washback Effect vis-à-vis teacher and context factors**

As reported in Chapter One, section 1.3.3, Chapter Two sections 2.6.1.2, 2.6.1.4 and 2.6.3, and Chapter Three, section 3.3.2.1, this section looks into the washback intensity of the new BECE vis-à-vis teacher and context factors, i.e. independent variables. The aim is twofold: firstly, to complement and hopefully support the data reported in the previous sections. Secondly to explore the effect of the main independent variables (IVs), on the three main dependent variables (DVs) as specified below. This section will concurrently present results and discuss them in relation to the current study's aims and queries, mainly to address the fourth research question. .

Such discussions will increase our understanding of what directions these factors push washback, and add more insights into the nature of this educational phenomenon, washback, attempting to respond to the question raised by Spratt (2005: 23): “would a [male and/or female] well-educated and experienced teacher be more or less likely to adhere to the content of the exam in their lessons?”

There are five factors, i.e. independent variables considered in this study. These included: academic qualification (Bachelor of Arts or BA, Bachelor of Education or BE, Diploma of Teaching or DT), teaching experience (3-6, 7-10, 11-15, or more than 15 years), the grade(s) teachers teach (5, 6, 7 or 8) in addition to grade 9, classroom size (20-30, 31-40, or above 40 students per class), and finally gender. The effect of the stated independent variables will be assessed on the following main dependent variables that are divided into sub dependent variables:

1- Teachers’ teaching practices, teaching methods and techniques

Items related to this variable are: teaching according to the test content, adopting new teaching methods, using a more communicative approach in teaching, emphasizing listening and speaking skills, and the similarity between teachers’ teaching to grade 9 and other grades.

2- Teachers’ classroom testing practices and evaluation, on-going tests

This variable concerns the following items: familiarizing students with the content and format of the revised exam; organizing more focused activities that reflect exam activities; using mock exams to review the drills of the revised exam; identifying consistency between teachers’ on-going tests and the revised exam;

and identifying consistency between teachers' on-going tests to grade 9 and to other grades.

3- Teachers' choice, selection and use of teaching materials and curriculum

This variable comprises the following items: revising the existing teaching materials, focusing more on the reading passages in the textbooks, putting more emphasis on writing aspects, concentrating on the grammatical structures provided in textbooks, using teaching materials other than their current textbooks, and putting more emphasis on speaking and listening skill sections.

***Inferential statistics***

All items were scored on an agreement Likert scale (5-points), time scale (5-points) or consistency scale (4-points). The Likert scale measurement is often considered an interval, hence reporting the mean scores in this chapter will help in understanding the differences between groups. The averages were reported for each independent variable based on the DVs. The effects of the IVs were measured through inferential statistics that can be used to generalize the outcomes from the current sample to the bigger population.

Using inferential statistics, this section will only report significant results and will ignore insignificant ones, with alpha set at .05. Here is a brief explanation of each test used.

1- Analysis Of Variance (ANOVA): the one-way ANOVA was used to measure the effect of IVs, which have more than two levels, on the DVs. Academic qualification, teaching experience, grade(s) and age are IVs that have

more than two levels. The ANOVA was, firstly, run to identify overall differences between groups, i.e. significant statistics. Then the mean scores were compared to find out the significant differences on the dependent variable across the stated groups/levels. An effect size (i.e. eta squared) was calculated to indicate the relative magnitude of the differences between the means of the levels of each independent variable. This gives an indication of practical rather than statistical significance. Formulaically, according to Brown (2008),  $\eta^2$  was calculated as follows:

$$\eta^2 = \frac{SS: \text{sum of squares effect}}{SS: \text{sum of squares total}}$$

Cohen (1988, cited in Pallant, 2010) proposed the following guidelines for effect size applied in this study: “.01 as a small effect, .06 as a medium effect and .14 as a large effect” (254). For ease of interpretation, the decimal point was moved two places to the right, and read as a percentage (Brown, 2008). For this study, the higher the percentage the more effect the indicated independent variable will have on the dependent variable; conversely, the lower percentage gained the less effect that independent variable will have on the dependent variable.

To identify where the differences among the different levels of each variable lie, i.e. which of the groups differ, Post-hoc comparisons using the Tukey HSD tests were conducted (Pallant, 2010).

2- Independent sample t-test: similar to the ANOVA, the t-test was used to measure whether there was a statistically significant difference in the mean scores for two groups/levels (i.e. gender, male vs. female; and class size, 20 – 30 vs. 31



– 40) in relation to their effect on the main three DVs stated above. To gain an indication of the magnitude of the differences between the groups, males vs. females for instance, an effect size (eta squared) was calculated using the data provided in the output. As the SPSS manipulated for statistical analyses in this study does not provide eta squared values for t-tests, Pallant's (2010: 242 – 243) formula for calculating the effect size for t-tests was adopted:

$$\text{eta}^2 \text{ or } \eta^2 = \frac{t^2}{t^2 + (N1 + N2 - 2)}$$

The same guidelines for effect size applied to the ANOVA above were used.

It is important to mention here that the key aspects of the output obtained from the above procedures are presented in the ANOVAs and t-tests tables preceded by the output interpretations of each table (due to space constraints some of the generated outputs mainly the tables of descriptives and the Tukey Post-hoc tests are stated in the indicated appendices).

#### **4.2.4.1 Qualification effect**

Academic qualification was a variable that has three levels: Bachelor of Arts (BA), Bachelor of Education (BE) and Diploma of Teaching (DT). To test the effect of this on the main DVs, a one-way ANOVA was used.

##### **4.2.4.1.1 Teachers' teaching practices, teaching methods and techniques**

One-way ANOVA, as indicated below in Table 4.17, showed a statistically significant effect (at the  $p < .05$  level) of "qualification" on the first dependent variable, i.e. teaching practices, particularly with reference to the three sub-

dependent variables: “teaching according to the test content”,  $F(2,99)=7.38$ ,  $p=0.001$ ,  $\eta^2 = .13$  (BA=4.38; BE=4.00, DT=4.09); “adopting new teaching methods”  $F(2,99)=88.91$ ,  $p=0.000$ ,  $\eta^2 = .65$  (BA=3.81; BE=2.00; DT=4.00); and “emphasizing listening and speaking skills”  $F(2,99)=7.55$ ,  $p=0.001$ ,  $\eta^2 = .13$  (BA=1.40; BE=1.78; DT=1.81).

Despite reaching statistical significance, the actual difference in mean scores between the groups was to some extent small (see table of descriptives in appendix 20a). The effect size, calculated using eta squared, showed varied values for the three mentioned variables. The effect of qualification, under the conditions of the introduced exam, was large on the sub-dependent variable “adopting new teaching methods” (65%). This means that the type of qualification had a large effect on teachers’ adoption and use of new teaching methods under the conditions of the new exam. However, as for the other two variables the effect of the independent variable was quite small, i.e. 13%, which does not seem worthy of concern/comment.

The Tukey Post-hoc test (see appendix 20b) indicated significant difference between the three levels/groups, especially between BE and BA, and between BE and DT across the three sub-dependent variables, while the difference between BA and DT was not significant. DT teachers, however, showed a higher agreement mean score compared to the BA or BE teachers. This means that the significant difference among teachers having different academic qualification is contributed by this mean difference. This indicates that teachers who earned DT appeared to be more affected by the exam, while BE teachers seemed the least affected, suggesting that BE teachers performed less exam-oriented practices

compared to DT and BA teachers. This result seems to correspond with the results obtained from measuring the effect of “experience” on the same sub-dependent variables (see section 4.2.4.2 below).

Table 4-17 Results obtained from one-way ANOVA with eta squared values for qualification in relation to teaching practices.

	Sum of Squares	df	Mean Square	F	Sig.	eta <sup>2</sup>
Teaching according to the test content	2.665	2	1.333	7.389	.001	.132
Within Groups	17.495	97	.180			
Total	20.160	99				
Adopting new teaching methods	52.404	2	26.202	88.912	.000	.647
Within Groups	28.586	97	.295			
Total	80.990	99				
Emphasizing listening and speaking skills	3.366	2	1.683	7.559	.001	.135
Within Groups	21.594	97	.223			
Total	24.960	99				

#### 4.2.4.1.2 Teachers’ Classroom Testing Practices and Evaluation, On-going Tests

Similarly, concerning teachers’ classroom testing practices, the ANOVA test , as shown in Table 4.18 below, revealed a significant impact of “qualification” on this second dependent variable, relating to the main two relevant sub-dependent variables: “familiarising students with the content and format of the revised exam”,  $F(2,99)=17.51$ ,  $p=0.000$ ,  $\eta^2 = .26$  (BA=4.10; BE=4.00; DT=4.63), and “organizing more focused activities that reflect exam activities”,  $F(2,99)=7.37$ ,  $p=0.001$ ,  $\eta^2 = .13$  (BA=4.44; BE=4.00; DT=4.45). The actual statistical difference between the three groups was quite small, as indicated by the mean scores shown in the table of descriptives (see appendix 21a). However, eta squared values indicated practical significant difference, but to a certain degree, 26% and 13% for

the two variables respectively. Tukey Post-hoc test (see appendix 21b) suggests that the three groups were affected differently according to the qualification teachers have. A significant difference appeared between BA vs. BE and between DT vs. BE. BA and DT groups did not differ significantly from each other with  $p > .05$ . Thus, it can be concluded that DT teachers were more affected by the exam as they seemed relatively higher in significance in comparison with the BA and BE teachers, the latter of which scored lowest.

Table 4-18 Results obtained from one-way ANOVA with eta squared values for qualification in relation to teachers' classroom testing practices

	Sum of Squares	df	Mean Square	F	Sig.	eta <sup>2</sup>
Familiarising students with the exam content and format of the revised exam	3.195	2	1.597	17.516	.000	.265
Between Groups	8.845	97	.091			
Within Groups	12.040	99				
Total						
Organising more focused activities that reflect exam content	3.041	2	1.521	7.376	.001	.131
Between Groups	19.999	97	.206			
Within Groups	23.040	99				
Total						

#### 4.2.4.1.3 Teachers' Choice, Selection and Use of Teaching Materials and Curriculum

Finally, teachers' qualifications significantly affected the third dependent variable relating to teaching materials and curriculum use. As presented in Table 4.19 below and shown in appendix 22a, there was a statistically significant difference at the  $p < .05$  level for the three groups of teachers, mainly relating to the five sub-dependent variables. These include "revising the existing teaching materials",  $F(2,99)=83.15$ ,  $p=0.000$ ,  $\eta^2 = .64$  (BA =3.75; BE=2.00; DT=4.09); "focusing more on reading specifically the reading passages in the textbook",  $F(2,99)=13.82$ ,  $p=0.000$ ,  $\eta^2 = .22$  (BA=4.42; BE=3.57; DT=4.36); "using teaching

materials other than the current textbooks”  $F(2,99)=24.13$ ,  $p=0.000$   $\eta^2= .33$  (BA=2.50; BE=1.31; TD=3.63); “concentrating on the grammatical structures provided in the textbooks”,  $F(2,99)=5.95$ ,  $p=0.004$ ,  $\eta^2= .11$  (BA=4.35; BE=3.89; DT=4.27); and “putting more emphasis on writing aspects”,  $F(2,99)=4.22$ ,  $p=0.017$ ,  $\eta^2= .08$  (BA=1.68; BE=1.26; DT=1.90).

Although the size of variance (eta squared) associated with most stated variables was quite small (ranging from 8% - 33%), it was practically large on the variable “revising the existing teaching materials” ( $\eta^2= .638$ ). This means that the kind of qualification has a nearly 64% effect that will make teachers revise the content of the curriculum under the conditions of the introduced exam. This reflects the influence of the exam on the content of the current English curriculum. Mean scores (see appendix 22a) showed different results between certain groups the significance of which was identified by running a test of multiple comparisons. While the Tukey Post-hoc test (see appendix 22b) revealed that the difference between BA and DT was not statistically significant, it showed a significant difference between BA and DT combine and BE group. This again, interestingly, indicates that DT teachers scored significantly higher than both BA and BE teachers, the latter of which registered lower scores, implying that DT teachers were more affected by the introduced exam, while other groups were minimally affected, particularly BE group. The above results complement and substantiate the results elicited from interviews and questionnaires descriptive analyses reported in the above three main sections.

Table 4-19 Results obtained from one-way ANOVA with eta squared values for qualification in relation to teaching materials and curriculum use.

		Sum of Squares	df	Mean Square	F	Sig.	eta <sup>2</sup>
Revising the existing teaching materials	Between Groups	51.059	2	25.530	83.155	.000	.638
	Within Groups	29.781	97	.307			
	Total	80.840	99				
Focusing on reading specifically the reading passages in the textbook	Between Groups	10.920	2	5.460	13.821	.000	.221
	Within Groups	38.320	97	.395			
	Total	49.240	99				
Using teaching materials other than their current textbooks	Between Groups	39.849	2	19.925	24.113	.000	.332
	Within Groups	80.151	97	.826			
	Total	120.000	99				
Concentrating on the grammatical structures provided in textbooks	Between Groups	3.197	2	1.599	5.954	.004	.109
	Within Groups	26.043	97	.268			
	Total	29.240	99				
Putting more emphasis on writing aspects	Between Groups	3.631	2	1.815	4.225	.017	.080
	Within Groups	41.679	97	.430			
	Total	45.310	99				

According to the above results it can be concluded that the qualification variable has had a significant effect on all the three dependent variables. However, it was noted that BE teachers appeared to feel less impact of the new examination on their curriculum planning and instruction, particularly on issues pertinent to exam-based activities and tasks. DT teachers, on the other hand, had higher scores than both BA and BE teachers, and showed more impact due to the new examination. This result may be attributed to the experience factor, as all DT teachers have long years of classroom teaching in schools (see Table 4.21 below), a claim justified by the results reported in the following sub-section.

#### 4.2.4.2 Experience Effect

As an independent variable, teachers' years of teaching experience had four levels (3-6, 7-10, 11-15, and more than 15 years). Using a One Way ANOVA, a number

of significant results were found when assessing the effect of experience on the three main dependent variables.

#### **4.2.4.2.1 Teachers' Teaching Practices, Teaching Methods and Techniques**

Similar to the qualification factor, the results showed a significant effect of “experience” on teachers’ teaching practices, mainly on “teaching according to the test content”,  $F(3,99)=74.53$ ,  $p=0.000$ ,  $\eta^2 = .69$  (3-6=4.00; 7-10=5.00; 11-15=4.11; more than 15 years=4.15); “adopting new teaching methods”,  $F(3,99)=840$ ,  $p=0.000$ ,  $\eta^2 = .96$  (3-6=2.00; 7-10=4.00; 11-15=4.02; more than 15 years=4.00); “using a more communicative approach in teaching”,  $F(3,99)=17.05$ ,  $p=0.000$ ,  $\eta^2 = .34$  (3- =1.35; 7-10=1.00; 11-15=1.79; more than 15 years=1.42); “emphasizing listening and speaking skills”,  $F(3,99)=16.83$ ,  $p=0.000$ ,  $\eta^2 = .34$  (3-6=1.84; 7-10=1.00; 11-5=1.58; more than 15 years=1.52); and “similarity between teachers’ English teaching to grade 9 classes and other grades”,  $F(3,99)=5.75$ ,  $p=0.001$ ,  $\eta^2 = .34$  (3-6=3.46; 7-10=2.71; 11-15=2.88; more than 15 years=2.78).

Although the ANOVA revealed statistical significance, as presented in Table 4.20, the actual difference in mean scores between the groups varied (see appendix 23a). According to the Tukey Post-hoc test (see appendix 23b), the difference lay between teachers of longer years of teaching and teachers of fewer years’ experience. For instance, regarding “adopting new teaching methods” there was a significant difference between 3-6 vs. 7-10 ( $p=0.000$ ), 3-6 vs. 11-15 ( $p=0.000$ ), 3-6 vs. above 15 ( $p=0.000$ ). However, there was no significant difference among teachers of more than seven years of experience. Regarding how teachers’ teaching in low-level grades contrast with their teaching in upper-

level grades, a statistically significant difference appeared again between 3-6 vs. 7-10 ( $p=0.003$ ), 3-6 vs. 11-15 ( $p=0.011$ ), 3-16 vs. above 15 ( $p=0.011$ ).

It can be noted from the eta squared values that there was a 96% effect of experience (which deemed very high) on the way teachers adopted or used new methods in their teaching as a result of exam change. This means that veteran teachers were not similar in their teaching to novice teachers, a result that can be attributed to the different degrees of influence generated by the introduced exam on teachers. This implies that experienced teachers, according to this result, were more influenced by the introduced exam than their novice colleagues. This finding seems to match those reported in the previous section (section 4.2.4.1.1) relating to the same dependent variable. Particularly, DT teachers who have the longer years of experience (as shown in Table 4.20) appeared to be more affected by the exam. These results, in practical terms, seem to be confirmed by the interview data which was elicited from teachers who have many years of experience.



Table 4-20 Results obtained from one-way ANOVA with eta squared values for experience in relation to teachers' teaching practices

	Sum of Squares	df	Mean Square	F	Sig.	eta <sup>2</sup>
Teaching according to the test content	14.104	3	4.701	74.531	.000	
Within Groups	6.056	96	.063			.699
Total	20.160	99				
Adopting new teaching methods	78.019	3	26.006	840.447	.000	
Within Groups	2.971	96	.031			.963
Total	80.990	99				
Using a more communicative approach in teaching	8.565	3	2.855	17.050	.000	
Within Groups	16.075	96	.167			.347
Total	24.640	99				
Emphasizing listening and speaking skills	8.603	3	2.868	16.831	.000	
Within Groups	16.357	96	.170			.344
Total	24.960	99				
Similarity between teachers' teaching to grade 9 and to other grades	8.525	3	2.842	5.751	.001	
Within Groups	47.435	96	.494			.152
Total	55.960	99				

Comparing the four levels of the experience factor with the three levels of the qualification factor tackled in the previous section, it can be noted, as shown in Table 4.21 below, that the BE participant teachers are the only teachers with the least amount of experience ranging from 3-6 years, and DT teachers have the longest years of experience, i.e. above 15 years. Although all the three levels of teacher qualification have affected significantly teachers' teaching practices in classrooms vis-à-vis the new examination, the BE teachers whose qualification considered the best among others (see section 1.3.3 for illustration), seemed the least affected. Assuming that the new introduced exam has generated negative rather than positive consequences in the context, as appeared from the results reported in this chapter thus far, BE teachers are deemed the least negatively

affected. This finding supports our earlier claim, stipulated in sections 1.3.3 and 2.6.1.4, that BE teachers (teachers who have gained better educational and pedagogical practices in their pre-service education) might be differently affected by the administered exam. BE teachers, according to this finding, thus, appeared less affected than their counterparts (i.e. BA and DT teachers) and seemed more fit in their classroom planning and instruction as far as the current English curriculum is concerned.

Thus, it is to argue here that the kind of qualification and the educational backgrounds that teachers have can play an important role in the kind of instruction teachers implement, especially in relation to high-stakes public examinations. This result may have important implications for investigations related to whether teachers of different academic qualifications might react to testing reform and enact curriculum innovation differently, especially in the current context.

Table 4-21 Cross-tabulation: Academic Qualification - Teaching Experience

Academic Qualification - Teaching Experience Cross-tabulation						
		Teaching Experience				Total
		3-6 yrs.	7-10 yrs.	11-15 yrs.	above 15 yrs	
Academic Qualification	Bachelor of Arts	7	21	34	8	70
	Bachelor of Education	19	0	0	0	19
	Diploma of Teaching	0	0	0	11	11
Total		26	21	34	19	100

#### **4.2.4.2.2 Teachers' Classroom Testing Practices and Evaluation, On-going Tests**

As presented in Table 4.22 below and in appendix 24a, there was a statistically significant effect (at the  $p < .05$  level) of “experience” on teachers’ classroom testing practices in relation to the introduced exam. Particularly, significant effects were found in: “familiarizing students with the content and format of the revised exam”,  $F(3,99)=10.80$ ,  $p=0.000$ ,  $\eta^2 = .25$  (3-6=4.00; 7-10=4.00; 11-15=4.14; more than 15 years=4.47); “organizing more focused activities that reflect exam activities”,  $F(3,99)=8.66$ ,  $p=0.000$ ,  $\eta^2 = .21$  (3-6=4.00; 7-10=4.52; 11-15=4.52; more than 15 years=4.36); “using mock exams to review the drills of the revised exam”  $F(3,99)=5.31$ ,  $p=0.000$ ,  $\eta^2 = .14$  (3-6=4.69; 7-10=5.00; 11-15=4.94; more than 15 years=4.68); and “consistency between teachers’ on-going tests and the revised exam”,  $F(3,99)=5.97$ ,  $p=0.00$ ,  $\eta^2 = .25$  (3-6=4.00; 7-10=4.00; 11-15=3.70; more than 15 years=3.73). Despite gaining statistical significance for all sub-dependent variables, the practical significance between the means of the levels (see descriptives table in appendix 24a) as indicated by eta squared values provided in the ANOVA Table 4.21, was quite small, ranging between 16% to 25%. The small percentages accounted for by these variables/effects may lead us to understand that these effects are not of great importance even though there are statistical significances.

However, unlike the interpretations reported in the previous section, the Tukey Post-hoc test (see appendix 24b) showed negligible difference between the four levels/groups for most of variables. But regarding the first mentioned variable i.e. familiarizing students with the new exam, it can be noted that the most

experienced teachers (above 15 years) differ significantly from other group teachers who have less years of experience , 3-6 vs. above 15 ( $p=0.000$ ), 7-10 vs. above 15 ( $p=0.000$ ), 11-15 vs. above 15 ( $p=0.002$ ). This indicates that highly experienced teachers were affected similarly by the introduced exam. Likewise, regarding the second variable, i.e. organizing exam-related activities, the test showed that the least experience teachers (3 – 6 years) acted essentially differently from other groups, 3-6 vs. 7-10 ( $p=0.000$ ), 3-6 vs. 11-15 ( $p=0.000$ ), 3-6 vs. above 15 ( $p=0.030$ ). This implies that novice teachers did not work in a similar way as their veterans colleagues did. Considering that novice teachers are teachers who earn BE qualification, as shown in Table 4.21 above, this finding supports and reflects on the results reported above in sections 4.2.4.1.1 and 4.2.4.1.2 where BE teachers appeared to teach differently—less focus on exam-related activities—in comparison with DT and BA teachers who have the longer years of experience.

Table 4-22 Results obtained from one-way ANOVA with eta squared values for experience in relation to teachers' classroom testing practices

	Sum of Squares	df	Mean Square	F	Sig.	eta <sup>2</sup>
Familiarising students with the exam content and format of the revised exam	3.038	3	1.013	10.802	.000	
Within Groups	9.002	96	.094			.252
Total	12.040	99				
Organising more focused activities that reflect exam content	4.910	3	1.637	8.667	.000	
Within Groups	18.130	96	.189			.213
Total	23.040	99				
using mock exams to review the drills of the revised exam	1.914	3	.638	5.314	.002	
Within Groups	11.526	96	.120			.142
Total	13.440	99				
consistency between teachers' on-going tests and the revised exam	2.007	3	.669	5.978	.001	
Within Groups	10.743	96	.112			.157
Total	12.750	99				

#### **4.2.4.2.3 Teachers' Choice, Selection and Use of Teaching Materials and Curriculum**

As shown in Table 4.23, a statistically significant effect at the  $p < .05$  level for the sub-dependent variables. These included: “revising the existing teaching materials”,  $F(3,99)=407.77$ ,  $p=0.000$ ,  $\eta^2 = .93$  (3-6=2.00; 7-10=4.00; 11-15=3.94; more than 15 years=4.00); “focusing more on reading specifically the reading passages in the textbook”,  $F(3,99)=29.24$ ,  $p=0.000$ ,  $\eta^2 = .48$  (3-6=3.61; 7-10=4.00; 11-15=4.82; more than 15 years=4.24); “concentrating on the grammatical structures provided in the textbooks”,  $F(3,99)=15.45$ ,  $p=0.000$ ,  $\eta^2 = .33$  (3-6=3.92; 7-10=4.00; 11-15=4.64; more than 15 years=4.31); “using teaching materials other than their current textbooks”,  $F(3,99)=18.27$ ,  $p=0.000$ ,  $\eta^2 = .36$  (3-6=1.42; 7-10=2.19; 11-15=2.85; more than 15 years=3.15); and “putting more emphasis on writing aspects”,  $F(3,99)=15.17$ ,  $p=0.000$ ,  $\eta^2 = .32$  (3-6=1.19; 7-10=1.28; 11-15=2.05; more than 15 years=1.84).

Moreover, eta squared values indicated reasonable effects for most variables (ranging from 32% - 48%), and very large (93%) for the first variable “revising the existing teaching materials”. The 93% accounted for this variable suggests that “experience” as a dependent factor had a tremendous effect on teachers reactions to the exam in terms of their selection of the content of current curriculum/textbooks. This result mirrors those obtained when the effect of the qualification factor was measured on the same dependent variable (see section 4.2.4.1.3). Both results are compatible with those reported above in section 4.2.2.

Although a cursory look at the mean scores of the three groups of teachers (see appendix 25a) showed a negligible difference between means (except the first

sub-dependent variable), the Tukey Post-hoc test (see appendix 25b) indicated that the mean score for 3 – 6 years group was significantly different from both 11 – 15 years group and above 15 years group, the two latter of which did not differ significantly from each other. Regarding the first variable “revising the existing teaching materials” the test showed a significant difference between 3 – 6 years group teachers on the one hand and the other three groups on the other. This indicates that, even though all teachers appeared to be influenced by the introduced examination in relation to this dependent variable, novice teachers and veterans were affected and reacted differently. That is to say, washback intensity differs from one teacher to another, the greater the job experience, the greater the difference among teachers. Indeed, this suggests that experienced teachers appeared to be more vulnerably influenced by the exam than inexperienced teachers as far as their selection of the teaching contents and curriculum planning are concerned.

Table 4-23 Results obtained from one-way ANOVA with eta squared values for experience in relation to teachers’ selection of teaching materials and curriculum

		Sum of Squares	df	Mean Square	F	Sig.	eta <sup>2</sup>
Revising the existing teaching materials	Between Groups	74.958	3	24.986	407.770	.000	.927
	Within Groups	5.882	96	.061			
	Total	80.840	99				
To focus on reading specifically the reading passages in the textbook	Between Groups	23.513	3	7.838	29.247	.000	.477
	Within Groups	25.727	96	.268			
	Total	49.240	99				
To concentrate on the grammatical structures provided in textbooks	Between Groups	9.524	3	3.175	15.458	.000	.325
	Within Groups	19.716	96	.205			
	Total	29.240	99				
To use teaching materials other than their current textbooks	Between Groups	43.625	3	14.542	18.278	.000	.363
	Within Groups	76.375	96	.796			
	Total	120.000	99				
To put more emphasis writing aspects	Between Groups	14.577	3	4.859	15.178	.000	.321
	Within Groups	30.733	96	.320			
	Total	45.310	99				

Finally, it is important to mention here that compatible results were found when the age factor was measured for an effect on the same main dependent variables mentioned above. This indicates that a reciprocal relationship exists between age factor and experience factor, an argument predicated by the assumption that the older a teacher is, the more likely they are experienced in a particular job. This seems to apply to Libyan teachers of English because the vast majority of them graduate from universities when they are around twenty-two years old.

#### **4.2.4.3 Grade Effect**

The grade teachers teach was another factor measured for an effect on the dependent variables. It should be noted that teachers teaching these grades are concurrently teaching grade 9, i.e. a teacher is represented twice, one as a grade 9 teacher and again as a teacher of grade 5, 6, 7, or 8.

##### **4.2.4.3.1 Teachers' teaching practices, teaching methods and techniques**

As presented in Table 4.24 below, there was a statistically significant effect ( $p < .05$ ) of the grade factor on sub-dependent variables related to teachers' classroom teaching practices. These include: "adopting new teaching methods",  $F(3,99)=8.19$ ,  $p=0.000$ ,  $\eta^2 = .20$  ( $G5=4.00$ ;  $G6=2.93$ ;  $G7=3.08$ ;  $G8=3.59$ ); "emphasizing listening and speaking skills",  $F(3,99)=4.15$ ,  $p=0.008$ ,  $\eta^2 = .11$  ( $G5=1.41$ ;  $G6=1.60$ ;  $G7=1.79$ ;  $G8=1.37$ ); "similarities between teachers' teaching to grade 9 classes and other grades",  $F(3,99)=314.59$ ,  $p=0.000$ ,  $\eta^2 = .91$  ( $G5=2.00$ ;  $G6=3.33$ ;  $G7=3.91$ ;  $G8=3.00$ ). From the mean scores (see appendix 26a), it appeared that there was a statistically significant difference somewhere among the groups/grades across the three sub-dependent variables. Particularly, the Tukey

Post-hoc test as shown in appendix 26b unveiled significant difference between groups/levels with regard to the two variables: “adopting new teaching methods” and “similarity between teaching grade 9 and other grades”. For the former, grade five group was significantly different from both grade six group ( $p=0.001$ ) and grade seven group ( $p=0.001$ ). However, as for the latter, all groups differ from each other significantly ( $p=0.000$ ).

Eta squared values provided in the ANOVA Table 4.24 showed varied results. While the effect size was quite small for the first and second variables (.203 and .114), it was very large (.907) with regard to the third variable, i.e. “similarity between teaching grade 9 classes and other grades”. The latter result indicates that there was a greater magnitude of differences between the means of the levels/grades as shown at the bottom of appendix 26a. This means that there was a 91% effect of the grade factor on teachers’ teaching in relation to the introduced exam. That is, grade 9 teachers applied different methodologies when they teach lower-level classes, i.e. their teaching to lower grades is not the same as their teaching to other upper grades. Practically, this result is supported by and, thereby, cross-referenced to the results reported in section 4.2.2.1.1., under Table 4.8, TQ27.



Table 4-24 Results obtained from one-way ANOVA with eta squared values for grades effect in relation to teachers' classroom teaching practices

	Sum of Squares	df	Mean Square	F	Sig.	eta <sup>2</sup>
Adopting new teaching methods						
Between Groups	16.505	3	5.502	8.190	.000	
Within Groups	64.485	96	.672			.203
Total	80.990	99				
Emphasizing listening and speaking skills						
Between Groups	2.867	3	.956	4.153	.008	
Within Groups	22.093	96	.230			.114
Total	24.960	99				
Similarity between teachers' teaching to grade 9 and other grades						
Between Groups	50.793	3	16.931	314.59	.000	
Within Groups	5.167	96	.054			.907
Total	55.960	99				

#### 4.2.4.3.2 Teachers' Classroom Testing Practices and Evaluation, On-going Tests

Similar to the above findings, there was a significant effect of the grade factor on the way teachers implement their classroom testing practices vis-à-vis the new exam. As presented in Table 4.25, these included: “familiarising students with the content and format of the revised exam”,  $F(3,99)=2.94$ ,  $p=0.037$ ,  $\eta^2 = .08$  ( $G5=4.27$ ;  $G6=4.13$ ;  $G7=4.00$ ;  $G8=4.12$ ); and “using mock exams to review the drills of the revised exam”,  $F(3,99)=3.72$ ,  $p=0.014$ ,  $\eta^2 = .10$  ( $G5=3.89$ ;  $G6=3.73$ ;  $G7=4.00$ ;  $G8=4.00$ ).

Despite gaining statistical significance, eta squared values, as shown in Table 4.25, did not indicate practical significant difference between the means of the levels, i.e. small effect size (see also appendix 27a). The very low percentages 8% and 10% accounted for by the two variables/effects may lead us to understand that these effects are not of great importance even though there were statistical significances. Also, the Tukey Post-hoc test (see appendix 27b) appeared to indicate that there were no significant differences occurred between lower and

upper grade classes, though negligible differences between certain groups/levels occurred. The difference only lay between grade five teachers and grade seven teachers for the first variable, whereas for the second variable the difference occurred only between grade six on the one hand and both grade seven and eight. Thus, it can be concluded that grade 9 teachers were considerably affected by the impact of the introduced exam in relation to their classroom testing practices whatever other grade(s) they teach. This finding seems to be incongruent with the results of the previous section where grade 9 teachers appeared to implement different teaching methodologies hinging upon the grade(s) they teach.

Table 4-25 Results obtained from one-way ANOVA with eta squared values for the grade factor in relation to teachers' classroom testing practices

		Sum of Squares	df	Mean Square	F	Sig.	eta <sup>2</sup>
Familiarising students with the exam content and format of the revised exam	Between Groups	1.014	3	.338	2.941	.037	.084
	Within Groups	11.026	96	.115			
	Total	12.040	99				
	Within Groups	11.758	96	.122			.104
Using mock exams to review the drill of the revised exam	Total	13.440	99				
	Between Groups	.887	3	.296	3.724	.014	
	Within Groups	7.623	96	.079			
	Total	8.510	99				
	Within Groups	32.769	96	.341			
	Total	62.750	99				

#### 4.2.4.3.3 Teachers' Choice, Selection and Use of Teaching Materials

Regarding whether teachers revised the existing teaching materials in relation to the grade(s) they teach, the one-way ANOVA as presented in Table 4.26 showed a statistically significant effect ( $p < .05$ ) of the grade factor on this dependent variable. The result was:  $F(3,99)=6.89$ ,  $p=0.000$ ,  $\eta^2 = .18$  ( $G5=2.39$ ;  $G6=2.93$ ;  $G7=3.08$ ;  $G8=3.56$ ). Despite obtaining statistical significance, the effect size was quite small (0.177). Indeed, this means that there existed a low/small proportion of

variance of the dependent variable (teaching materials) that is explained by the independent variable (grade). That is, the effect (i.e. 18%) was not high enough to warrant a practical impact of the independent variable on the dependent variable.

Table 4-26 Results obtained from one-way ANOVA with eta squared values for the grade factor in relation to teachers' selection of teaching materials and curriculum use.

		Sum of Squares	df	Mean Square	F	Sig.	eta <sup>2</sup>
Revising the existence of teaching materials	Between Groups	14.336	3	4.779	6.898	.000	.177
	Within Groups	66.504	96	.693			
	Total	80.840	99				

However, the mean scores presented in Table 4.27 below demonstrate that there was a difference somewhere among the four levels/groups. Particularly, post-hoc comparisons (see appendix 28) identified that difference. It indicated that G5 group was significantly different from all other groups at the  $p < .005$  level, whilst other groups (G6, G7 and G8) did not significantly differ from each other. Hence, the statistics imply that grade 9 teachers appeared to be less impacted by the exam in their teaching to lower-level grade (particularly grade 5) than when they teach upper-level grades. It can be concluded thus that grade factor does have a direct effect on teachers' selection of the content of teaching and curriculum use as far as the revised exam is concerned.

Table 4-27 The descriptives generated from the one-way ANOVA for grade effect in relation to teachers' selection of teaching materials and curriculum use

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Min.	Max.
						Lower Bound	Upper Bound		
Revising the existence teaching materials	Grade 5	29	2.3910	.45756	.08497	3.7570	4.1051	2.00	5.00
	Grade 6	15	2.9333	1.03280	.26667	2.3614	3.5053	2.00	4.00
	Grade 7	24	3.0833	1.01795	.20779	2.6535	3.5132	2.00	4.00
	Grade 8	32	3.5625	.84003	.14850	3.2596	3.8654	2.00	4.00
	Total	100	3.4600	.90364	.09036	3.2807	3.6393	2.00	5.00

#### 4.2.4.4 Classroom Size Effect

An independent samples t-test was run to test the effect of class size on the DVs. The class size is a nominal variable with two levels (20-30 and 31-40); although another category was provided (above 40 students), none of teachers registered a class of this size, the dependent variables are considered of an interval scale.

The outputs generated by running the t-tests, as presented in appendices 29a and 29b, indicated that there was no statistically significant effect of classroom size factor on most of teachers' instructional practices, particularly on the first dependent variable "teachers' teaching practices, teaching methods and techniques". The mean scores of both groups, as provided in appendix 29a and depicted in the beige colour, were to a great degree close. This indicates that teachers for these variables were acting towards the exam similarly in whatever setting—small or big classes—they were involved.

Nonetheless, regarding the second dependent variable "teachers' classroom testing practices" which is shown in the blue colour in appendix 29a, the class size

appeared to have a significant effect on some sub-variables. These included “organising more focused activities that reflect exam content” and “using mock exams to review the drills of the revised exam”. As presented in appendices 29a and 29c, the mean scores of the smaller classes group and the larger classes group were quite different. The smaller sized classes generally showed a lower score, whereas the larger sized classes showed a higher score.

As indicated in Table 4.28, the significance level is less than .05, meaning the variances for the two class groups were not the same, so the information in the second line of the t-test table was used (Pallant, 2010). The t values for both mentioned variables are -2.530 and -2.624, and the corresponding p values are .018 and .019, respectively. On the premises that  $p < .05$  and 0 is included in the 95% confidence interval of the difference, the null hypothesis can be rejected that there would be a difference between teachers’ practices in these two groups. This implies that there was a significant difference on these two variables for each of the two groups. That is, teachers in smaller classes were using less exam-oriented activities regarding this particular practice. Although the scores showed a statistically significant difference, Eta squared values indicated that the magnitude of differences in the means was very small for the two sub-dependent variables. ( $\eta^2 = .061$ , and  $\eta^2 = .065$ ). This means that class size had only a 6% effect on the dependent variable.

Furthermore, regarding the dependent variable “teachers’ choice, selection and use of teaching materials and curriculum”, one item appeared to be significantly affected by the class size: “to concentrate on the grammatical structures provided in their textbooks”,  $t(98)=3.19$ ,  $p=0.005$ ,  $\eta^2 = .094$ , again the smaller sized classes

showed a lower score (20-30=3.86) compared to the larger classes (31-40=4.32).

The effect size was very small  $\eta^2 = .094$  (9%), however.

Generally, it could be concluded that a low association existed between class size and the teachers' impact of the administered exam in relation to their teaching methods and techniques; however, teachers who were teaching a class with more students appeared relatively more impacted by the exam on some of their classroom testing practices and curriculum use.

Table 4-28 Results obtained from the Independent Samples Test for some teachers' testing and teaching practices.

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Organising more focused activities that reflect exam content	Equal variances assumed	44.276	.000	-2.004	98	.048	-.26667	.13309	-.53079	-.00255
	Equal variances not assumed			-2.530	24.873	.018	-.26667	.10541	-.48382	-.04952
Using mock exams to review the drills of the revised exam	Equal variances assumed	23.696	.000	-3.716	98	.000	-.36078	.09710	-.55347	-.16810
	Equal variances not assumed			-2.624	15.821	.019	-.36078	.13749	-.65253	-.06904
Concentrating on the grammatical structures provided in textbooks	Equal variances assumed	9.634	.002	-3.177	98	.002	-.46275	.14566	-.75180	-.17369
	Equal variances not assumed			-3.196	19.371	.005	-.46275	.14480	-.76543	-.16006

#### 4.2.4.5 Gender Effect

Teachers' gender was the last factor that was measured for an effect on the dependent variables (comparing the results of male versus female scores). Using

an independent samples t-test, no significant effect of gender ( $p > 0.05$ ) was reported on the teaching practices, teaching methods and techniques vis-à-vis the new examination (see appendix 30b). It can also be noted that gender had neither significant effects on most of teachers' testing practices nor on teachers' selection of the content of teaching materials and curriculum use. Also the Table provided in appendix 30a showed that the mean scores on the dependent variables for both male and female groups were quite close. This means that male and female participants reacted to and were affected similarly by the exam in these particular aspects of classroom instruction.

Nonetheless, when considering the item: "familiarising students with the content and format of the revised exam"—which is related directly to the second main dependent variable, "teachers' classroom testing practices"—a significant effect of gender was noted as provided in Table 4.29 below.

Table 4-29 Results obtained from the Independent Samples Test for some teachers' testing and teaching practices

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Familiarising students with the exam content and format of the revised exam	Equal variances assumed	30.855	.000	-2.960	98	.004	-.22661	.07656	-.37854	-.07468
	Equal variances not assumed			-2.320	31.195	.027	-.22661	.09768	-.42578	-.02744
To use teaching materials other than their current textbooks	Equal variances assumed	10.900	.001	-3.398	98	.001	-.81081	.23861	-1.28432	-.33730
	Equal variances not assumed			-2.980	35.655	.005	-.81081	.27204	-1.36272	-.25890
Writing aspects	Equal variances assumed	1.341	.250	-2.644	98	.010	-.39605	.14977	-.69326	-.09884
	Equal variances not assumed			-2.564	41.494	.014	-.39605	.15447	-.70790	-.08420

Also, as shown below in Table 4.30, male teachers showed a higher agreement score (Male=4.30) compared to the female group (Female=4.01), the results were significant at  $t(98)=2.32$ ,  $p=0.027$ ,  $\eta^2 = .045$ . This implies that males and females differed significantly in terms of their reactions to the exam on this particular variable. However, this difference was not practically essential as indicated by eta squared values; a mere 4.5% of the variance is accounted for by gender.

Similarly, a significant effect was generated by gender regarding the third studied variable: “teachers’ choice, selection and use of teaching materials and curriculum”. As shown in Table 4.30, male teachers showed more agreement (Male=3.00) in recommending “the use of teaching materials other than the current textbooks provided” compared to female teachers (Female=2.18), the outcomes were significant and resulted in  $t(98)=2.98$ ,  $p=0.005$ ,  $\eta^2 = .083$ . Also, male teachers showed more focus (Male=1.92) on the “emphasizing writing aspects” as compared to their female counterparts (Female=1.52), the t-test showed a significant result of  $t(98)=2.64$ ,  $p=0.010$ ,  $\eta^2 = .066$ . In percentage terms, it can be seen that only 8% and 6% of the variance is accounted for by gender for this particular dependent variable.



Table 4-30 Group statistics for gender effect with eta squared values in relation to some of teachers' testing and teaching practices

Gender		N	Mean	Std.	Std. Error Mean	eta <sup>2</sup>
Familiarising students with the exam content and format of the revised exam	Female	74	4.0811	.27482	.03195	.045
	Male	26	4.3077	.47068	.09231	
To use teaching materials other than their current textbooks	Female	74	2.1892	.96053	.11166	.083
	Male	26	3.0000	1.26491	.24807	
Emphasizing writing aspects	Female	74	1.5270	.64581	.07507	.066
	Male	26	1.9231	.68836	.13500	

In summation, the results of this section (i.e. 4.2.4) apparently complement and support the findings previously reported in the previous sections, and substantiate that the introduced exam does exert washback effect on the ways teachers plan and implement their classroom instruction. The findings also corroborate the prevailing assumption that there are some factors other than the exam per se that may intervene in determining the washback of examinations, the most influential of which relate to teacher and context characteristics. Teachers' characteristics examined in this study were: qualification, teaching experience and gender.

As summarized in Table 4.31 below, firstly; qualification factor and experience factor significantly influenced the dependent variables relating to teachers teaching practices, testing practices and use of teaching materials and curriculum. However, the significance varies between the levels. As for the qualification factor, teachers who hold DT showed higher scores than those who hold BA and

BE, suggesting that washback associated with the first qualification was more perceptible than that associated with the other two, although all are significantly affected. Secondly, similarly, based on the analyses presented in this section, the table demonstrates that teachers with more experience appeared more influenced than novice teachers as their scores were higher. This suggests that the greater the job experience, the greater the difference among teachers, and thereby the more associated with washback.

Thirdly, the gender variable, notwithstanding, registered the lowest degree of effect on the dependent variables, suggesting that gender differences as an independent variable has no direct effect on the degree and kind of washback, and that gender, according to this study's findings, cannot be considered as an intervening variable in investigating washback effect from public examinations.

Fourthly, regarding the effect of the context characteristics (class size and grades teachers teach) on the three identified dependent variables, it was found that the grade(s) teachers teach had a significant effect on all dependent variables; however, it differed from one level to another. That is, teachers teaching upper-level grades registered a higher degree of impact than the lower-level grades.

Fifthly, the class size factor, on the other hand, recorded different effects. While there was no significant effect on teaching practices, fairly significant effects were found on some sub-dependent variables relating to teachers' testing practices. But, regarding to teachers' selection of teaching materials and curriculum use, there was a very limited effect of grade factor on this variable. In this respect, teachers

teaching classes with more students appeared to be more affected by the exam than those teaching smaller classes.

The conclusion of this section agrees entirely with the findings reported earlier in this chapter. Also, it corroborates the views posed in Chapter One section 1.3.3 and Chapter Two sections 2.6.1.2, 2.6.1.4 and 2.6.3, that teacher and context characteristics can be considered intervening variables in mediating washback, and may influentially decide the direction and degree of washback effect of the BECE. Thus, there seems to exist a linearly related relationship between the IVs and the main DVs vis-à-vis the washback effect of the studied examination.

Table 4-31 Summary of the effects of the independent variables on each of the main dependent

Independent variable IVs. and their levels	Dependant variables DVs.	Degree of effect of the IVs on DVs	Notes of variance in significance(scores) between levels/IVs	
Qualification (BA – BE – DT)	Teaching practices	Significant	DT teachers higher	BE teachers lower
	Testing practices	Significant	DT teachers higher	
	Teaching materials & curriculum use	Significant	DT teachers higher	
Experience (3 – 6, 7 – 10, 11 – 15, above 15 yrs.)	Teaching practices	Significant	Experienced teachers higher	
	Testing practices	Significant	Experienced teachers higher	
	Teaching materials & curriculum use	Significant	Experienced teachers higher	
Grade(s) (lower grades=5&6 vs. upper grades= 7&8)	Teaching practices	Significant	Upper grades teachers higher	
	Testing practices	Significant	Similar, but uppers slightly higher	
	Teaching materials & curriculum use	Significant	Upper grades teachers higher	
Class size (small classes=20-30 vs.large classes=31-40)	Teaching practices	Insignificant	-	
	Testing practices	Fairly significant	Larger classes higher	
	Teaching materials & curriculum use	Fairly significant	Larger classes higher	
Gender (males vs. females)	Teaching practices	Insignificant	-	
	Testing practices	Negligibly significant	Males higher	
	Teaching materials & curriculum use	Negligibly significant	Males higher	

#### **4.2.5 Documentary Study**

Further to the exam and curriculum review reported in chapter one, this section provides a brief analysis to the content (questions and items) of the new BECE, highlighting its pros and cons in relation to the content of the current curriculum of English. It is important to note here that due to time and space constraints, and considering the research aims and the research questions investigated, a validation study of the new exam (i.e. detailed technical analysis to test items regarding validity and reliability issues) is beyond the scope of this study. This section also casts light on the relationship between teachers' on-going tests and the studies exam in terms of design and content. Discussion in this section along with the curriculum and exam review discussed in sections 1.3.1.2 and 1.3.2.3 lend themselves to be utilized as documentary study, which is adopted in this research as a complementary tool of data collection and analysis.

The exam paper, as shown in appendix 2a, comprises sixty questions/items in total. A perusal of the content of the exam will reveal that the exam is objective testing techniques based, i.e. discrete-point/selected-response items including multiple-choice with 28 items, true-false with 22 items, and matching with 10 items, representing 48%, 36% and 16% of the exam questions respectively. It appears that all exam questions/items are either testing grammar knowledge or information based on the reading texts stipulated in the students' course-book prescribed.

Furthermore, review of the exam sample attached to this thesis and some other samples of the same exam, particularly vis-à-vis the textbooks prescribed,

revealed that the exam suffers from some defects. First, the writing skills are untested in this testing syllabus, students are not asked to produce written English in any form, unlike the old examination. Second, the communicative language testing items are entirely ignored; however, the prescribed teaching syllabus is based on the theory of a communicative approach to language teaching, suggesting that the exam lacks construct validity, as the less consistent a test is with the theory that underlies the course of study, the less a test has construct validity (Heaton, 1998: 161). Third, the exam seems to lack content validity. Henning (1987) defines content validity as being “concerned with whether or not the content of the test is sufficiently *representative* and *comprehensive* for the test to be a valid measure of what it is supposed to measure” (94; original italics). The primary issue here is the extent to which the exam questions/items can adequately represent the content of the current English curriculum. In other words, the question is how a 60-item test, mostly testing reading and grammar objectively, can adequately represent the content of the whole prescribed syllabus, which is described as a good representation of a positive language curriculum programme (Orafi and Borg 2009).

So, although a more communication-oriented curriculum in basic education schools was introduced by the authority of education in 2000, it appears that the new exam had not shifted the focus from testing students' grammar knowledge towards communicative competences. It seems that despite this test being considerably different from the previous one in terms of format and design, they are virtually similar in terms of content. Thus, the content of both the new exam and curriculum reviewed indicate that congruence between the two is tenuous, in

the sense that some aspects, which are deemed pivotal, writing, speaking and listening, are untested in the new examination. This not only violates the objectives of the curriculum prescribed (see 1.3.1.2), but also contravenes one important objective of the exam: “to cover comprehensively and equally all the components and contents of the curriculum”, outlined in section 1.3.2.3.1.

Notwithstanding the new exam seems to rate highly in terms of practicality and usability as it can be completed in a reasonable amount of time and put in an electronic format facilitating its scoreability (Schmitt *et al.*, 2001: 72), the exam sample in appendix 6 reveals some technical problems. For instance, there are spelling mistakes as shown in Q43, with “furion” instead of “furious”. There is also item repetition, for example Q22 and Q34 testing the same thing, and Q9 and Q38 testing similar information.

Also, ambiguity in some items can be observed, as indicated in questions 14, 23 and 39, the latter of which appears to have more than a key. All of this will likely confuse students, and thus cause them to misunderstand what they are required to do. This will likely cause students to lose the marks for these items if the chosen answers are incompatible with the appointed answers inserted into the computer, as the computer programme is designed to accept only one answer for each question/item. Similarly, some exam questions, especially true-false items, tend to be vague in terms of the construct it tests – for instance, “Water boil at 100 °C.”, a true or false question (Q7 in the exam sample). Grammatically, this sentence is incorrect regarding the verb form; but knowledge-wise, based on information taken from the students' course-book, it is correct. Conversely, the item-stem of Q11, a true or false question, asks students “Libya is in the south of Africa”. In

this case, the student might be confused about whether this question tests grammar or knowledge about geography. If it tests grammar, it is true, and if it tests geographical knowledge it is false, as Libya is located in the north of Africa. Here the student's answer can be wrong or might be correct, depending on the exam designer's intention, and the answer inserted into the computer. Interesting, this particular issue was raised, and thus underpinned, by one of the interviewed teachers (see section 4.2.1.2, excerpt from T8 Alghoul).

Moreover, as the exam items reviewed seem to be thrown into a state of disarray, in some cases students might become confused (as mentioned above), specifically as the instructions of the questions do not tell students to provide their answers in terms of grammar rules (i.e. exam questions are not split according to the construct(s) measured); instead, it was included among general information questions assigned for testing reading comprehension. So, this suggests that the exam question in this case does not test what it was intended to test, a drawback which might undermine test validity.

Regarding the time allotted for students to answer the exam, which is three hours, it appears that the exam may take much less time to be completed by students. This claim is empirically supported by Beglar and Hunt (1999) who carefully analysed and validated a revised version of a high-stake national examination in Japan (based on objective-item techniques). The authors found that "less than 35 minutes were required to complete all 72 items, confirming that a large number of this type of item can be completed in a fairly short period of time" (136). Thus, a three-hour examination of this type in our case would likely harm the practicality of the exam discussed above. It should be noted here that, according to

examination board instructions, students/candidates are not allowed to hand over their answer sheets and leave the examination room unless half of the exam time has elapsed, in this case 90 minutes. This may encourage some students to look around the examination room with the intention of cheating, an activity that contradicts one of the new main purposes of the exam outlined in section 1.3.2.3, ‘to minimise the risk of cheating’.

Furthermore, it can be claimed here that answering some exam questions depends largely on luck or the student's knowledge of general information rather than the student's linguistic competence. Most of the exam questions ask students to retrieve information taken directly from texts stipulated in the prescribed course-book as in the case with questions: 8, 15, 16, 18, 22, 23, 46, and many others. Questions 16 and 46, for instance, were designed on the basis of information taken from Unit Five, pages 42 and 39 respectively (see appendix 34 which shows a unit example from the prescribed course-book).

Further, skimming the exam sample brings about a feeling that the candidate in such exams seems to act as if they are solving crossword puzzles, rather than taking a language test (See Q41, and Q44 where all options seem distracters, with no key). In many cases with a purely objective-item exam, especially when students do not know the key answer, they have a 50% chance of being right in true-false questions (Ward, 1981: 31); by analogy, hence, students will have a 25% chance of being right in four-option question items, as is the case in the current exam. As a consequence, students' scores in such exams might not reflect the students' actual levels in English (Currie and Chiramanee, 2010: 485). Students may gain high scores, but be unable to express themselves in real life



situations or even write sentences in English correctly. This was attested above during teachers' and inspectors' interviews (see 4.2.1.2), when they were asked about their students' levels of English, and whether this has been a direct effect of the new examination. For example, one of the teachers said:

“The exam results do not reflect the students' real levels in English. That is, we find some students who succeeded the exam with a very high mark, but he or she cannot express himself or herself in a real life situation or even write a very short well-structured paragraph” (T8, Alghoul).

This was corroborated by most inspectors, claiming that learning the language under the conditions of this exam would hinder students' acquisition of the language; one inspector, for instance, commented: “many students can get a very good mark at this exam, but in reality their language levels are lagged quite behind” (Insp.2, Albukbak). As this was established, it could be argued that the new exam has a low predictive validity as the correlation between students' results and their future performance in the target language is minimised.

Further confirmation of the above review was sought by asking educators (PhD holders graduated recently from some UK universities) who know about the examination and curriculum implementation in Libyan schools, to compare the test's actual content to a statement about that content issued by those educators. This strategy is recommended by Alderson *et al.* (1995), and deployed by Fulcher (1997: 118), who investigated the content validity of an English placement test at the English language institute of the University of Surrey. Two educators, who responded to the request, generally hold negative views about the exam. In brief, they stated that the exam content and the content of the current textbooks seem incongruent. They further reaffirmed that the course of study calls for all four

language skills to be treated equally, whereas the exams described here demand only grammar rules, and more heavily, reading skills which are tested improperly. According to their views the old pattern of examination, despite its limitations, is far better than the new one because the former tested more aspects of language, the most important of which is the skill of writing. As a consequence, they concluded, it seems likely that teachers may adapt the guidelines in curriculum and content to mirror the existing pattern of the examination.

Additionally, comparing the exam sample with the on-going tests that grade 9 teachers write for their students during the school year (appendix 31 contains a sample of these tests), high consistence between the two types of exam can be noted in terms of design and content. Particularly, it was found that teachers copied verbatim some questions from the previous year's final exam, the exam attached to this thesis in appendix 2a. For example, the teacher wrote Q2 in his test which was Q6 in the main exam, and Q3 which was Q26 (also questions 5, 9, 24, 31, 35 and 47 in the teacher's test were identical to the exam's questions 4, 17, 56, 27, 27, 15 and 23 respectively). This supports and confirms teachers' statements during the interviews when they were asked about the kind of relationship between the content and design of their on-going and the content of the final examination, the new exam studied. For instance, one teacher demonstrated:

I usually write and design my own exams similar to the final/new exam. I administer these tests to students as weekly, monthly or periodically tests ... my classroom tests are to a large extent consistent with the new final exam" (T7, Badi).

In view of the above discussion, the researcher might contend that the revised exam can be evaluated as a test of some successes and more failures. Regarding the exam content, in the first instance, heterogeneity can be observed not only between the exam content and its aims (see section 1.3.2.3), but also between the exam content and the content of the current curriculum of English, a drawback which undermined the exam content validity. Also, the exam tends to divide language into discrete points such as grammar and reading comprehension items, and tests them separately, generally in the form of multiple-choice questions. “It has been proved by evidence that students who take multiple-choice tests can significantly increase their scores ‘artificially’” (Alderson *et al.*, 2000: 45) (see also Currie and Chiramanee, 2010: 485). Hence, the test mission seems less authentic, which may lead to inaccurate deductions about students’ language abilities according to test scores, and thus produce low levels of construct validity (Han *et al.* 2004: 18).

Furthermore, it appears that the main reason for the overflow of selected-response items in the revised BECE paper is their convenience for machine grading. This convenience may sacrifice the test validity, which in turn would mislead EFL teaching in Libyan schools, and it may contradict other important objectives of the test. Thus, one may argue here that although the old examinations suffered from some defects, they appear to be more authentic than the new examination pattern in measuring students’ linguistic abilities as they included integrative/construct-response items; besides, the literature has shown the increased pervasive use of constructed response testing items in most recent educational reforms (Bejar, 2011: 319 ). The new exam, conversely, seems to limit the constructs to be

measured because they depend solely on discrete-response testing with deliberate inclusions of selected-response items, a process which may have far-reaching repercussions for language education in the context.

### **4.3 Summary of the Chapter**

The results reported in this chapter were elicited from different sources, triangulating data from teacher questionnaires, teacher and inspector interviews, pilot observations, as well as document analysis. The findings were presented on the basis of four main themes stemmed from the aims and research questions.

They are:

- Teachers' perceptions of and reaction to the new exam;
- Washback on teachers' classroom instructional behaviour;
- Washback on teaching materials and curriculum;
- The degree of washback vis-à-vis teacher and context factors.

Firstly, findings revealed that most teachers and inspectors shared similar views towards the exam. They expressed relatively negative attitudes and opinions towards the new revised exam (especially in terms of content), and the reasons for which it was implemented, revealing a gap between the new policy and teachers' perceptions. It was claimed that the exam was not comprehensive, and lack test validity and practicality. However, the exam was welcomed in terms of format and design. Indeed, results showed that while the exam has achieved some objectives announced by the authority of education, it has failed to accomplish

others. For example, the new policy has helped to prevent impersonation and other forms of cheating; make marking the answer sheets more reliable; and make the results more transparent. On the other hand, it has failed to fulfil important objectives such as covering the components of the syllabus, motivating students to use integrated skills, and providing a real indication of students' levels of achievement regarding the curriculum. Additionally, the new exam has brought about unintended consequences, such as driving down costs, and reducing the number of appeals test-takers were normally lodging with the old examination system.

Moreover, it was also found that the majority of teachers knew what changes had been made for the new exam. Teachers perceived that the exam placed on them extra demands and work such as being more responsible for their students' performances in the exam, organising exam-focused activities and familiarising their students with the content and format of the new exam. Thus, questionnaire data revealed that teachers' beliefs and perceptions served as the foundation for their behaviour, indicating that 'external exams' as well as 'learners' expectations' were important in terms of teachers' classroom instruction.

Secondly, aspects of daily teaching were explored in the context of the new exam. Although 73% of teachers said they adopted new teaching methods by teaching test-taking strategies, this was not clearly identified. However, data elicited from interviews and pilot observations regarding teachers' classroom activities, teacher talk and the medium of instruction indicated that teachers tailored their everyday teaching practices to meet the policy of the new exam, and thus taught to the test, especially when the exam is close. Results also indicated that teachers' actual

classroom practices exhibited traditional teaching characteristics (i.e. the Grammar Translation Method) which violate the principles of the current curriculum, in which adherence to a learner-centred approach is recommended. Additionally, grade 9 teachers' classroom assessment practices – on-going tests – were affected by the new exam as the content of both kinds of test were closely congruent. Teachers coached students on items similar to those that will be on the test, present facsimile test items that teachers construct themselves and present actual test items before the new exam is administered. Nevertheless, this kind of washback appeared less clearly in teachers' practices with other grades, such as grades 5 and 6.

Thirdly, regarding the influence of the new exam on teachers' curriculum planning, washback effect in this particular area was evident. However, the influence was on selection rather than on choice, as teachers overwhelmingly relied on the prescribed textbooks and no special courseware for the new exam had been published. Nevertheless, clear evidence of washback effects from the new exam on the content of the existing textbooks/curriculum was found. Indeed, inspectors' statements attested those of teachers' who overwhelmingly confessed that they revised the content of the current English syllabus, narrowing the curriculum, in order to meet the demands and content of the new exam. Most importantly, the interviews and observations evinced what materials some teachers made use of. Some teachers wrote summaries of units studied and prepared booklets containing all possible/anticipated questions. Few others resorted to use other relevant teaching resources. All of these actions seems to

compromise the quality of the classroom curriculum, and may cause validity problems.

Fourthly, results reported in this chapter have unveiled that not only is the exam potent in affecting stakeholders' actions and reactions, but also there are other factors that mediate and coexist and interact with the exam to make the washback effect from such exams viable. These include teacher and context related factors such as teacher qualification, teaching experience, class size and the grade(s) teachers teach. Interestingly, inferential statistics recorded differences in effects among these independent variables/mediators.

Finally, document analysis to a sample of the introduced exam (BECE), underpinned by interviews, indicated that the disadvantages of this exam outweigh the advantages. It showed that the exam not only has some practical problems, but it revealed a mismatch between the exam content and the content of the current curriculum of English. It does not seem to be a good representative to the curriculum as it ignores testing some of the curriculum important elements such as oral and aural aspects of language and the writing skill as well. Also, it was revealed that the exam is exclusively confined to testing some grammar points and reading comprehension; however, the latter of which is tested improperly. Moreover, document study revealed a very compatible relationship between the administered exam and the on-going tests teachers prepare and administer to their students in terms of design and content.

The findings of this study reported in this chapter showed that identifying the washback effect of the new revised exam was quite complicated. However,

triangulating the different sources, data triangulation (teachers, inspectors and documents) and methodological triangulation (questionnaires, interviews and observations), have contributed to unveil what was happening in the context of the revised exam—what kind of washback the exam has induced and how it operated. The following chapter will synthesise the main findings in relation to the research questions and with reference to the pertinent literature.



## Chapter 5 Discussion

---

### **5.1 Introduction**

This study set out to investigate the influence of a recently reformed high-stakes public examination (the BECE) on teachers' teaching practices, materials and curriculum. Specifically, the aims were:

- to examine teachers' attitudes and perceptions of the reform in relation to their classroom instructional behaviour;
- to explore the extent to which the introduced exam induced washback effect on teachers instructional practices—teaching and testing methods;
- to explore the influence of this exam on teachers' choice and selection of teaching materials and curriculum;

- to investigate how teacher and context characteristics may have an influence on the degree of effect induced by the exam;
- to explore whether the administered exam exerted the intentions—intended washback—as envisaged by its constructors; and
- to find out what kind of washback—positive or negative—this exam induced on EFL classrooms in Libya.

As stated in chapter two, the study tested seven related versions of the 15 hypotheses raised by Alderson and Wall (1993) that a test will influence teaching; a test will influence what teachers teach; a test will influence how teachers teach; tests that have important consequences will have washback, and vice versa; tests will have washback effects on all ... teachers; and tests will have washback effects for ... some teachers, but not for others (120-21). Discussion will be guided by Hughes' (1993) trichotomy: participants, processes and product (see Chapter Two, 2.4.2).

The findings of this study were grounded in data gathered in three phases. The research methods employed were pilot observations of two teachers (phase one), an attitudinal questionnaire completed by 100 teachers (phase two), interviews with eleven teachers and seven inspectors and the Examination Board Directors (phase three), complemented by document analysis.

Discussions of the main findings, in this chapter, will be guided by the research questions, covering the following areas:

- Teachers' perceptions of and reactions to the new exam (attitudinal and perceptual washback);

- Washback on teachers' classrooms instructional behaviour (methodology washback);
- Washback on teaching materials and curriculum (curriculum washback).
- Effects of independent variables on dependent variables vis-à-vis the studied examination (washback intensity).

Discussions on the above areas will contribute to shed light on two other important areas:

- Intended and unintended consequences/washback induced by the exam;
- Positive or negative washback the exam has induced.

The findings have brought to light both compatible and divergent standpoints of the main participants of this study when compared to and linked with the literature reviewed in Chapter Two. The discussions will be sequentially presented in line with the six posed research questions.

## **5.2 Synthesis of the Research Findings**

### **5.2.1 Discussions on Findings to Research Question One**

What is the nature and scope of the washback effect on grade 9 teachers' perceptions of aspects of teaching in relation to the new examination?

The aspects explored were:

- teachers' perceptions of the reasons behind implementing the new exam;
- teachers' attitudes towards the new exam and its quality; and
- teachers' perceived classroom practices in response to the new exam.

Addressing this research question paved the way for answering the second and the third research questions (see sections 5.2.2 and 5.2.3).

Questionnaire results showed that most teachers were aware of the change in the new BECE pattern, though none had taken part in preparatory workshops related to this kind of examination. Teachers' understanding and awareness of the new exam paved the way for inducing washback effect from the introduced exam, as was speculated would happen in Chapter Two (see section 2.6.1). This result confirms findings of other studies (e.g. Alderson and Wall, 1993; Wall, 2005; Chapman and Snyder, 2000) (see 2.6.1.1 and 2.6.1.3).

Most teacher-participants, however, held negative views about the new examination policy. Their perceptions of the reasons for implementing the new examination were to some extent inconsistent with the underlying theories behind the new policy. The incompatibility between teachers' perceptions and policymakers' intentions suggested negative reactions towards the implementation. Most teachers stated that the new examination was of low quality. They criticised the exam for not redressing the balance between its content and that of the existing curriculum. Teachers, generally, stated that the new examination could not be an effective tool to assess students' integrative skills or prepare students for their future education, and nor did it cover the major components of the prescribed teaching syllabus. These negative views were further corroborated by most inspectors' interviews, according to which the exam was considered much more limited; and thus may be described as construct under-representative, in relation to curriculum content.

The results of the interviews in particular suggested that the new BECE showed a minimum standard level of students' competence in English and that it put more pressure on teachers although they were able to cope with it, because the success or failure of their students in the exam reflected on their performance as teachers. Some teachers, however, stated that the exam was not hard enough, stressing the fact that it did not adequately discriminate among students in general. Others considered the old BECE more operational than the new one, even though the former had also been criticised by some educators (see section 1.3.2.2). Indeed, some interviewees suggested recasting the exam to include the oral and aural aspects of language as well as writing skills. Should these suggestions be seriously taken by policy makers, this would have an important implication that washback from public examinations play an integral part in examination reform.

Conversely, some teachers welcomed the new policy in terms of the exam pattern, as automation ensured that the exam papers were marked accurately and the results disseminated quickly and transparently. The exam policy also helped to considerably reduce cheating associated with the administration of the old BECE. Interviews, furthermore, revealed that most teachers liked the exam because more students passed it and it was easier to teach to. Questionnaire responses, however, showed that the new exam placed extra demands and more pressure on some teachers, especially when the exam was imminent—the seasonality of washback (Baily, 1999; Cheng, 2005). These findings support the belief of some authors that there can be washback from examination on teachers' feelings, perceptions and attitudes, and in turn these outcomes may affect what and how teachers conduct their classroom instruction (Wall and Alderson, 1993; Shohamy *et al.*, 1996; Cheng, 2005; Spratt, 2005).

Teachers' attitudes to the new policy in this study seemed quite different from other studies. Wall (2005), Watanabe (2000) and Amengual-Pizarro (2009), for instance, highlighted that teachers had mixed but positive reactions and attitudes to the introduced exams, whilst in this study, teachers also had mixed attitudes, but they were generally negative. The reasons in this case can be ascribed to stakeholders' perceptions of the quality of the test introduced. It may thus be argued that the quality of the introduced exam can have direct effects on determining the nature of washback induced by that exam. Indeed, washback can be predictable and stable; sound tests would likely induce positive washback and poor ones would probably have negative washback.

Regarding teachers' perceived classroom instruction and curriculum use, the majority of teachers perceived that introducing the new exam influenced the way they shaped their classroom practices to reflect its different demands. Although most teachers' questionnaire responses showed that teacher altered their teaching practices by adopting new teaching and testing methods as a result of exam change, this behaviour could not be further elaborated due to limitations of close-ended questionnaires. Interviews and observations provided further details on these issues (see 5.2.2).

In addition, in terms of whether teachers employed test-oriented textbooks and what teaching and learning resources were used, questionnaire responses showed that the overwhelming majority of teachers did not use any test-oriented textbooks, instead, relying exclusively on the content and organization of the current prescribed textbooks, the *English for Libya* series, as the source of their lessons plans and implementations. This may be ascribed to the fact that these textbooks were well prepared and organized, in that they provided the syllabus, in terms of what and how

to teach; besides teachers did not have other alternatives. Nevertheless, washback from the new exam on the selection of teaching content from these textbooks materialised substantially, as most teachers responded that they revised the content of these textbooks to mirror the content of the new examination. This finding was confirmed during the interviews (see 5.2.3)

To conclude, in light of the above discussion, attitudinal/perceptual washback is discernible. This particular finding is similar to those of Cheng *et al.* (2011) discussed in section 2.3.1), so confirming Alderson and Wall's (1993: 120) assumption: "a test will influence attitudes to the content, methods, etc. of teaching ...".

### **5.2.2 Discussions on Findings to Research Question Two**

What is the nature and scope of the washback effect on teachers' classroom instructional behaviours as a result of implementing the new revised examination?

To complement the data from teacher questionnaires, this research question was principally addressed through pilot observations and informative interviews with teachers and inspectors, supplemented by document analysis. Teachers' classroom instructional behaviour refers to practices pertinent to:

- teachers' teaching methods and techniques;
- teachers' classroom activities, teacher talk and medium of instruction;
- and teachers' classroom testing practices

These points are discussed sequentially in the following paragraphs.

Firstly, consistent with questionnaire responses, observations and interviews data revealed that teachers adhered to teacher-centred approaches to language teaching where teachers are dominant in classroom instruction and were reluctant to apply the communicative practices upon which the principles of the current curriculum of English are based. Most teachers, in response to test demands and their students' expectations, appeared to apply traditional teaching methods, putting more emphasis on teaching reading and the grammatical points of language, during which L1 was noticeably used as a medium of instruction. The way teachers taught the reading texts, for instance, was consistent with the way this skill was tested in the new BECE. For example, most teachers made summaries of the reading texts stipulated in the textbooks, unit by unit, on the basis of discrete questions, true-false or multiple-choice items, and they taught grammar similarly. Appendix 32 shows how some teachers exploited this practice in a unit from the course-book prescribed.

Thus, results suggest that the new BECE pattern impacted, to a certain extent, upon the instructional practices teachers carry out in their instruction, tailoring it to the test. However, the impact seems to be more evident in terms of the teaching activities and tasks teachers implement inside classrooms rather than the methodologies they adhere to, suggesting that teachers' teaching methods in classrooms may remain unchanged. This finding seems compatible with that of Cheng (2005) "the way the teachers carried out their teaching remained more or less the same, whether the testing syllabus was the old one or the new one" (246) as well as the findings of other studies (e.g. Wall and Alderson, 1993 and Qi, 2004), which found limited clear-cut evidence of the relationship between tests introduced and teachers' teaching methods. But they, to some extent, contravene other studies' findings (e.g. Watanabe, 1996, Burrows



2004 and Amengual–Pizarro 2009; see 3.2) that teachers changed their teaching methods considerably.

However, considering the incompatible results of previous studies in this regard, one may argue that in this study it is not discernibly evident that the new exam has been the solitary reason behind these practices, suggesting other variables may come into play. According to the results of this research, these may include teacher characteristics and context characteristics: the former including teachers' beliefs, qualification and experience, the latter including insufficient time allotted for instruction, large classes, the grades teachers teach, and students' low levels of English (see section 6.2.4 for details on some of these issues). Although it seems that it all depends on how these variables interact with the test, the test *per se* still, in many cases, remains the overriding variable that does have a direct effect on washback; substantiating that: "tests that have important consequences will have washback" (Alderson and Wall, 1993: 120).

Secondly, teacher-participants reported that although they were not pedagogically self-satisfied with exam-tailored tasks they performed, they were compelled to use Arabic during English lessons as the medium of instruction and talk to the whole class the most. This was also noted during the pilot observations. The reason behind this behaviour, according to teachers, was two-fold: considering students' low levels of English and enabling students pass the final exam. So teachers felt extra pressure from the exam, resulting in "a tension between pedagogical and ethical decisions" (Spratt, 2005: 24) to meet their students' expectations, i.e. doing well on the exam. Consistent with the results of questionnaires, interviews and observations revealed that teachers orientated their classroom activities towards the pattern of the reformed

examination, especially when the exam was imminent . This seasonality has also been noted in other washback studies. For instance, Wall and Alderson (1993), Watanabe (1996) and Shohamy (1996) noted that timing is essential for conducting washback research, finding, as did this study, that as the exam date approaches, classroom instruction becomes substantially intensified. This particular finding corroborates Bailey's (1999) suggestion that seasonality can be an "applicable concept in washback investigations" (40). However, this may require more comprehensive investigation.

As reported by teachers and observed by inspectors, the most frequent activity carried out by teachers was the use of mock exams, especially when the exam approached. Teachers used previous years' tests to review the drills and format of the revised exam, whereas activities such as organising group work or discussion and organising integrated language tasks were the least used or were non-used activities by teachers, even though the curriculum expects students to develop those skills in more learner-centred classrooms (see section 1.3.1.2). As a result, teachers spent too much of the instructional time speaking to the whole class, most of which was dedicated to exam-related activities such as reading texts and grammar work within the various topics which were included in the textbooks. This suggests that students received less practice opportunities on more meaningful constructs such as the oral and aural aspects of the language. Hence, students were demotivated to learn English, which is deemed a detrimental consequence of the new exam on teaching and learning (see section 5.2.6). This finding echoes that of Andrews (1995) who noted teachers spent too much instruction time on exam-related materials, which he considered

“a limiting of focus for teachers and students rather than a broadening of horizons” (80).

Nevertheless, research data presented in the previous chapter (section 4.2.2.1 table 4.8, TQ27, and section 4.2.4.3) revealed that washback effect from the new examination varied from one grade to another. Grade 9 teachers who concomitantly teach grades/levels 5 - 8 acted differently depending on the grade(s) they teach, i.e. the higher the level teachers teach the clearer evidence of washback from the exam can be noted, and vice versa. (section 5.2.4 sheds more light on this particular issue).

Finally, washback of the exam on teachers classroom testing practices was pervasive and more discernible than that on teachers’ teaching methods and techniques. Interpretations of interview data and documentary analysis reported in the previous chapter revealed that teachers made use of mock exams very similar to the new exam. It was found that the sole reason for that was to acclimatise students to the pattern of the new examination in terms of format and content. This result correlates highly with the results of the teacher questionnaire discussed above in section 5.2.1.

Particularly, it was revealed that grade 9 teachers designed their own tests akin to the introduced exam, exploiting them as on-going tests to assess students in classrooms as quizzes, and monthly or periodical tests. Document study (the exam reviewed in Chapter One with the analysis of a sample of the new examination in Chapter Four, section 4.2.5; and reference to some samples of teachers’ on-going tests shown in appendix 31), revealed remarkable homogeneity between the two types of exam. These tests mirrored the reformed examination to a great extent in design and format, length and type of questions/items, weight given to different components/constructs,

and the objectives covered. While teachers in these tests gave a higher weighting to reading comprehension questions and grammatical issues 'cloning' them from the textbooks prescribed, they ignored any form of communicative language testing. Thus, these results recorded evidence of more intensive and perceptible washback regarding teachers' testing practices than appeared on teachers' teaching methodologies.

### 5.2.3 Discussions on Findings to Research Question Three

What is the nature and scope of the washback effect on the content of the curriculum of English, as the change has been only on the syllabus of testing?

Discussion related to this question, by extension to the above, constitutes another form of the *processes of participants* that may be affected by the nature of the exam (Hughes, 1993). Two major issues regarding washback on teaching materials and curriculum explored in this study were related to:

- teaching and learning resources teachers used (i.e. teachers' choice of textbooks and teaching materials); and
- teachers' arrangements and selection of teaching contents from those resources vis-à-vis the content of the current English curriculum.

Regarding whether teachers employed test-oriented textbooks and what teaching and learning resources were used, results evidenced that there were no special textbooks published for the new examination, and the majority of the teachers did not employ test-oriented textbooks as a source of their lesson plans. Instead, their choice relied exclusively on the current textbooks, with only one exception when a teacher added a

commercially-written and test-oriented textbook to his teaching. However, this was not practiced by other teachers nor observed by inspectors during their classroom visits to teachers.

Thus, as teachers did not resort to other external textbooks other than the existing one, washback effect of the exam on the choice of teaching resources was very minimal. This particular finding seems incompatible with those of other studies (Cheng, 2005 and Shohamy *et al.*, 1996) where new textbooks were purposively published nation-wide to meet the content of the recently reformed examinations. In our case, the reason may be that the reform was only in the examination system, while the curriculum remained unchanged. Also publishing and prescribing teaching materials for schools has been the exclusive responsibility of the Authority of Education, and no one is allowed to do so without an officially prior appointed permission from the authorities. Moreover, the current prescribed textbooks play an important role in Libyan schools, being almost the sole source to teachers—and they have no other choices—to deliver the instruction. Also, it is assumed that teachers' full awareness of the content of the new examination made them not to resort to other alternative materials, as they noticed that what was included in the exam was available in the existing textbooks.

Notwithstanding, the effect of the exam on teachers' selection and arrangements of teaching contents from the current curriculum/textbooks was strongly negative. Research data indicated that teachers narrowed the content of the existing teaching materials to mirror the content of the new examination. Some teachers prepared unit and lesson summaries on the basis of questions to reflect the content of the exam (see appendix 32), neglecting other important lessons/contents. A brief perusal of the

content of this material, which was based on Unit Five from the course-book prescribed, and the content of this unit as stipulated in the course-book (see appendix 34) shows how negatively (i.e. narrowing the curriculum) test demands affected teachers' practices. This finding is compatible with those of Watanabe (2000), who noted, as did Tsagari (2009a) and Lam (1995), that teachers were "using a variety of self-made materials" (44). In our case, this practice seemed to induce pervasive washback on the content of the current curriculum as teachers focused on some activities/lessons from some units in the prescribed textbooks, and neglected others that were deemed important in the curriculum. This, as explained by some teachers, was a result of students' resistance to teachers' attempts to implement certain activities. For example, as the writing, oral and aural skills were untested, they were not included in teachers' repertoire of instructional activities. On the other hand, to make them fit the exam requirement, aspects such as reading and grammar were emphasised.

This result supports the researcher's claim raised early in Chapter One regarding the exam studied, that what is tested is taught, which was assumed to lead to a narrowing of the curriculum. Such findings were also recorded by researchers in other contexts (Stecher, 2002; Cheng, 2005; Abu-Alhija, 2007; Choi, 2008) where it was found that the examinations investigated had the effect of restricting the curriculum to only those aspects most likely to appear in the examination.

The findings reflect those of Gorsuch (1999) and Agrawal (2004), who investigated teachers' practices in relation to EFL curriculum reform in two different contexts (see Chapter Two, section 2.5.3). The authors noted that although the reform urges teachers to use communicative and integrated language skills, the examination

emphasised testing the knowledge of vocabulary, grammar and language usage. Consequently, instructors tended to marginalize skills not tested. Thus, this study provides evidence that high-stakes public examinations can restrict the curriculum, substantiating one more further assumption made by Alderson and Wall (1993), namely: “A test will influence what teachers teach” (120).

To conclude, the discussions so far seem to support the prevailing assumption that a reform in an examination system acts as an engine for teachers’ classroom instructional behaviour change. The discussion appears to attest the plausibility of most of the seven versions of Alderson and Wall’s washback hypotheses highlighted at the beginning of this chapter (section 5.1), and earlier in the literature review (section 2.4.1). The discussion substantiates that: “a test will influence teaching”; “a test will influence what teachers teach”; “a test will influence how teachers teach”; “tests will have washback effects on all ... teachers”. And due to the importance and influence of the exam on students’ future education, it was also demonstrated that “tests that have important consequences will have washback”; however, the version “tests that do not have important consequences will have no washback” was not attested as it fell beyond the scope of the research aims and questions.

In addition, the results of this research do not clearly demonstrate that “tests will have washback effects for [...] some teachers, but not for others”; instead the study rather revealed that all teachers who participated in the study were influenced by the administered exam, although the degree of effect—washback intensity—varied from one teacher to another, likely contingent upon some intervening factors (see section 5.2.4). Regarding Hughes’ distinction between participant, process, and product

(see section 2.4.2), the findings also suggest that these trichotomies work inseparably to make washback viably perceptible.

#### **5.2.4 Discussions on Findings to Research Question Four**

To what extent do the teacher and exam context characteristics have an influence on the kind and degree of washback that materialized?

Posing this research question was predicated on the assumption that washback from exam change can be mediated by other intervening variables relating to teacher and context characteristics. Teacher variables included academic qualification, teaching experience and gender, while context characteristics included class size and the grade(s) teachers teach.

Inferential statistics reported and discussed in Chapter Four (section 4.2.4, as summarised in table 4.31), complemented by interview data, showed that these variables, irrespective of gender, have significant effects on the main dependant variables (teachers' teaching practices: teaching methods, techniques and activities; testing practices: classroom on-going tests; and teachers' selection and use of teaching materials and curriculum). Interestingly, however, the degree of effect varied between levels/sub-independent variables within independent variables. For instance, teachers who hold a Diploma of Teaching (DT) were more affected by the exam than colleagues holding Bachelor of Arts (BA) or Bachelor of Education (BE) qualifications. This means that washback associated with the former was more perceptible than that associated with the other two types of qualification.



This result is attributable to the greater experience DT teachers possess: veteran teachers were much more influenced by the exam than novice teachers. This implies that the greater the job experience, the greater the difference among teachers, and hence the more associated with exam impact, and vice versa. Indeed, exam-oriented activities and tasks were less implemented in novice teachers' classes compared with their experienced counterparts.

Unlike experience and qualification factors, results indicated that teacher gender was of no account in mediating the occurrence of washback. That is, it recorded the lowest degree of effect amid other factors, as male and female teachers almost reacted similarly to the revised exam. Thus, it is difficult to isolate test washback from such variables. That is to say, for instance, the longer years of teaching teachers have the clearer direct evidence of washback from an examination reform can be noted, and vice versa.

Regarding context-related factors reported and reported in sections 4.2.4.3 and 4.2.4.4, results indicated that class size and the grade(s) teachers teach had influences on teachers' curriculum planning and instruction in relation to the introduced exam. Class size factor registered varied results: while it showed little account on teachers' teaching methods and curriculum use, class size indicated a fairly significant effect on issues related to teachers' classroom testing practices, with bigger numbers appearing to be more affected by the exam than teachers teaching smaller classes. That is, the bigger the class the more testing practices in line with the revised exam were conducted by teacher.

For the second factor i.e. the ‘grade’, results from different sources indicated that teachers who taught upper-level grades (grades 7 and 8) were more affected than the lower-level-grade teachers in their adherence to the demands of the exam. Indeed, with the lower grades 5 and 6, for instance, they did not normally teach to the test, i.e. washback did not materialize. But with upper grades, such as grade 8 and particularly grade 9, teachers did teach to the test, i.e. washback was evident. That is, some teachers were usually adopting communicative language teaching in their instruction in the lower-level settings, making their classes less test-oriented, while in the upper-level teachers adhered to more focused exam-related methodologies and activities.

Similar findings were echoed by other washback studies reviewed in chapter two, section 2.6.3. For instance, Shohamy *et al.*, 1996 found that teachers who taught the grades to which the exam was introduced were more perceptibly influenced by the exam than those who were teaching lower-level grades to which the exam was not given in the same school. It should be noted that, in those studies the teachers were different teachers teaching different grades, but in our case, in this study, participants were the same teachers teaching different grades, providing more solid evidence of exploring washback in this particular issue. This finding suggests that washback effect from public examinations can sometimes be indiscernible, supporting, to a certain degree, the claim that the relationship between examinations and instruction is a “ripple effect” (Barnes *et al.*, 2000: 632).

In the Libyan context, “the ripple effect” metaphor would mean the pebble thrown into the water at one point represents the new BECE pattern, while the ripples suggest its consequences on instruction and curriculum in grade 9 classes, and across other classes as well. However, as the ripples spread across the surface of the water, they

seem to gradually become weaker, until there is little or even no ripple. Accordingly, in our case, lower-level grades such as grades 5 and 6 were less affected by the new exam than the upper-level grades, mainly grade 9, even though they were taught by the same teachers (it should be noted that grades 5 and 6 sit the same exam after three or four years, i.e. when they get to grade 9). Although this proves that tests can dictate what and how teachers teach, but to different degrees, teachers may acclimatise and alter their instruction according to the situations and circumstances they are involved in. This finding seems to be inconsistent with the proposition raised by Alderson and Wall (1993) that “tests will have washback on all learners and teachers” (121), but conversely, the finding is in strict conformity with the other view that “tests will have washback effects for **some** ... teachers, but **not** for others” (121) (bold in original).

In summation, these findings support the view that “washback is not easy to predict or control, and the shape it assumes is influenced not only by tests but by the interaction of numerous factors” (Wall, 2012: 83). So, it is suggested that the kind and degree of washback is inextricably associated with other factors besides the exam *per se*; particularly, it appeared that teachers’ factors seemed to outweigh the other factors including the introduced exam. Indeed, teachers with more teaching experience have been able to alter and modify their instructional and curriculum planning to meet the requirements of the introduced exam. The academic qualification and the educational background teachers possess also appeared to be important in determining washback.

This study’s findings in this regard seem to be in line with findings of other washback studies reviewed in chapter two, sections 2.5 and 2.6 (e.g. Fish, 1988;

Shohamy *et al.*, 1996; Lam, 1995; Watanabe, 1996; Cheng, 2005). Watanabe (1996), for instance, found that the exam studied in Japan did not affect teachers in the same way, proposing external factors, such as teachers' educational background, experiences, beliefs and the timing of the research, played a role. This study, thus, supports some of these suggestions empirically.

Having addressing the above four questions, the answer to these questions contribute towards answering two further questions, question five and question six.

### **5.2.5 Discussions on Findings to Research Question Five**

To what extent did the new revised exam exert the intended washback envisaged by its constructors, and did it bring about any other unintended consequences?

Early in Chapter One of this thesis (1.3.2.3), it was reported that, according to its designers, the intended washback of introducing the exam was: to facilitate the way in which candidates answer the exam questions; to cover, comprehensively and equally, all the components and contents of the curriculum; to provide students with a gauge of their language learning achievement as far as the material of the prescribed syllabus is concerned; to minimize the risk of cheating; and to score the answer sheets mechanically, and hence to disseminate the results quickly, accurately and as transparently as possible. Research question number five attempted to find out whether the announced intentions/objectives were achieved, and whether there were any other hidden intentions or unintended consequences. This, in relation to the present study, was described by Hughes (1993) as the *product*, i.e. outcomes.

Results showed that while the exam succeeded in achieving some intentions, it failed in others. The exam succeeded in facilitating the way students answer the exam questions, in that it was easier for students to deal with this exam than the old examination. This was reflected by the good marks and the results students gained in comparison with their scores in the previous examination system as mentioned by teachers during the interviews, and indicated by the figures shown in appendix 17, bringing about unintended but positive washback (see section 5.2.6).

Also, further to the better security measures and identification checks carried out by examination boards and committees, the exam succeeded to diminish the previously rampant phenomenon of cheating, and helped to prevent impersonation (see section 1.3.2.4). Scoring the exam answer sheets by automation helped make the marking of exam papers more reliable, and hence, the results were accurately, transparently and swiftly disseminated, thus accomplishing another aim of the exam constructors.

On the other hand, the exam did not meet its intended outcomes in relation to the following important objectives: to comprehensively cover equally the major components and contents of the curriculum, and to provide students with a gauge of their language learning achievement vis-à-vis the content of curriculum. The more general point being made is that although the intention was to encourage a more communicative approach and teaching a broader range of skills, the test has not had this effect. Wall (2012), as discussed in section 2.2.1 raised an important concern: “why the test has *not* produced the desired washback” (87) (*italics in original*). In our case, the answer lies in several reasons. The first is that there seems to be a gap between the content of the exam on the one hand and the content of the curriculum and the exam aims and purposes on the other. Indeed, the exam did not take into

account all or even most of the components of the prescribed syllabus. It focused on the discrete points of the content of teaching, such as grammar points and reading texts, through discrete testing items, and neglected the communicative aspects of language not evaluated by the exam, such as the practice of oral interaction in classrooms, and equally important, the assessing of writing aspects.

Furthermore, according to questionnaires and interviews, research results revealed a mismatch between students' test scores and their English proficiency (this was one of the reasons why most questionnaire respondents expressed negative views of the test). This, in turn, would likely lead to inaccurate inferences about students' language abilities according to test scores, and thus produce low levels of construct validity (i.e. negative washback see 5.2.6 below), another answer to Wall's above question. This failure, consequently, prevented the exam from achieving another important aim "to provide students with a gauge of their language learning achievement as far as the material of the prescribed syllabus is concerned". So, as advised by Wall (2012), such understanding of these issues/shortcomings informs exam designers when they reconstruct or improve such exams.

Moreover, the exam brought about unintended consequences/outcomes that may be deemed advantageous: firstly, the noticeably positive increase in students' final results compared to their results with the old examinations as shown in appendix 17. This may support interviewees' claims that the new examination is easier for students than the old ones. As noted by many teachers and inspectors, however, this increase in scores was not concomitant with an increase in students' language proficiency. Secondly, administering the new pattern of the BECE helped to drive down cost, in that the large amounts of money previously paid for marking exam papers, and

preparing and disseminating the results, have been reduced considerably. Although this was not admitted by the examination authorities, some teachers and inspectors in the interviews contended that this was one underlying aim of the new examination, a claim neither rejected nor confirmed by the Examination Board Director during the interview. This finding adds to previous studies (for example, Toch, 2006: 14-16; Haggie, 2008: 4), that report that high-stakes language tests are sometimes automated, especially in terms of scoring, in order to reduce the cost, despite their limitations.

Another further favourable ‘unannounced’ outcome from implementing the new BECE is that it reduced the number of appeals some candidates would lodge to scrutinise their marks/results under the conditions of the old examination patterns (these appeals were very annoying and disturbing to the examination authority), since marking and preparation and dissemination of results are undertaken by computers.

### **5.2.6 Discussions on Findings to Research Question Six**

What quality of washback effect—negative or positive—did the new revised exam induce on EFL classrooms in Libyan schools?

Discussion in Chapter Two highlighted that washback from examinations, especially when they undergo changes, can be negative or positive, an issue addressed by the sixth research question in this study. This was addressed by triangulating the results obtained from the three phases of the research study, mainly by combining the answers of the five previously discussed research questions; comparing and linking them to the pertinent literature (see section 2.3).

The results support the researcher's claim made in Chapter Four, section 4.5, that the new administered examination tended to be a test of some successes and more failures. These failures are associated particularly with issues related to the process of teaching and learning. An analysis of a sample exam revealed confounding of content and language in the exam, though it indicated that the exam tended to measure content area knowledge rather than language proficiency. This shortcoming casts serious doubt on the legitimacy of using the exam as an assessment for EFL learners as far as the content and principles of the current English curriculum are concerned.

The findings indicate that while textbooks homogeneously correspond to the curriculum, the new BECE does not, and that there is a harmful washback effect of this exam on EFL teaching and learning in the context. This, as noted by Au (2011: 30), will likely cause the exam to have a net effect of standardizing and pre-packaging the content of the curriculum in teachers' classroom practices. Indeed, as discussed in the previous sections, teachers taught to the test and narrowed the content of the curriculum, restricting their instruction to the most frequently occurring elements of language in the exam. This negative practice is conceived as a degree of cheating, though usually justifiable (Amrein-Beardsley *et al.*, 2010). In effect, the new BECE deleteriously compelled most teachers to follow the "hidden syllabus", and hence, altered the curriculum in a negative way in that teachers instrumentally ignored teaching the communicative aspects of language because they are not directly related to passing the exam, even though these aspects and activities are highly recommended by the curriculum.

Teachers were found to apply teacher-centred approaches instead of learner-centred approaches of instruction, adopting the 'spoon-feeding' methods, in which students



were confined to remain at their desks listening passively for the whole school year. This behaviour, undesirably, has a negative impact on the development of students' cognitive and communicative abilities. This, in turn, may have serious repercussions for the future of EFL classrooms in Libya, an effect that contradicts the intentions, objectives and prospects of the policy makers (and also all concerned in the society at large), one of which is to "enable students to acquire the appropriate knowledge of skills and positive attitudes and cultural and social values appropriate to the needs of the student, and the needs and aspirations of the society" (ME, 2008: 4).

One further negative effect of the exam on the context is that it induced *test score pollution* (Haladyna *et al.*, 1991: 4). That is to say, since the exam being orientated towards more discrete-point testing rather than integrative testing, students would not acquire real-life knowledge, but instead learn the tested discrete points of knowledge, and their scores would only reflect the quality of language performance on relevant tasks of the exam. Consequently, students' scores can be considered invalid measurements in mirroring students' real levels in using language communicatively as recommended by the existing curriculum. In effect, students may gain high marks in the exam but their real levels of language lag behind, and they are unable to express themselves in real-life situations, or even produce any type of writing.

It can be concluded that the new BECE induced a harmful washback in EFL classrooms in Libyan schools, rather than having positive washback. In other words, the negative effects of the new BECE seemed to outweigh the positive effects as far as promoting the communicative and cognitive competence of students is concerned, the achievement of which is contingent upon what teachers perform in their daily teaching and the type of exam students sit.

## Chapter 6 Conclusions

---

This chapter outlines the implications of the study and its contributions to wider knowledge, its limitations, and presents some recommendations and suggestions for further studies.

### **6.1 Recapitulation**

This study was conducted in state schools in the city of Misrata, Libya, to explore the washback of a revised high-stakes public examination on teachers' teaching practices, materials and curriculum use. After developing an appropriate conceptual framework, the study deployed a mixed research design using both quantitative and qualitative approaches. Empirically, the study was conducted in three main phases: Phase One in which a pilot study was carried out to gain a general picture of the studied phenomenon and, importantly, to trial the research methods/tools designed

for the main study. The Second Phase, the perception stage, involved questionnaires to teachers to elicit their perceptions and attitudes towards the introduced exam in relation to their classroom instruction and curriculum planning. Phase Three, the implementation stage, was undertaken to gather more in-depth data to supplement and complement the questionnaire data by means of interviews with teachers, inspectors and the exam board director, as well as a document study.

The study demonstrated, first, how teachers perceived the new exam (i.e. perception washback). Then, it explored the effects of the exam on teachers' classroom behaviour, teaching and testing practices (i.e. methodology washback) and teachers' selection, choice and use of teaching materials (i.e. curriculum washback). The study showed how teacher and context-related factors may influence the degree of washback induced by the exam (washback intensity), and whether this washback was positive or negative. In addition, it explained which of the intentions of the exam constructors, have or have not been met, and showed other unintended consequences emerged because of the exam.

This study showed findings mainly concordant with the recent literature, though, discrepant findings support the idea that washback is quite context-oriented and complex. For instance, teachers expressed negative rather than positive views towards the new exam; despite its advantages (such as ensuring accurate, quick and transparent results), most teachers surveyed perceived the exam to be of low quality, specifically in terms of the mismatch between test scores and students' English attainment vis-à-vis the requirements prescribed by the English curriculum. This was supported by document analysis of both the content of the exam and content of the current curriculum.

The study indicated that teachers' perceptions of the new policy have been reflected along in practices, and findings clearly evidenced that the present EFL examination came to dominate classroom work. It caused grade 9 teachers to change their instruction planning and curriculum use, especially as the exam date approached (i.e. seasonality in washback). These effects were more discernible in curriculum washback rather than methodology washback, particularly in teachers' selection of the content of the teaching material rather than on teachers' choice of textbooks, as most of teachers tended to rely on the "hidden syllabus" (test papers, model questions, summaries of some units/sections, etc.). Indeed, teachers tended to skip certain topics in the textbook because they were less likely to be tested in the examination. Additionally, teachers were heavily influenced by the new examination in terms of their classroom assessment practices in that they designed their on-going classroom tests to be consistent with the content and format of the new BECE. Teachers' teaching methods and techniques were also, to a certain degree, affected by the new exam, as they were in many cases found to be teaching to the test.

The study substantiates the earlier findings of Wall and Alderson (1993) and Cheng (1997), notably that the washback effect can occur at the level of content rather than at the level of methodology. Thus findings further corroborates the claim that washback is neither simple nor direct, but circuitous and complex (Wall and Alderson, 1993). The current author, therefore, suggests some areas need further research, especially with reference to the studied exam (see section 7.4 below).

The study supports the prevailing assumption that a change in evaluation can leverage changes in classroom instruction. It suggests that tests could be a good lever for change, in that stakeholders will plan their instruction around the language

aspects that most frequently occur in the introduced exam. The findings of this study, by contrast, indicated that the introduced exam may seriously alter the future of EFL teaching in Libya, in that its effects are, by and large, unfavourable, because teachers concentrated on practicing test-related tasks which narrowed the curriculum, rather than developing students' communicative abilities (aural and oral skills) as recommended by the prescribed syllabi in Libyan schools.

This study shows, however, that there are some variables other than the exam *per se* that mediate the existence and degree of washback, some of which pertinent to teacher characteristics—experience and academic qualification—and some context characteristics relating to students' expectations, class size and grades taught.

## **6.2 Implications and Contributions**

The implications and contributions of the study are presented here under two areas: local context and the relevant literature.

### **6.2.1 Local Context**

This study has not only produced evidence of what and how Libyan EFL teachers perceived the new exam in terms of its format and content in relation to their classroom practices, but it has also led to some ideas on how this exam can be improved: for instance, some teachers and inspectors suggested students' English speaking skills should be assessed in a new BECE, to drive improvement in the ability of students to speak English. Thus, the findings of this study affirm the role public examinations have in shaping EFL classroom practices, particularly in Libyan

schools (and potentially in other contexts), which will likely inform school and English examination system for future reforms.

This appears to be the first study which examines the influence of the new BECE on EFL teaching and learning in Libya. So, pedagogically speaking, findings of this study may have important implications for curriculum implementation and prospective assessment changes. At several points throughout this study, a mismatch between the principles of the English curriculum and teachers' classroom instruction planning was identified. In other words, teachers generally adhered to more traditional approaches to language teaching, making their teaching practices at odds with the aims of the communicative syllabus prescribed by the education authority. This mismatch tended to be ascribed to the constraints imposed by the format and content of examination change; though, this relationship sometimes may not be so straightforward. This was echoed by some recent studies which have noted that Libyan English teachers failed to implement communicative language teaching at classroom levels because some of them lack the required degree of competence, especially in the oral and aural aspects of the language (Ali, 2008; Orafi and Borg, 2009; Shihiba, 2011; Elabbar, 2011).

Nevertheless, it should be noted that all teachers in the questionnaire sample recognized the importance of assessing students' ability in the four language skills. Interviews also indicated that if changes were made to the BECE (test reform), by including the speaking skill, teachers would be in a better position to change not only the content of the lessons but also the methods of teaching. Thus, certainly, it would appear that in the case of some teachers, a more communicative approach to language testing would encourage a correspondingly communicative approach to

language teaching. Hence the researcher suggests including communicative aspects in the current examination to promote communicative language teaching and learning in Libyan EFL classrooms. Nevertheless, admitting the lack of linguistic and communicative competence among many teachers, this proposed reform should occur in combination with both theoretically and practically focused in-service teacher education and comprehensively intensive in-service teacher training programmes designed to boost familiarity with communicative teaching and testing methodologies. This would aid a shift to learner-centered communicative language from teacher-fronted approaches.

In addition to the several implications for the EFL teaching and testing in Libya, this study has implications and contributes to language education in general and washback literature in particular. The following section sheds light on these issues.

### **6.2.2 Relevant Literature**

This study has important implications for and adds fresh insights to the body of the pertinent literature describing how exam conditions influence teachers' instructional planning and implementation. These include: a) exam proximity—the closer the test time is, the more its stakeholders are affected by it; b) the level of stakes of an exam—the higher the stakes of a test, the more likely the teachers' perceptions of instruction planning and curriculum use will be affected; c) the purposes and uses of the test scores and the results on which important decisions may be based; d) test's familiarity to teachers and students; and e) the status of the language the exam tests. However, these are not the only ingredients of the recipe, as discussed below.

The literature indicates that positive washback is caused by factors other than the new examination. The intensity and direction of washback may be influenced by mediating variables such as the teacher and the context. As this has not been given ample attention in the literature, this study affords detailed insight into how these factors, such as teacher qualification, experience and the grade(s) they teach play a major role in determining washback. While other factors, such as class size and gender, may be of lesser or no account. This implies that washback from high-stakes tests cannot be considered as an inevitably automatic consequence of a test alone.

Thus, this suggests that there is no way of isolating factors that govern washback, or predicting what kind of washback will be observed in different contexts. In other words, the mediator variables operate with and within washback settings like good seeds and fertilizers work when cultivating a piece of land. The more of these variables that coexist with the implementation of an innovation such as a new or revised examination in an educational system, the more likely washback is to appear. Although, the findings of this research confirm the existence of washback, this particular finding may provide an answer to Fulcher and Davidson's (2012) question: "[...] is washback a theory?" (4). The answer would be: no, it is not a theory, because it does not have predictive power—as it cannot be verifiable or controllable, and definite predictions cannot be made all the time. That is, because most washback studies have been in different contexts, used different methodologies, and so came to rather differing conclusions. However, if we understand which variables are at play in the creation of washback, and their relative strengths within a particular context, we are able at least to try and create the conditions under which favourable washback



might occur. Investigating washback from a high-stakes public examination needs a context tailored-made study in which the exam is administered.

These findings may also have important implications for investigations pertinent to English language curriculum innovation implementation. That is, in addition to examinations, the influences reported in this study, such as teaching experience, academic qualification, educational backgrounds, and class size may have direct or indirect effects on teachers' enactment to these innovative programmes in language education. Thus, the findings of this research may have direct associations with conclusions from investigations conducted on curriculum reforms referred to in sections 2.6.1.1 and 2.6.3 (such as those of Gorsuch, 2000; Agrawal, 2004; Waters and Vilches, 2008; Orafi and Borg, 2009; and Basturkmen, 2012;).

Additionally, this study looked at teachers' concerns, beliefs and perceptions of an EFL examination change, and proved that these have profound effects on their practices. This study highlighted that EFL high-stakes tests are one of the greatest challenges encountered by teachers in classroom situations. It was found that these tests induce changes in the way that teachers alter and perform their curriculum planning and instruction in response to the expectations of students and parents. The most important expectation is that the learner will pass the examination that is the gateway to the next level of education.

Methodologically, this study provided empirical evidence for the necessity of employing both qualitative and quantitative approaches in washback research and supported the supposition that the good timing of gathering research data and participants' personal empathy with the researcher, can contribute to successful data

collection (Cohen *et al.*, 2011), where the research is impeded by obstacles or is interrupted due to political or military unrest. It also has important implications for space triangulation in educational research, where the research or the researcher attempts to overcome the parochialism of studies conducted in the same country or within its subculture, by making use of cross-cultural techniques and perspectives.

Furthermore, this study appears to be the first washback study that utilised a survey questionnaire in the area of washback that can be drawn upon in washback studies: the teacher questionnaire was an adaptation of the one designed and used by Cheng (2005) in her washback study in Hong Kong. This study suggests the possibility of using a single uniform questionnaire to survey teachers about language testing washback, keeping in mind the context and the issues investigated, the results of which could be compared across these studies. This seems to fulfil Bailey's (1999: 38) prediction that such a questionnaire would be an important contribution to the available methodological instruments to washback study to develop such a tool for teachers.

This study shows that high-stakes language tests do not necessarily generate the positive washback intended by exam administrators, that unintended consequences may occur. It also contributes to furthering our understanding of the reasons that hinder the desired washback from occurring, and meanwhile induced unintended, harmful washback. Indeed, when a gap exists between the purposes and specifications of the reformed test on the one hand and the actual application of the test on the other, a mismatch is caused between the aims of the prescribed curriculum and its actual implementation.

This study also has produced practical implications: surveying stakeholders to acquire quantitative data about examinations, such as the new BECE, is worthwhile because their responses afford insight into the educational and pedagogical consequences of that test; it demonstrates that collecting qualitative data from stakeholders provides rich and important information about the broad validity of a test and its impact at the micro and macro levels; and it extends our understanding of how public examinations are sometimes intrinsically considered leverage to improve an education and examination system. Although it is difficult to change a test once it is fully operational (Kane, 2002 cited in Winke, 2011: 654), the results of this study cast light on how a high-stakes public examination could be improved by means of washback testing research.

### **6.3 Limitations of the Study**

This study has some limitations that need to be acknowledged. The literature shows that washback can evolve over time (Shohamy *et al.*, 1996; Wall and Horak, 2011), so a longitudinal study would perhaps be better able to capture and monitor the ripple effect of the new BECE. This was not possible in view of the time limit allocated for PhD research. As it was, the data for the main study of this research was collected over a period of five months, and conducting follow-up studies was beyond the practicalities of this study.

Furthermore, as this study was confined to investigating the washback effect of the new BECE on grade 9 teachers' instruction and curriculum planning only when the BECE was initiated, the second limitation is associated with the lack of base-line data, i.e. a "before and after" study (Wall, 2012: 87). This kind of data is deemed

essential in washback studies (Wall, 2005), and entails researchers gathering data before an innovation or a reform takes place to make comparisons and links between the kinds of data elicited later on. But the top-down policy used by the Authority of Education prevented this, as evidenced in the responses to TQ8 where all teachers stated that the time between the announcement of replacing the old BECE and the implementation of the new one ranged between one to three months. Thus, data about teachers' classroom instructional behaviour before introducing the new examination, for this study, was largely contingent upon teachers' anecdotes and inspectors' statements, and also on previous language education studies conducted in the same context reviewed in Chapter One. However, importantly, the data obtained in this research could be utilised as base-line data for any future washback study on the new BECE.

The third limitation is from a methodological point of view. Although pilot observations were utilised along with the data collected for the main study, this research would have benefited more from classroom data based on more regular observation visits to the same teachers participated in the main study. But it was not possible to collect such data, due to the difficult circumstances encountered by the researcher during fieldwork (see 3.6 – 3.8). Nonetheless, this shortcoming was, to some extent, compensated for through interviews with some EFL inspectors who regularly supervise and observe the work of the same participant teachers in their classes. This study, however, is one of the few washback studies that has used both quantitative and qualitative data to study exam impact.

Finally, another limitation is pertinent to the translation of data. Some of the interviews were conducted in Arabic, and then translated into English. Despite the

fact that the researcher paid considerable attention to the translations, and asked one of his Libyan colleagues (a TESOL doctoral student at a UK university) to check the translated data, the process is not without its drawbacks. Translations, however, were compared and found to be highly consistent, and I am confident that the translated data faithfully captured the meanings that interviewees expressed.

#### **6.4 Recommendations and Suggestions for Future Research**

The conclusions and limitations of this study bring forth some fruitful and interesting possible avenues for further directions and future research in relation to the themes of the study. Researchers in the realm have previously called for overall coherence of washback studies. Cheng (2008: 360) suggested: “[i]t would be the best use of resources if a group of researchers could work collaboratively and cooperatively to carry out a series of studies around the same test within the same educational context”, and “to see research that truly builds on work that has been done previously” (Wall, 2012: 89). In the first instance, what is left to be done in this particular context is to collect and benefit from classroom observation data in order to add some insights and to make the evidence of washback crystal clear. Also, this first washback study conducted in the context, could be extended by carrying out longitudinal studies on the washback effect of the new BECE. The purpose is two-fold: to explore the degree of stability of the exam impact observed in this study over an extended period of time; and to explore the extent to which the exam impacts not only upon teachers, but also upon other stakeholders, including students, parents and even policy makers. Doing so would enable “to cross-reference their findings to build up a comprehensive picture of the same washback operating therein” (Wall,

2012: 89). Fortunately, this study can be utilized as a baseline, and thus, paves the way for such pertinent future studies to be conducted.

In Chapter One of this thesis, it was mentioned that the Libyan examination system, particularly final exams, has undergone changes across all school subjects, one of which is English language. Two major school-leaving exams were reformed—the BECE and SECE (see 1.3.2.1). Due to constraints of time and cost, this research was limited to studying only the first type of exam. It would therefore be worth replicating this study on the SECE, on different population and classroom settings. Equally important would be the need for replications of the study on different school subjects (e.g. Arabic, Science, IT, Maths, History, etc.) on the same populations. This could provide greater generalizability regarding washback of high-stakes public examinations not only in English language education but also in general education.

One other direction for future research could be a validation study of the new examination in terms of its construct validity and reliability. It is to design a performance-based test paper on the basis of a BECE paper, for the convenience of comparison. The overarching aim is to find any significant differences in the construct validity between the BECE paper and the performance test paper.

Further, if the status quo of the examination system continues to divorce instruction from content, teachers may simply continue to accommodate it to their current and future modes of teaching, a situation that will undoubtedly have serious repercussions for Libyan EFL classrooms. Therefore, due to the pivotal role that sound language tests play in determining the *what* and the *how* of teaching and learning, the researcher recommends that the EFL examination system in Libyan

EFL classrooms should be reconsidered. The proposed reform calls for ambitious shifts in classroom assessment and instruction to make language education more consistent. That is, the administered exam questions/items should not be limited to the measurement of recall or recognition of information and rote learning. Further, the system should endeavour to quantify higher-order outcomes based on understanding and meaningful learning, in order to evaluate students' ability to apply their knowledge, not just in pedagogic and educational settings, but also in real-life contexts. There will be much to gain and little to lose in taking steps forward in the direction of test reform. In essence, the performance of such examinations should be analysed and monitored by expert bodies to feed back into the process of language education and assessment in schools, and contribute to bridging the existing gap between the examination and English curriculum.

In conclusion, since this research study has been sponsored by the Libyan Ministry of Education, this study's implications, suggestions and recommendations will hopefully benefit the education system of the country in general, and EFL education in Libyan schools in particular. That is, it may aid the relevant government authorities to develop appropriate policies and administration for improving the education and examination system, and finding better ways to enhance teachers' performance and capabilities at all education stages. This in turn will likely boost learners' motivation levels to learn the language effectively. I would recommend taking a bottom-up rather than a top-down approach, giving more weight to teachers' as well as inspectors' suggestions and start building future plans from the existing situation instead of following instructions from the top—"today is the best guide to tomorrow" (Cohen and Spillane, 1992:41).

## References

- Abdulali, A.J. (1986) *Changing the role of inspector in the Libyan educational system*, USA: University of Pittsburgh.
- Abu-Alhija, F.N. (2007) 'Large-scale testing: Benefits and Pitfalls', *Studies in Educational Evaluation*, 33, 50-68.
- Agrawal, M. (2004) 'Curricular reform in schools: the importance of evaluation', *Journal of Curriculum Studies*, 36(3), 361-379.
- Akle, F. (2005) *An evaluation of English language secondary school teachers' qualifications*, Unpublished M.A dissertation ed. Tripoli University, Department of Graduate Studies, Tripoli-Libya.
- Alderson, I. C. and Wall, D. (1993) 'Does washback exist?', *Applied Linguistics*, 14(2), 115-129.
- Alderson, J.C., Clapham, C. and Wall, D. (1995) *Language Test Construction and Evaluation* Cambridge: Cambridge University Press.
- Alderson, J.C. and Hamp-Lyons, L. (1996) 'TOEFL preparation courses: A study of washback', *Language Testing*, 13(3), 280-297.
- Alderson, J.C. Nagy, E. and Overges, E. (eds.) (2000) *English Language Education in Hungary, Part II; Examining Hungarian Learners' Achievements in English*. Budapest: The British Council.
- Aldridge, A. and Levine, K. (2001) *Surveying social world: principles and practice in survey research*. Buckingham: Open University Press.
- Alhmali, R. (2007) *Students Attitudes in the Context of the Curriculum in Libyan Education in Middle and High Schools*, unpublished PhD thesis. University of Glasgow.
- Ali, M.A.A. (2008) *The Oral Error Correction Techniques Used by Libyan Secondary School Teachers of English*, unpublished PhD thesis. University of Sunderland.
- Amengual-Pizarro, M. (2009) 'Does the English Test in the Spanish University Entrance Examination Influence the Teaching of English?', *English studies*, 90(5), 582-598.
- Amrein-Beardsley, A. Berliner, D. and Rideau, S. (2010) 'Cheating in the first, second, and third degrees: Educators' responses to high-stakes testing.', *Education Policy Analysis Archives*, 18(14), 1-36.



- Andrews, S. (1995) 'Washback or washout? The relationship between examination reform and curriculum innovation', in D. Nunan, R. Berry and V. Berry (ed.) *Bringing about change in language education*, Hong Kong: University of Hong Kong, 67-81.
- Andrews, S. (2004) 'Washback and Innovation'. in L. Cheng Y. Watanabe and A. Curtis (eds.) *Washback in Language Testing*, Mahwah, New Jersey: Lawrence Erlbaum Association, 37-50.
- Andrews, S., Fullilove, J. and Wong, Y. (2002) 'Targeting washback-a case study', *System*, 30 207-223.
- Angrosino, M.V. (2008) 'Recontextualizing observation: Ethnography, pedagogy, and the prospects for a progressive political agenda', in Denzin, N. K. and Lincoln, Y. S. (ed.) 3rd ed. Thousand Oaks, CA.: Sage, 161-183.
- Atta, N. Ahmed, N. Ahmed, M. and Ali, Z. (2012) 'Role of Gender and Teaching Experience on Teachers' Self-Efficacy', *Language in India*, 12(9), 251-262.
- Bachman, I. and Palmer, A. S. (1996) *Language testing in practice*. Oxford: Oxford University Press.
- Bachman, L.F. and Palmer, A. (2010) *Language assessment in practice : developing language assessments and justifying their use in the real world*, Oxford: Oxford University Press.
- Bailey, K. (1996) 'Working for washback: A review of the washback concept in language testing.', *Language Testing*, 13(3), 257-279.
- Bailey, K. (1999) *Washback in Language Testing*, Princeton, New Jersey: ETS.
- Baker, C. (1992) *Attitudes and language*, Clevedon: Multilingual Matters Ltd.
- Bandura, A. (1997) *Self-efficacy: The exercise of control*. New York: Freeman.
- Barksdale-Ladd, M.A. and Thomas, K.F. (2000) 'What's at stakes in high-stakes testing: teachers and parents speak out', *Journal of Teacher Education*, 51 384-394.
- Barnes, M. and Clarke, D. and Stephens, M. (2000) 'Assessment: the engine of systemic curricula reform?' *Journal of Curriculum Studies*, 32(5), 623-650.
- Basturkmen, H. (2012) 'Review of research into the correspondence between language teachers' stated beliefs and practices', *System*, 40 282-295.
- Beglar, D. and Hunt, A. (1999) 'Revising and validating the 2000 Word Level and University Word Level Vocabulary Tests', *Language Testing*, 16(2), 131-162.

- Bejar, I.I. (2011) 'A validity-based approach to quality control and assurance of automated scoring', *Assessment in Education: Principles, Policy & Practice*, 18(3), 319-341.
- Berg, B.L. (2009) *Qualitative research methods for the social sciences*, 7th ed. Boston, Mass.; London: Ally & Bacon.
- Biggs, J. (1995) 'Assumptions underlying new approaches to educational assessment: Implications for Hong Kong' *Curriculum Forum*, 1995, 1-22.
- Bilgrami, A. (2002) 'Realism and Relativism'. in E. Sosa and E. Villanueva (ed.) *Realism and Relativism: Philosophical issues Volume 12*, Oxford and Ma: Blackwell, 1-25.
- Black, P. and Wiliam, D. (2006) *Inside the black box: Raising standards through classroom assessment*, Granada Learning.
- Borg, S. (2006) *Teacher cognition and language education: research and practice*, London: Continuum.
- Brannen, J. (2005) 'Mixing methods: The entry of qualitative and quantitative approaches into the research process', *International Journal of Social Research Methodology*, 8(3), 173-184.
- Brindley, G. (2008) 'Educational reform and language testing'. in E. Shohamy and N. H. Hornberger (ed.) *Encyclopaedia of language and education, Language Testing and Assessment*, 2nd ed. New York: Springer, 365-378.
- British Educational Research Association (BERA) (2011) *Ethical Guidelines for Educational Research*. London: BERA.
- Brophy, J. (1985) 'Interactions of male and female students with male and female teachers.'. in L.C. Wilkinson and C.B. Marrett (eds.) *Gender influences in classroom interaction*, New York: Academic press, 115-142.
- Brown, H.D. (2006) *Principles of language learning and teaching*, 5th ed. White Plains, NY: Longman.
- Brown, H.D. and Abeywickrama, P. (2010) *Language assessment : principles and classroom practices*, 2nd ed. Upper Saddle River, N.J. ; Harlow: Pearson Education.
- Brown, J.D. (2001) *Using surveys in language programs*, Cambridge: Cambridge University Press.
- Brown, J.D. (2008) 'Effect size and eta squared', *JALT Testing & Evaluation SIG Newsletter*, 12(2), 38-43.

- Bryman, A. (2008) *Social Research Methods*, 3rd ed. Oxford: Oxford university Press.
- Burrows, C. (2004) 'Washback in classroom-based assessment: A study of the washback effect in the Australian adult migrant English program', in L. Cheng, Y. Watanabe and A. Curtis (ed.) *Washback in language testing: research context and methods*, Mahwa, New Jersey: Lawrence Erlbaum Associates., 113-128.
- Chapman, D. W. and Snyder Jr, C. W. (2000) 'Can high stakes national testing improve instruction: reexamining conventional wisdom', *International Journal of Educational Development*, 20(6), 457-474.
- Charmaz, K. (2006) *Constructing grounded theory: a practical guide through qualitative analysis*, London ; Thousand Oaks, CA.: Sage.
- Cheng, L. (1997) 'How Does Washback Influence Teaching? Implication for Hong Kong.', *Language and Education*, 11(1), 38-54.
- Cheng, L. (1998) 'Impact of a Public English Examination Change on Students', *Studies in Educational Evaluation*, 24(3), 279-301.
- Cheng, L. (1999) 'Changing assessment: Washback on teacher perceptions and actions', *Teaching and teacher education*, 15(3), 253-272.
- Cheng, L. (2001) 'Washback studies: Methodological considerations' *Curriculum Forum*, 10 (2), 2001, 17-32.
- Cheng, L. (2005) *Changing language teaching through language testing: A washback study*, Cambridge University Press.
- Cheng, L. (2008) 'Washback, impact and consequences', in E. Shohamy and N.H. Hornberger (eds.) *Encyclopaedia of language and education, V. 7: language testing and assessment*, 2nd ed. New York: Springer, 349-364.
- Cheng, L. and Curtis, A. (2004) 'Washback or washout: A review of the impact of testing on teaching and learning.', in Cheng, L. Watanabe, Y., and Curtis, (ed.) *Washback in Language Testing: Research Context and Methods*, , Mahwah New Jersey USA: Lawrence Erlbaum Associates., 3-17.
- Cheng, L. Andrews, S. and Yu, Y. (2011) 'Impact and consequences of school-based assessment (SBA): Students' and parents' views of SBA in Hong Kong', *Language Testing*, 28(2), 221-249.
- Choi, I. (2008) 'The impact of EFL testing on EFL education in Korea', *Language Testing*, 25(1), 39-62.
- Cohen, D.K. and Spillane, J.P. (1992) 'Policy and practice: The relations between governance and instruction', *Review of research in education*, 18 3-49.

- Cohen, J. (1988) *Statistical power analysis for the behavioral sciences*, Lawrence Erlbaum.
- Cohen, L. and Manion, L. (1985) *Research methods in education*, 2nd ed. London: Croom Helm.
- Cohen, L. and Manion, L. (1994) *Research methods in education*, 4th ed. London: Routledge.
- Cohen, L. Manion, L. and Morrison, K. (2007) *Research methods in education*, 6th ed. London: Routledge.
- Cohen, L. Manion, L. and Morrison, K. (2011) *Research Methods in Education*, 7th ed. London: Routledge.
- Creswell, J.W. (1994) *Research design: qualitative and quantitative approaches*, Thousand Oaks, CA.; London: Sage.
- Creswell, J.W. (2005) *Educational research: planning, conducting, and evaluating quantitative and qualitative research*, 2 International ed. Upper Saddle River, N.J.: Pearson/Merrill Prentice Hall.
- Creswell, J.W. (2007) *Qualitative inquiry & research design: Choosing among five approaches*, CA: Sage.
- Creswell, J.W. (2009) *Research design : qualitative, quantitative, and mixed methods approaches*, 3rd ed. London: Sage.
- Creswell, J.W. Plano Clark, V.L. Gutmann, M.L. and Hanson, W.E. (2003) 'Advanced mixed methods research designs', *Handbook of mixed methods in social and behavioral research*, 209-240.
- Creswell, J.W. and Plano Clark, V.L. (2007) *Designing and conducting mixed methods research*, Thousand Oaks, CA. ; London: Sage.
- Crotty, M. (1998) *The foundations of social research : meaning and perspective in the research process*, London: Sage.
- Currie, M. and Chiramanee, T. (2010) 'The effect of the multiple-choice item format on the measurement of knowledge of language structure', *Language Testing*, 27(4), 471-491.
- De Vaus, D.A. (2001) *Research design in social research*, London: Sage.
- De Vaus, D.A. (2002) *Surveys in social research*, 5th ed. London: Routledge.
- Denscombe, M. (2003) *The good research guide : for small-scale social research projects*, 2nd ed. Buckingham: Open University Press.

- Denzin, N. (1978) *The research act: A theoretical introduction to sociological methods*, New York: McGraw-Hill.
- Denzin, N.K. (2009) *The research act: A theoretical introduction to sociological methods*, Aldine De Gruyter.
- Denzin, N.K. and Lincoln, Y.S. (2003) 'Part II: Paradigms and perspectives in transition '. in N. Denzin and Y.S. Lincoln (eds.) *The landscape of qualitative research: theories and issues*, 2nd ed. London: Sage, 245-252.
- Dörnyei, Z. (2003) *Questionnaires in second language research: construction, administration, and processing*, London: Erlbaum.
- Dörnyei, Z. (2007) *Research methods in applied linguistics : quantitative, qualitative, and mixed methodologies*, Oxford: Oxford University Press.
- Drew, C. Hardman, M. and Hart, W. (1996) *Designing and conducting research inquiry in educational and social science*. Boston: University of Utah.
- Drudy, S. (2008) 'Gender balance/gender bias: the teaching profession and the impact of feminisation ', *Gender and Education*, 20(4), 309-323.
- Dunkley, D. (2010) 'Washback in language testing', *Foreign Languages & Literature*, 35(1), 115-128.
- Elabbar, A.A. (2011) *An investigation of influences affecting Libyan English as Foreign Language University Teachers (LEFLUTs), teaching approaches in the language classrooms*, unpublished PhD thesis. University of Glasgow.
- El-Ebyary, K. (2009) *'Deconstructing the complexity of washback in relation to formative assessment in Egypt'*. 35. Cambridge: Cambridge ESOL.
- Erickson, F. (1985) *Qualitative methods in research on teaching*, Michigan State University: The Institute for Research on Teaching.
- Ernest, P. (1994) *An introduction to research methodology and paradigms*, Exeter: School of Education, University of Exeter.
- Falvey, P. (1996) 'The education of teachers of English in Hong Kong: A case for special treatment.' *Proceedings of ITEC' 95*, 5th-7th June 1995 1996, International teacher education conference, Hong Kong. 107-113.
- Farrel S. C. T. (2012) 'Novice-service language teacher development: bridging the gap between preservice and in-service education and development.', *TESOL Quarterly*, 46(3), 435-449.
- Ferman, I. (2004) 'The washback of an EFL national oral matriculation test to teaching and learning', in Cheng, L. Watanabe, Y., and Curtis (ed.) *Washback in*

*Language Testing: Research Context and Methods*, Mahwah New Jersey USA: Lawrence Erlbaum Associates., 191-210.

Fink, A. (2003) *The survey handbook*, 2nd ed. London: Sage.

Fish, J. (1988) *Responses to mandated standardised testing*, unpublished PhD dissertation. University of California: Los Angeles.

Fraenkel, J.R. and Wallen, N.E. (2005) *How to design and evaluate research in education*, 6th ed. Boston ; London: McGraw-Hill.

Fredericksen, J.R. and Collins, A. (1989) 'A system approach to educational testing', *Educational Researcher*, 18(9), 27-32.

Frino, L. Mhochain, R.N. O'Neil, H. and McGarry, F. (2008) *English for Libya, Preparatory 3: Teacher's Book*, Reading-UK: Garnet Publishing Ltd.

Fröhlich, M. Spada, N. and Allen, P. (1985) 'Differences in the communicative orientation of L2 classrooms', *TESOL Quarterly*, 19(1), 27-57.

Fulcher, G. (1997) 'An English language placement test: issues in reliability and validity', *Language Testing*, 14(2), 113-139.

Fulcher, G. (2011) 'Cheating gives lies to our test dependence: Policymakers are using language tests to carry a larger social burden than they can reasonably bear'. *theguardian Weekly, Tuesday 11th October 2011*, Learning English(4), <http://www.guardian.co.uk/education/2011/oct/11/why-more-language-test-cheating?INTCMP=SRCH>. Online accessed on 22 October 2012.

Fulcher, G. and Davidson, F. (2007) *Language testing and assessment*, Routledge.

Fulcher, G. and Davidson, F. (2009) 'Test architecture, test retrofit', *Language Testing*, 26(1), 123-144.

Fulcher, G. and Davidson, F. (2012) *The Routledge handbook of language testing*, London: Routledge.

Fulcher, G. (2012) Assessment Literacy for the Language Classroom, *Language assessment Quarterly*, 9 (2), 113-132.

Gay, L. & Airasian, P. (2003) *Educational research: competences for analysis and applications*. 7th ed. New Jersey: Merrill Prentice Hall.

Ghuma, M.A. (2011) *The Transferability of Reading Strategies between L1 (Arabic) and L2 (English)*, unpublished PhD thesis. University of Durham.

Gipps, C. (2011) *Beyond Testing: Towards a Theory of Educational Assessment*, Routledge.

- Gorard, S. (2001) *Quantitative methods in educational research: the role of numbers made easy*, London: Continuum.
- Gorsuch, G.J. (1999) *Exploring the relationship between educational policy and instruction in Japanese high schools EFL classrooms*, unpublished PhD thesis. The Temple University: Japan.
- Gorsuch, G.J. (2000) 'Educational policies and educational cultures: influences on teachers' approval of communicative activities.', *TESOL Quarterly*, 34(4), 675-710.
- Gray, D.E. (2004) *Doing research in the real world*, London; Thousand Oaks, CA.: Sage.
- Gray, D.E. (2009) *Doing research in the real world*, London: Sage.
- Green, A. (2007) *IELTS washback in Context: Preparation for academic writing in higher education*, *Studies in language testing* 25, Cambridge: UCLES/Cambridge University Press.
- Gregory, C. and Burg, S. (2006) *Addressing test anxiety in high-stakes environment: Strategies for classrooms and schools*, London: Sage.
- Guba, E.G. (1990) 'The alternative paradigm dialog', *The paradigm dialog*, 17-27.
- Guba, E.G. and Lincoln, Y.S. (1994) 'Competing paradigms in qualitative research.'. in N. K. Denzin and Y.S. Lincoln (ed.) *Handbook of qualitative research*, Thousand Oaks, London.: Sage, 105-117.
- Haggie, D. (2008) 'The strategic use of marking technologies to support innovation and diversity in assessment.' *Paper presented at the International Association of Educational Assessment*. September, in Singapore.  
[http://www.rm.com/RMVirtual/Media/Downloads/IAEA\\_Paper\\_Sept\\_2008\\_David\\_Haggie.pdf](http://www.rm.com/RMVirtual/Media/Downloads/IAEA_Paper_Sept_2008_David_Haggie.pdf) 2008, 1-12.
- Haladyna, T.M. Nolen, S.B. and Haas, N.S. (1991) 'Raising Standardized Achievement Test Scores and the Origins of Test Score Pollution', *Educational Researcher*, 20(5), 2-7.
- Hammersley, M. (1992) *What's wrong with ethnography?: methodological explorations*, London: Routledge.
- Hamp-Lyons, L. (1997) 'Washback, impact and validity: Ethical concerns', *Language testing*, 14(3), 295-303.
- Han, B. Dai, M. and Yang, L. (2004) 'On the Problems of CET.', *Foreign Languages and Their Teaching*, 2, 17-23.

Hayes, B. and Read, J. (2004) 'IELTS Test Preparation in New Zealand: Preparing Students for the IELTS Academic Module', in L. Cheng, Y. Watanabe and A. Curtis (ed.) *Washback in language testing: research contexts and methods*, Mahwah, NJ US: Lawrence Erlbaum Associates Publishers, 97-112.

Heaton, A. (1998) *Writing English Language Tests*, New York: Longman Inc.

Henn, M. Weinstein, M. and Foard, N. (2006) *A short introduction to social research*, London: Sage.

Henning, G. (1987) *A Guide to Language Testing: Development, Evaluation, Research*. Cambridge: Newbury House.

Hogan, T. and Murphy, G. (2007) 'Recommendations for Preparing and Scoring Constructed-Response Items: What the Experts Say', *Applied Measurement in Education*, 20(4), 427-441.

Hopf, D. and Hatzichriston, C. (1999) 'Teacher gender-related influences in Greek schools.', *British Journal of Educational Psychological Society*, 69 1-18.

Howe, K.R. (1998) 'The interpretive turn and the new debate in education', *Educational Researcher*, 27(8), 13-20.

Hsu, H. (2009) *The impact of implementing English proficiency tests as a graduation requirement at Taiwanese universities of technology*, unpublished PhD thesis. University of York.

Hughes, A. (2003) *Testing for language teachers*, 2nd ed. Cambridge: Cambridge University Press.

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

Hughes, A. (1993) *Backwash and TOEFL 2000*, unpublished manuscript. University of Reading.

Hung, T.A. (2012) 'A washback study on e-portfolio assessment in an English as a foreign language teacher preparation program.', 25(1), 21-36.

Hyland, K. (2002) *Teaching and researching writing*, Harlow: Longman.

Isiksal, M. and Cakiroglu, E. (2005) 'Teacher efficacy and academic performance', *Academic Exchange Quarterly*, 9(4), 28-32.

Johnson, D. (1994) *Research methods in educational management*, Longman Harlow.

Johnson, R.B. and Onwuegbuzie, A.J. (2004) 'Mixed methods research: A research paradigm whose time has come', *Educational researcher*, 33(7), 14-26.



- Jones, B.A. and Johnston, A.F. (2002) *The effects of high-stakes testing on instructional practices*. A paper presented at the 2002 annual meeting of the American Educational Research Association. New Orleans.
- Jones, G., Jones, D. B. and Hargrove, Y. T. (2003) *The Unintended Consequences of High-Stakes Testing*, USA: Rowman & Littlefield.
- Kane, M.J. (2002) 'Validating high-stakes testing programs', *Educational Measurement: Issues and Practice*, 21, 31-41.
- Lam, H. (1995) 'Methodology washback-an insider's view'. in D. Nunan R. Berry and V. Berry (eds.) *Bringing about change in language education*, Hong Kong: University of Hong Kong, 83-102.
- Larson-Hall, J. (2010) *A guide to doing statistics in second language research using SPSS*, New York ; London: Routledge.
- Lincoln, Y.S. and Guba, E. G. (2003) 'Paradigmatic controversies, contradictions, and emerging confluences', in N. Denzin and Y. Lincoln (ed.) *The landscape of qualitative research : theories and issues*, 2nd ed. Thousand Oaks, London: Sage, 253-291.
- Long, M.H. (1980) 'Inside the "black box": Methodological issues in classroom research on language learning', *Language Learning*, 30(1), 1-42.
- Madaus, G. (1988) 'The influence of testing on the curriculum'. in L. Tanner (ed.) *Critical issues in curriculum*, University of Chicago Press, 83-121.
- Mark, S.H. (2011) 'Tension between conflicting beliefs of an EFL teacher in teaching practices.', *RELC Journal*, 42(1), 53-67.
- Marshall, C. and Rossman, G.B. (2010) *Designing qualitative research*, 5th ed. London: Sage.
- Matsuda, P.K. and Silva, T.J. (2005) *Second language writing research: Perspectives on the process of knowledge construction*, Mahwah, New Jersey: Lawrence Erlbaum.
- ME (Ministry of Education) (2008) 'The Development of Education, the national report of Libya presented to the international conference on education, session (48), Geneva, November,' 2008, Libya Authority of Education 25 - 28.
- ME (Ministry of Education) Centre for Educational Curricula and Pedagogical Researches, Administration of Curricula (2012a) *School Studying Plan, Syllabi divisions on weekly basis, and Textbooks prescribed*, Tripoli-Libya: Nour Advertising Center.

- ME (Ministry of Education), Centre for Educational Curricula and Pedagogical Researches, Syllabi divisions on weekly basis, and Textbooks Prescribed (2012b) *A booklet outlines the rules of the divisions and distributions of the marks for the curriculum of the basic education stage*. Tripoli, Libya: Nour Advirtising Center.
- Merriam, S.B. (2002) *Qualitative research in practice: Examples for discussion and analysis*, Jossey-Bass.
- Mertens, D.M. (1998) *Research methods in education and psychology: Integrating diversity with quantitative & qualitative approaches*, Thousand Oaks, CA.: Sage.
- Messick, S. (1996) 'Validity and washback in language testing.', *Language Testing*, 13(3), 241-256.
- Miles, M.B. and Huberman, A.M. (1994) *Qualitative data analysis : an expanded sourcebook*, 2nd ed. Thousand Oaks, CA. ; London: Sage.
- Morris, T. (2006) *Social work research methods : four alternative paradigms*, Thousand Oaks, CA.; London: Sage.
- Morrison, M. (2007) 'What do we mean by educational research'. in A. Briggs and M. Coleman (eds.) *Research methods in educational leadership and management*. 2nd ed. London: Sage, 13-36.
- Muijs, D. (2004) *Doing quantitative research in education with SPSS*, London: Sage.
- Munoz, A. and Alvarez, E, (2010) 'Washback of an oral assessment in the EFL classroom', *Language Testing*, 27(1), 33-49.
- Nardi, P. (2006) *Doing survey research: a guide to quantitative methods*. 2nd ed. Boston, Mass and London: Pearson.
- Neuman, W.L. (2000) *Social research methods : qualitative and quantitative approaches*, 4th ed. Boston, Mass.: Allyn and Bacon.
- Newton, P.E. and Meadows, M. (2011) 'Marking quality within test and examination systems', *Assessment in Education: Principles, Policy & Practice*, 18(3), 213-216.
- Odo, D.M. (2012) 'The impact of high school exit exams on ESL learners in British Columbia', *English Language Teaching*, 5(9), 1-8.
- Onaiba, E.A. (2006) *An Evaluation of the Final English Examinations of Year Nine of the Basic Education Stage schools in Misrata, Libya*, unpublished M.A. dissertation. Academy of Postgraduate studies, School of Languages, English Department, Tripoli-Libya.
- OPEC (2011) *'Annual Statistical Bulletin 2010/2011'*. 2. Austria: Ueberreuter Print und Digimedia.

- Oppenheim, A.N. (2000) *Questionnaire design, interviewing and attitude measurement*, New ed. London: Continuum.
- Orafi, S.M.S. (2008) *Investigating teachers' practices and beliefs in relation to curriculum innovation in English language teaching in Libya*, Unpublished PhD thesis. University of Leeds.
- Orafi, S.M.S. and Borg, S. (2009) 'Intentions and realities in implementing communicative curriculum reform', *System*, 37(2), 243-253.
- Otman, W. and Karlberg, E. (2007) *The Libyan economy: economic diversification and international repositioning*, Berlin: Springer.
- Pallant, J. (2010) *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using the SPSS Program*, 4th ed. Maidenhead, Berkshire: Open University Press.
- Pan, Y.C. (2009) 'A review of washback and its pedagogical implications', *VNU Journal of Science, Foreign Languages*, 25 257-263.
- Patton, M.Q. (1987) *How to use qualitative methods in evaluation*, Sage.
- Patton, M.Q. (1990) *Qualitative evaluation and research methods*, Newbury Park, CA., Sage.
- Pearson, I. (1988) 'Tests as levers for change', in Chamberlain, D. and Baumgardner, R. (ed.) *ESP in the classroom: Practice and Evaluation, ELT Document 128*, London: Modern English Publications and The British Council., 98-107.
- Popham, W.J. (1987) 'The merits of measurement-driven instruction', *Phi Delta Kappa*, 68(9), 679-682.
- Popham, W.J. (1991) 'Appropriateness of teachers' test-preparation practices', *Educational Measurement: Issues and Practice*, 10(4), 12-15.
- Prodromou Luke (1995) 'The backwash effect: from testing to teaching', *ELT Journal*, 49 13-25.
- Punch, K.F. (2005) *Introduction to social research : quantitative and qualitative approaches*, 2nd ed. London: Sage.
- Qi, L. (2004) 'Has a high-stakes test produced the intended changes'. in L. Cheng, Y. Watanabe and A. Curtis (ed.) *Washback in language testing: Research context and methods*. 171-190.
- Qi, L. (2007) 'Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes test in China', *Assessment in Education: Principles, Policy & Practice*, 14(1), 51-74.

Read, J. and Hayes, B. (2003) 'The impact of IELTS on preparation for academic study in New Zealand', *IELTS research reports*, 4(5),.

Richards, J. (1990) 'The dilemma of teacher education in second language teaching'. in J.C. Richards and D. Nunan (eds.) *Second language teacher education*, Cambridge, UK: Cambridge University Press, 3-15.

Richards, J.C. and Nunan, D. (1990) *Second language teacher education*, Cambridge University Press.

Robson, C. (2002) *Real world research : a resource for social scientists and practitioner-researchers*, 2nd ed. Oxford: Blackwell.

Rocco, T.S. Bliss, L. Gallagher, S. Perez-Prado, A. Alacaci, C. Dwyer, E. Fine, J. and Pappamihel, N. (2003) 'The pragmatic and dialectical lenses: Two views of mixed methods use in education', in A. Tashakkori and C. Teddlie (ed.) *The Handbook of Mixed Methods in Social and Behavioural Sciences*, Thousand Oaks, CA: Sage, 595-615.

Rossmann, G.B. and Rallis, S.F. (1998) *Learning in the field: An introduction to qualitative research*, Thousand Oaks, CA: Sage.

Saif, S. (2006) 'Aiming for positive washback: a case study of international teaching assistants', *Language Testing*, 23(1), 1-34.

Sarantakos, S. (2005) *Social research*, 3rd ed. New York: Palgrave Macmillan.

Sari, M. (2012) 'Exploring gender roles' effects of Turkish women teachers on their teaching practices', *International Journal of Educational Development*, 3 418-425.

Saunders, M.N.K. Lewis, P. and Thornhill, A. (2009) *Research methods for business students*, 5th ed. Harlow: Financial Times Prentice Hall.

Schmitt, N. Schmitt, D. and Clapham, C. (2001) 'Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test', *Language Testing*, 18(1), 55-88.

Schwandt T. A. (2003) 'Three epistemological stances for qualitative inquiry: interpretivism, hermeneutics, and social constructionism', in N. Denzin and Y. Lincoln (ed.) *The landscape of qualitative research: theories and issues*, 2nd ed. Thousand Oaks, London: Sage, 292-331.

Seale, C. (1998) *Researching society and culture*, London: Sage.

Shazadi, T. Khatoon, S. Aziz, S. and Hassan, H. (2011) 'Determining factors affecting teachers' self-efficacy at secondary schools level', *Language in India*, 11(10), 385-395.

- Shepard, L.A. (1993) 'The place of testing reform in educational reform: A reply to Cizek.', *Educational Researcher*, 22(4), 10-14.
- Shih, C. (2007) 'A new washback model of students learning', *The Canadian Language Review*, 64(1), 135-161.
- Shihiba, E. (2011) *An Investigation of Libyan EFL Teachers' Conceptions of the Communicative Learner-Centred Approach in Relation to their Implementation of an English Language Curriculum Innovation in Secondary Schools*, unpublished PhD thesis. University of Durham.
- Shohamy, E. (1992) 'Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning', *The Modern Language Journal*, 76(4), 513-521.
- Shohamy, E. (1993) *The power of tests: The impact of language tests on teaching and learning.* NFLC Occasional Paper. Washington: National Foreign Language Center.
- Shohamy, E. (2001) *The power of tests : a critical perspective on the uses of language tests*, Harlow: Longman.
- Shohamy, E. (2007) 'Tests as Power Tools'. in J. Fox M. Wesche D. Bayliss L. Cheng C. Turner and C. Doe (eds.) *Language Testing Reconsidered*, Ottawa: University of Ottawa Press, 141-152.
- Shohamy, E. Donitsa-Schmidt, S. and Ferman, I. (1996) 'Test impact revisited: washback effect over time', *Language Testing*, 13(3), 298.
- Silverman, D. (2005) *Doing qualitative research : a practical handbook*, 2nd ed. Thousand Oaks, CA. ; London: Sage.
- Silverman, D. (2010) *Doing qualitative research: a practical handbook*, 3rd ed. London: Sage.
- Smith, J.K. (1983) 'Quantitative versus qualitative research: An attempt to clarify the issue', *Educational Researcher*, 12(3), 6-13.
- Smith, M.L. (1991a) 'Put to the test: The effects of external testing on teachers', *Educational Researcher*, 20(5), 8-11.
- Smith, M.L. (1991b) 'Meanings of test preparation', *American Educational Research Journal*, 28(3), 521-542.
- Spada, N. and Fröhlich, M. (1995) *COLT: Communicative orientation of language teaching observation scheme: Coding conventions and applications*, Sydney, Australia: National Centre for English Language Teaching and Research, Macquarie University.

- Spolsky, B. (1995) 'The examination of classroom backwash cycle: Some historical cases', in D. Nunan V. Berry and R. Berry (eds.) *Bringing about change in language education*, Hong Kong: University of Hong Kong, Department of Curriculum Studies, 55-66.
- Spratt, M. (2005) 'Washback and the classroom: the implications for teaching and learning of studies of washback from exams', *Language Teaching Research*, 9(1), 5-29.
- Stecher, B. (2002) 'Consequences of large-scales, high-stakes testing on schools and classroom practices.', in L. Hamilton, B. Stecher and S. Klein (ed.) *Making Sense of Test-Based Accountability in Education*, CA: Rand, USA.: Santa Monica, 79-99.
- Stecher, B. Chun, T. and Barron, S. (2004) 'The effects of assessment-driven reform on the teaching of writing in Washington State. ', in L. Cheng, Y. Watanabe and A. Curtis (ed.) *Washback in Language Testing: Research Context and Methods*, Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc., 53-72.
- Tashakkori, A. and Teddlie, C. (1998) *Mixed methodology : combining qualitative and quantitative approaches*, Thousand Oaks, CA. ; London: Sage.
- Taylor, S.J. and Bogdan, R. (1984) *Introduction to qualitative research methods: The search for meanings*, 2nd ed. New York: Wiley.
- Taylor, L.B. and Weir, C.J. (2009) *Language testing matters : investigating the wider social and educational impact of assessment : proceedings of the ALTE Cambridge Conference, April 2008*, Cambridge: Cambridge University Press.
- Teddlie, C. and Tashakkori, A. (2003) 'Major issues and controversies in the use of mixed methods in the social and behavioural sciences'. in A. Tashakkori and C. Teddlie (eds.) *Handbook of mixed methods in social and behavioral research*, Thousand Oaks, CA.: Sage, 3-50.
- Teddlie, C. and Tashakkori, A. (2009) *Foundations of mixed methods research : integrating quantitative and qualitative approaches in the social and behavioral sciences*, Thousand Oaks: Sage.
- The Guardian (2011) 'UN Security Council Resolution 1973 (2011) on Libya'. 1973. New York: UN security council.
- Toch, T. (2006) *Margins of error: The Education Testing Industry in the No Child Left Behind Era*, Washington DC, [http://www.educationsector.org/sites/default/files/publications/Margins\\_of\\_Error.pdf](http://www.educationsector.org/sites/default/files/publications/Margins_of_Error.pdf). Online accessed on 06 June 2012: Education Sector.
- Tsagari, D. (2009a) 'Revisiting the concept of test washback: Investigating FCE in Greek Language schools'. 35. Cambridge: Cambridge ESOL.

- Tsagari, D. (2009b) *The Complexity of Test Washback: An Empirical Study*. , Frankfurt: Peter Lang GmbH.
- Twati, J. (2008) 'The Influence of Societal Culture on the Adoption of Information System: The Case of Libya', *Communications of the IIMA*, 8(1), 1-12.
- Wall, D. (1996) 'Introducing new tests into traditional systems: Insights from general education and from innovation theory.', *Language Testing*, 13(3), 334-354.
- Wall, D. (1997) 'Impact and washback in language testing', *Encyclopaedia of language and education*, 7 291-302.
- Wall, D. (2000) 'The impact of high-stakes testing on teaching and learning-can this be predicted or controlled?', *System*, 28(4), 499-509.
- Wall, D. (2005) *The impact of high-stakes examinations on classroom teaching : a case study using insights from testing and innovation theory*, Cambridge: Cambridge University Press.
- Wall, D. (2012) 'Washback'. in G. Fulcher and F. Davidson (eds.) *The Routledge handbook of language testing*, London: Routledge, 79-92.
- Wall, D. and Alderson, J.C. (1993) 'Examining washback: the Sri Lankan impact study', *Language Testing*, 10(1), 41-69.
- Wall, D. and Horák, T. (2006) *The impact of changes in the TOEFL examination on teaching and learning in Central and Eastern Europe: Phase 1, the baseline study*, ETS.
- Wall, D. and Horák, T. (2011) 'The Impact of Changes in the TOEFL® Exam on Teaching in a Sample of Countries in Europe: Phase 3, The Role of the Coursebook Phase 4, Describing Change', ETS.
- Ward, C. (1981) *Preparing and Using Objective Questions*. Cheltenham: Stanley Thornes Ltd.
- Watanabe, Y. (1996) 'Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research.', *Language Testing*, 13(3), 318-333.
- Watanabe, Y. (2000) 'Washback effects of the English section of the Japanese university entrance examinations on instruction in pre-college level EFL', *Language Testing Update*, 27 42-47.
- Watanabe, Y. (2004) 'Methodology in Washback Studies'. in A. Curtis (ed.) *Washback in language testing: Research contexts and methods*. Mahwah, NJ US: Lawrence Erlbaum Associates Publishers, 19-36.

- Waters, A. and Vilches, L.C. (2008) 'Factors Affecting ELT Reforms : The Case of the Philippines Basic Education Curriculum', *RELC Journal*, 39(1), 5-24.
- Wiersma, W. and Jurs, S. (2009) *Research methods in education: An introduction*, 9th ed. Pearson International Edition.
- Winke, P. (2011) 'Evaluating the Validity of a High-Stakes ESL Test: Why Teachers' Perceptions Matter', *TESOL Quarterly*, 45(4), 628-660.
- Wiseman, S. (1961) *Examinations and English education*, University Press.
- Xi, X. (2010a) 'How do we go about investigating test fairness?', *Language Testing*, 27(2), 147-170.
- Xi, X. (2010b) 'Automated scoring and feedback systems: Where are we and where are we heading?', *Language Testing*, 27(3), 291-300.
- Yarbrough, T.L. (1999) *Teacher perception of the North Carolina ABC program and the relationship to classroom practice*. PhD ed. University of North Carolina and Chapel Hill.
- Yin, R.K. (2003) *Case study research : design and methods*, 3rd ed. Thousand Oaks, CA. ; London: Sage.
- Yin, R.K. (2009) *Case study research : design and methods*, 4th ed. Thousand Oaks, CA.: Sage.



## Appendices

### Appendix 1 A sample of the Old BECE in English

#### *Final Exam for Preparatory Schools*

#### *English Language Examination*

#### *First session, 2007*

*“Make sure that the exam includes 11 questions, and the number of pages is ‘5’” (Translated)*

**Time: 3 Hours**

**Total Marks: (60)**

---

**Q.1.: Read and answer the questions below:**

Libya is in north Africa. Its total area is 1.759.540 square kilometres and its population is about 5.5 million. Arabic is the official language in Libya, and some people also speak English or Italian. The country lies on the coast of the Mediterranean sea. This is also where the country's capital, Tripoli, is. Its neighbouring countries are Tunisia and Algeria in the west. Algeria is the second largest country in Africa. Its total area covers 2.382.740 square kilometres. There are about 28 million people in the country. The official language is Arabic and many people also speak French. The country's capital is Algiers.

Tunisia is a small country in north Africa, with a total area of 163.610 square kilometres and a population of about 9 million. Tunisia is further north than any other country in Africa. The country's capital is Tunis. The official language is Arabic. Many people speak French as a second language.

***Answer these questions:***

- a- What other language do some people in Libya speak?  
.....
- b- Which is smaller, Libya or Tunisia?  
.....
- c- Which country has a bigger population, Tunisia or Algeria?  
.....
- d- What are Libya's neighbouring countries in the west?  
.....
- e- Where is Tripoli?  
.....
- f- Which country is the furthest north?  
.....
- g- Is Algeria the biggest country in Africa?  
.....

(7 x 1 = 10 ½ marks)

**Q.2.: Fill in the space with these words:**

walk	help	hurt	climbing	hour	mountains
------	------	------	----------	------	-----------

Ali and Nuri went on a holiday to the ..... because they like ..... They climbed for about ..... Then Ali fell and ..... His leg. He couldn't ....., so Nuri went for .....

(6 x ½ = 3 marks)

---

**Q.3.: Choose the best answer:**

- a- There isn't (many – much) traffic in villages.
- b- There is (much – plenty) of entertainment.
- c- Do (many – much) people live in Tripoli.
- d- The girl (which – who) came is my sister.
- e- (Which – Who) cleaned the blackboard?
- f- This is the car (which – who) I want to buy.

(6 x ½ = 3 marks)

---

**Q.4.: Complete with correct words and right spelling:**

- a- ..... come before Sunday.
- b- ..... comes after Tuesday.
- c- Today is Friday, yesterday was .....
- d- ..... is the first month of the year.
- e- June, July and ..... are summer.
- f- ..... comes before Monday.

(6 x 1 = 6 marks)

---

**Q.5.: Join these sentences as it is shown in the brackets.**

- a- I like tea. I don't like coffee. (but)  
.....
- b- I was five. I started school. (when)  
.....
- c- you can come here. You can stay at home. (or)  
.....
- d- he was happy. He told everybody. (so ..... that)  
.....

(4 x 1½ = 6 marks)

---

**Q.6.: Write in words:**

- a- 21<sup>st</sup>= .....
- b- 90<sup>th</sup>= .....
- c- 100<sup>th</sup>= .....
- d- 1/6/98 = .....
- e- 13/11/1735 = .....
- f- 22/2/1542=.....

(6 x 1 = 6marks)

---

**Q.7.: Put ( ) or ( ) in front of each sentence, then correct the wrong ones:**

- a- I work on a farm. I look after sick people. ( )
- b- A fall can be very dangerous for old people. ( )
- c- We use (be + going to + infinitive) to talk about intentions. ( )
- d- People around the world eat the same kind of food. ( )
- e- The first explorers we know about were from America. ( )
- f- Sudan is the largest country in Africa. ( )
- g- Naguib Mahfouz was born in Tokyo. ( )
- h- Mahfouz began writing at the age of seventeen. ( )
- i- If you put a cork in water it sinks. ( )
- j- If you water plants they die. ( )

.....

.....

.....

.....

.....

(10 x ½ = 5 marks)

---

**Q.8.: write the vowels to complete the words.**

- a- H... nggl...d...ng
- b- Sc...b...d...y...ng
- c- p... .. m
- d- p... ..no
- e- sp ...gh...tt...
- f- ... .. dh(6 x 1 = 6 marks)

**Q.9.: Put these words in the right space**

conductor	electrical	electricity	atoms	equator
-----------	------------	-------------	-------	---------

- a- Salt water is good ..... of .....
- b- Lightning is a huge ..... spark in the sky.
- c- Tropical rainforests grow near the .....
- d- Everything around us is made up of .....

(5 x ½ = 2 ½ marks)

---

**Q.10.: Rewrite the sentences using “if”**

- a- Stay at home when there us a bad storm.  
If .....
- b- Don't touch anything electrical when you have wet hands.  
If .....
- c- Take plenty of water when you go camping in the desert.

If .....

- a- Don't drive when the weather is very bad.

.....

- b- Study hard the week before when you have a test.

If .....

- c- Ask your father for help when you can't do your homework.

If .....

(6 x 1 = 6 marks)

---

**Q.11.: Write 6 sentences about your lifestyle.**

These questions may help you.

- a- What kind of food do you eat?

- b- Do you ever miss meals?

- c- Do eat between meals?

- d- Do eat too much?

- e- How often do take exercise?

- f- What kind of exercise do you take?

.....

.....

.....

.....

.....

(6 sentences, one mark for each)

**GOOD LUCK**

**Appendix 2a A Sample of the New BECE in English**

***School Year: 2009 – 2010, Time allotment: 3 hours***

Subject: English language

Exam question paper/ page: 1

Student's name: .....

Student's sitting No.: .....

---

**(Choose "A" if the answer is correct and "B" if the answer is wrong) اسئلة الصواب والخطأ**

Q1 - A triangle has four sides and four angles.

A) True

B) False

Q2 - Carrots are fast food

A) True

B) False

Q3 - Take the egg out. Imperative for instruction

A) True

B) False

Q4 - Water travels up the plants.

A) True

B) False

Q5 - The word heavy is the opposite of rainy.

A) True

B) False

Q6 - Lettuce is usually eaten in soup.

A) True

B) False

Q7 - Water boil at 100 °C.

A) True

B) False

Q8 - In ancient times people made medicine from herbs and plants.

A) True

B) False

Q9 - During transpiration leaves lose soil

A) True

B) False

Q10 - If food is delicious, it does not taste very well.

A) True

B) False

Q11 Libya is in the south of Africa.

A) True

B) False

Subject: English language

Exam question paper/ page: 2

Student's name: .....

Student's sitting No.: .....

Q12 - We can use the present simple tense to talk about the future.

A) True

B) False

Q13 - Farming was an important industry in Libya.

A) True

B) False

Q14 - We shouldn't eat more fruit and less water during the day.

A) True

B) False

Q15 - Naguib Mahfouz had a successful writing career.

A) True

B) False

Q16 - If the sky is red at night, the next day will be fine.

A) True

B) False

Q17 - If you put a stone in water, it floats.

A) True

B) False

Q18 - In the 16th century doctors began to take temperature.

A) True

B) False

Q19 - Libya is three hours ahead of GMT.

A) True

B) False

Q20 - Plants take water from the soil through their roots.

A) True

B) False

Q21 - Have you ever eaten Chinese food? is a correct question.

A) True

B) False

Q22 - Yuri Gagarin was a scientist.

A) True

B) False

Q23 - Alhambra palace is the oldest.

A) True

B) False

Subject: English language

Exam question paper/ page: 3

Student's name: .....

Student's sitting No.: .....

Q24 - If you have plenty of exercise, you stay healthy.

A) True

B) False

**(Multiple-choice questions) أسئلة الاختيار المتعدد**

Q25 - Slow down means .....

- a) take it easy
- b) take care
- c) give and take
- d) take your time

Q26 - ..... of the world's rainforests are in Brazil.

- a) Two quarters
- b) One quarter
- c) Two thirds
- d) One third

Q27 - The noun of the verb treat is .....

- a) treating
- b) treatment
- c) treasure
- d) true

Q28 - There is a textile factory in .....

- a) Misurata
- b) Janzour
- c) Tajoura
- d) Sirt

Q29 - Edmund Hillary was an excellent .....

- a) climber
- b) farmer
- c) teachers
- d) player

Q30 - If people destroy rainforests, there will be more .....

- a) floods and droughts on Earth
- b) crops on Earth
- c) plants on Earth
- d) fruit on Earth

Subject: English language

Exam question paper/ page: 4

Student's name: .....

Student's sitting No.: .....

---

Q31 - The word drought means along period there is .....

- a) a lot of rain
- b) little or no rain
- c) hot weather
- d) cold weather

Q32 - The word water can be .....

- a) both a noun and a verb
- b) a verb
- c) an adjective
- d) a noun

Q33 - The word experience has ..... syllables.

- a) four
- b) three
- c) two
- d) five

Q34 - Yuri Gagarin was .....

- a) an astronomer
- b) an officer
- c) an astronaut
- d) a scientist

Q35 - We can use the internet more ..... now.

- a) easily
- b) easy
- c) easier
- d) easiest

Q36 - Mount Everest is ..... highest mountain in the world.

- a) a
- b) none
- c) the
- d) an

Q37 - Oil started to become important in Libya in .....

- a) the 1960s
- b) the 1940s
- c) the 1970s
- d) the 1950s

Q38 - Transpiration is when .....

- a) plants do not grow
- b) people move plants from one place to another
- c) plants lose water through their leaves
- d) plants move from one place to place



Subject: English language

Exam question paper/ page: **5**

Student's name: .....

Student's sitting No.: .....

---

Q39 - Sweets and biscuits .....

- a) aren't sweet
- b) are healthy
- c) aren't healthy
- d) are fruit

Q40 - The earth ..... round the sun.

- a) moves
- b) move
- c) moved
- d) is moving

Q41 – It is used when something is done sooner than I expected.

- a) yet
- b) of course
- c) just
- d) already

Q42 - The microscope was invented in .....

- a) the 16th century
- b) the 18th century
- c) the 17th century
- d) the 19th century

Q43 - Furion means .....

- a) very worried
- b) very sad
- c) very happy
- d) very angry

Q44 – You use it to carry things on your back, it's .....

- a) back pact
- b) big bag
- c) hand bag
- d) pact back

Q45 - The ice ..... to melt recently.

- a) starts
- b) start
- c) gas started
- d) started

Q46 - The average temperature in a tropical rainforest is .....

- a) 25-35 degrees
- b) 20-45 degrees
- c) 20-35 degrees
- d) 25-40 degrees

Subject: English language

Exam question paper/ page: 6

Student's name: .....

Student's sitting No.: .....

Q47 - At an international conference in ..... countries agreed to have a standard system of time keeping.

- a) 1886
- b) 1884
- c) 1882
- d) 1888

Q48 - I have twisted my ..... , I can't walk on it.

- a) uncle
- b) wrist
- c) arm
- d) ankle

Q49 - ..... discovered penicillin.

- a) Alexander Leopov
- b) Alexander Flemming
- c) Alexander Flemanov
- d) Alexander Gagarin

Q50 - Look at the sky. It ..... rain.

- a) should
- b) is going to
- c) would
- d) will

(Matching questions) اسئلة المزاوجة (عليك ان تختار من جمل العمود الایسر ما يتناسب مع جمل العمود الایمن)

Q51 - Era	a) How much a journey costs
Q52 - Steak	b) Usually happens in forests
Q53 - Fare	c) A period of time in history
Q54 - Conventional	d) I'd like it medium rare, please
Q55 - Transpiration	e) Usual, what has always been done

Q56 - Density	a) someone who works out directions
Q57 - Duration	b) change something from one thing to something else
Q58 - Navigator	c) the heaviness of an object for its size
Q59 - Convert	d) a material that is used a lot for building
Q60 - Concrete	e) the time something lasts

\*\*\*\*\***Finished** انتهت الاسئلة\*\*\*\*\*

## Appendix 2b The Answer Sheet Attached to the New BECE in English

ورقة الإجابة  
اللغة الانجليزية

الحكومة الليبية المؤقتة  
وزارة التربية والتعليم  
إدارة الإمتحانات

رمز الامتحان

0	0	0	0	0	0
1	1	1	1	1	1
2	2	2	2	2	2
3	3	3	3	3	3
4	4	4	4	4	4
5	5	5	5	5	5
6	6	6	6	6	6
7	7	7	7	7	7
8	8	8	8	8	8
9	9	9	9	9	9

يرجى عدم الكتابة في هذه المنطقة

يرجى تعبئة البيانات التالية خطياً والتوقيع

اسم الطالب : \_\_\_\_\_

التشعبة : \_\_\_\_\_

المدرسة : \_\_\_\_\_

الشعبة : \_\_\_\_\_

المادة : \_\_\_\_\_

التوقيع : \_\_\_\_\_

رقم الجلوس

0	0	0	0	0	0
1	1	1	1	1	1
2	2	2	2	2	2
3	3	3	3	3	3
4	4	4	4	4	4
5	5	5	5	5	5
6	6	6	6	6	6
7	7	7	7	7	7
8	8	8	8	8	8
9	9	9	9	9	9

1 A B C D E  
2 A B C D E  
3 A B C D E  
4 A B C D E  
5 A B C D E  
6 A B C D E  
7 A B C D E  
8 A B C D E  
9 A B C D E  
10 A B C D E  
11 A B C D E  
12 A B C D E  
13 A B C D E  
14 A B C D E  
15 A B C D E  
16 A B C D E  
17 A B C D E  
18 A B C D E  
19 A B C D E  
20 A B C D E  
21 A B C D E  
22 A B C D E  
23 A B C D E  
24 A B C D E  
25 A B C D E  
26 A B C D E  
27 A B C D E  
28 A B C D E  
29 A B C D E  
30 A B C D E  
31 A B C D E  
32 A B C D E  
33 A B C D E  
34 A B C D E  
35 A B C D E

36 A B C D E  
37 A B C D E  
38 A B C D E  
39 A B C D E  
40 A B C D E  
41 A B C D E  
42 A B C D E  
43 A B C D E  
44 A B C D E  
45 A B C D E  
46 A B C D E  
47 A B C D E  
48 A B C D E  
49 A B C D E  
50 A B C D E  
51 A B C D E  
52 A B C D E  
53 A B C D E  
54 A B C D E  
55 A B C D E  
56 A B C D E  
57 A B C D E  
58 A B C D E  
59 A B C D E  
60 A B C D E  
61 A B C D E  
62 A B C D E  
63 A B C D E  
64 A B C D E  
65 A B C D E  
66 A B C D E  
67 A B C D E  
68 A B C D E  
69 A B C D E  
70 A B C D E

71 A B C D E  
72 A B C D E  
73 A B C D E  
74 A B C D E  
75 A B C D E  
76 A B C D E  
77 A B C D E  
78 A B C D E  
79 A B C D E  
80 A B C D E  
81 A B C D E  
82 A B C D E  
83 A B C D E  
84 A B C D E  
85 A B C D E  
86 A B C D E  
87 A B C D E  
88 A B C D E  
89 A B C D E  
90 A B C D E  
91 A B C D E  
92 A B C D E  
93 A B C D E  
94 A B C D E  
95 A B C D E  
96 A B C D E  
97 A B C D E  
98 A B C D E  
99 A B C D E  
100 A B C D E  
101 A B C D E  
102 A B C D E  
103 A B C D E  
104 A B C D E  
105 A B C D E

106 A B C D E  
107 A B C D E  
108 A B C D E  
109 A B C D E  
110 A B C D E  
111 A B C D E  
112 A B C D E  
113 A B C D E  
114 A B C D E  
115 A B C D E  
116 A B C D E  
117 A B C D E  
118 A B C D E  
119 A B C D E  
120 A B C D E  
121 A B C D E  
122 A B C D E  
123 A B C D E  
124 A B C D E  
125 A B C D E  
126 A B C D E  
127 A B C D E  
128 A B C D E  
129 A B C D E  
130 A B C D E  
131 A B C D E  
132 A B C D E  
133 A B C D E  
134 A B C D E  
135 A B C D E  
136 A B C D E  
137 A B C D E  
138 A B C D E  
139 A B C D E  
140 A B C D E

141 A B C D E  
142 A B C D E  
143 A B C D E  
144 A B C D E  
145 A B C D E  
146 A B C D E  
147 A B C D E  
148 A B C D E  
149 A B C D E  
150 A B C D E  
151 A B C D E  
152 A B C D E  
153 A B C D E  
154 A B C D E  
155 A B C D E  
156 A B C D E  
157 A B C D E  
158 A B C D E  
159 A B C D E  
160 A B C D E  
161 A B C D E  
162 A B C D E  
163 A B C D E  
164 A B C D E  
165 A B C D E  
166 A B C D E  
167 A B C D E  
168 A B C D E  
169 A B C D E  
170 A B C D E  
171 A B C D E  
172 A B C D E  
173 A B C D E  
174 A B C D E  
175 A B C D E

14 13 12 11 10 9 8 7 6 5 4 3 2 1

يرجى عدم الكتابة في هذه المنطقة

14 13 12 11 10 9 8 7 6 5 4 3 2 1

لاستكمال لجنة المراقبة فقط

### Appendix 3 An Overview of most Washback Studies Addressed in this Study

Researcher(s)	Exam studied	The context	Main issues addressed	Methodological tools used
Amengual-Pizarro (2009)	The English test (ET) in the Spanish university entrance examination	Spain	ET impact on teachers' methods, curriculum and teachers' feelings and attitudes.	<ul style="list-style-type: none"> <li>Teacher questionnaire</li> </ul>
Andrews (1995)	Oral component of the Use of English test (UE)	Hong Kong secondary schools	Gaining some insights into the EU actual washback effect on curriculum innovation	<ul style="list-style-type: none"> <li>Two parallel questionnaires: to the Working Party members, and to teachers</li> </ul>
Andrews <i>et al</i> (2002)	Oral component of the Revised Use of English (RUE)	Hong Kong	The influence of the addition of oral component of the exam on students' oral performance	<ul style="list-style-type: none"> <li>Videotapes of mock oral tests</li> <li>Grading of oral tests</li> <li>Discourse analysis</li> </ul>
Burrows (2004)	A new classroom assessment system accompanied by introducing Certificate in Spoken and Written English (CSWE) programme	Australia	Relationship between the new system of assessment and teachers' classroom practices	<ul style="list-style-type: none"> <li>Teacher questionnaire</li> <li>Teacher interviews</li> <li>Classroom observations</li> </ul>
Cheng (1997)	The revised HKCEE in English	Hong Kong Secondary schools	How does washback influence classroom teaching	<ul style="list-style-type: none"> <li>questionnaires to teachers and students</li> <li>observations</li> </ul>

## References

Cheng (1998)	The revised HKCEE in English	Hong Kong Secondary schools	exam impact on students' attitudes and perceptions of their English learning	<ul style="list-style-type: none"> <li>• student questionnaires</li> </ul>
Cheng (1999)	The revised HKCEE in English	Hong Kong Secondary schools	Exam influence on teachers' perceptions and actions	<ul style="list-style-type: none"> <li>• classroom observations</li> <li>• teachers interviews</li> </ul>
Cheng <i>et al.</i> (2011)	A national-wide exam (SBA in English)	Hong Kong	Students' perceptions of the impact of the exam, and their parents' role in the process	<ul style="list-style-type: none"> <li>• Students' questionnaire</li> <li>• Parents' questionnaire</li> </ul>
Choi (2008)	Standardized EFL tests in Korean schools	Korea	Tests impact on EFL education across the education stages: elementary, secondary and university education	<ul style="list-style-type: none"> <li>• surveys to students and teachers across the three stage of education</li> </ul>
El-Ebyary (2009)	Classroom formative assessment	Egypt teacher college	How washback is linked to formative assessment	<ul style="list-style-type: none"> <li>• teacher and student questionnaires</li> <li>• interviews</li> </ul>
Ferman (2004)	National EFL matriculation test	Israel	The washback of a new EFL oral matriculation test to teaching and learning	<ul style="list-style-type: none"> <li>• student questionnaire</li> <li>• teacher and inspector interviews</li> <li>• documentary analysis</li> </ul>
Fish (2007)	A high-stakes standardized test	California, USA	Teachers' responses to mandated standardized testing	<ul style="list-style-type: none"> <li>• teachers' questionnaire</li> </ul>
Lam (1995)	Use of English TEST (UE)	Hong Kong	Investigating the extent to which exams induce positive washback in classroom teaching and learning	<ul style="list-style-type: none"> <li>• questionnaire to teachers</li> <li>• Documentary analysis (textbooks)</li> </ul>

## References

Munoz and Alvarez (2010)	Oral Assessment System (OAS)	USA, language centre <sup>a</sup>	Determine washback effects of an oral assessment system on some aspects of EFL education	<ul style="list-style-type: none"> <li>• teacher and student questionnaires</li> <li>• observations</li> </ul>
Qi (2004)	National Matriculation English Test (NMET)	China	Examined the reasons why the NMET test failed to induce washback in ELT.	<ul style="list-style-type: none"> <li>• in-depth interviews and follow-up contacts with English inspectors, teachers, and exam constructors</li> </ul>
Saif (2006)	Spoken language test designed for international teaching assistants (ITA test)	Canada	Aiming for positive washback	<ul style="list-style-type: none"> <li>• Interviews</li> <li>• Observations</li> <li>• Analysis of test scores</li> </ul>
Shohamy (1993)	<ul style="list-style-type: none"> <li>• ASL test</li> <li>• (EFL) oral test</li> <li>• L1 reading test</li> </ul>	Israel	The influence of examinations on various facets of language teaching and learning	<ul style="list-style-type: none"> <li>• Student questionnaire</li> <li>• Classroom observation</li> <li>• Interviews</li> <li>• Documentary analysis</li> </ul>
Shohamy <i>et al.</i> (1996)	<ul style="list-style-type: none"> <li>• ASL test</li> <li>• EFL oral test</li> </ul>	Israel	Revisiting Shohamy's (1993) study of washback effects over time	<ul style="list-style-type: none"> <li>• Student questionnaire</li> <li>• Teacher and inspector interviews</li> <li>• Documentary analysis</li> </ul>
Smith (1991)	Iowa Test of Basic Skills (BST)	Arizona, USA	The role of external testing in elementary schools	<ul style="list-style-type: none"> <li>• Long term, intensive observation of teachers</li> <li>• Clinical interviews with teachers and school administrators</li> </ul>
Stecher <i>et al.</i> (2004)	Washington Assessment of Student Learning (WASL) tests	Washington USA	How the WASL tests affected schools and teachers' methods to teaching writing	<ul style="list-style-type: none"> <li>• Questionnaires to teachers and school administrators</li> </ul>
Tsagari (2009a) (preliminary study)	First certificate in English (FCE) examination	Greece	Effects of FCE on participants' feelings and attitudes, teaching methods and contents	<ul style="list-style-type: none"> <li>• Documentary analysis</li> <li>• A diary study</li> </ul>

## References

---

Tsagari (2009b)	First certificate in English (FCE) examination	Greece	Investigating the effect of the FCE on participants' perceptions and material design	<ul style="list-style-type: none"> <li>• Documentary analysis</li> <li>• A diary study</li> </ul>
Wall (2005)	O-level Examination in English	Sri Lanka	Investigating the effects of changing the O-level examination	<ul style="list-style-type: none"> <li>• Teacher questionnaires</li> <li>• Classroom observations</li> <li>• Interviews with teachers</li> <li>• Documentary analysis</li> </ul>
Wall and Alderson (1993)	O-level Examination in English as an international language	Sri Lanka	Investigating washback of the exam on different areas of classroom teaching	<ul style="list-style-type: none"> <li>• Classroom observations</li> <li>• Interviews with teachers and students</li> </ul>
Watanabe (1996, 2000)	University Entrance Examination (UEE) in Japan	Japan	The relationship between the UEE and the grammar-translation method to ELT	<ul style="list-style-type: none"> <li>• Questionnaires</li> <li>• Interview and observation of two teachers</li> <li>• Documentary analysis</li> </ul>

## Appendix 4 Teacher Questionnaire for the Main Study

***Dear Teacher,***

I am conducting a small research survey for my doctorate at the University of Leicester in the United Kingdom. I would like to ask you for your opinions of the English examination revised recently, and administered to grade 9 students in your schools. To help me, please fill in this questionnaire based on your own experience. Your taking part will make a great contribution to the study. It should take no more than twenty minutes to complete. ***All information will be anonymous and treated in the strictest confidence.*** Thank you very much.

Please read carefully the instructions and guidance for the way of responding to each questions. If you have any doubts about any of the questionnaire questions/items, please contact me as this is very important to make your answers as adequate and accurate as possible. (Abdulhamid Onaiba: Tel: 00 44 7833 693444; 0913387007 Email: [amol0@le.ac.uk](mailto:amol0@le.ac.uk) or [hamidmo71@yahoo.com](mailto:hamidmo71@yahoo.com) )

***\*Please give yourself a code reference for future contact*** \_\_\_\_\_

### **PART I:**

**Please tick one appropriate answer or provide written answers.**

- (1) Your gender: ☐Female ☐Male School name and city: \_\_\_\_\_
- (2) Your age: ☐22 - 30 ☐31 - 40 ☐41 - 50 ☐above 50
- (3) Your academic qualification: ☐Bachelor of Arts ☐ Bachelor of Education  
☐Teacher's diploma ☐others \_\_\_\_\_
- (4) Number of years you have been teaching English:  
☐0 - 2 ☐3 - 6 ☐7 - 10 ☐11 - 15 ☐above 15
- (5) Grades/Classes you are currently teaching other than Grade Nine. You may tick more than one box:  
☐Grade five ☐Grade six ☐Grade seven ☐Grade eight
- (6) Number of periods you teach English per week:  
☐8 – 12 ☐14 – 18 ☐20 – 24
- (7) The typical size of each class you teach in terms of student numbers  
☐20 – 30 ☐31 – 40 ☐above40
- (8) The length of time between the announcement of the exam change and its first implementation was: ☐1 – 3months ☐4 – 8months ☐09 – 16months
- (9) Have you been involved in workshops relating to the revised exam? ☐yes ☐no
- (10) Have you taken courses specifically in testing and evaluation? ☐yes ☐no



**PART II:**

**In the brackets [ ], please mark the following on a five point scale as:**

**[1] Strongly disagree [2] Disagree [3] Undecided [4] Agree [5] Strongly Agree**

**(11)** What do you see as the major reasons for the education authority to change the old English exam by administering a new revised one?

- [ ] To prepare students for their future education
- [ ] To motivate students to use integrated skills
- [ ] To cover all the components and contents of the English curriculum
- [ ] To eliminate the risk of cheating
- [ ] To score the answer sheets mechanically
- [ ] To disseminate the results quickly, adequately and as transparently as possible
- [ ] To provide students with a gauge of their language learning achievement regarding the English curriculum
- [ ] If you think there are other reasons, please add them here:  
\_\_\_\_\_  
\_\_\_\_\_

**(12)** What are the key changes that you have perceived in the revised exam paper with comparison to the previous exam?

- [ ] More related to the principles of the current English curriculum
- [ ] More emphasis on reading passages stated in the textbooks
- [ ] More emphasis on grammatical usage
- [ ] More emphasis on essay writing
- [ ] Others, please specify: \_\_\_\_\_

**(13)** According to your own beliefs, the new revised exam is

- [ ] A good tool to gauge students' performance in classrooms
- [ ] A good indicator of students ability in using language in real life situation
- [ ] Not appropriate at all.
- [ ] Appropriate but it needs to include other aspects such as: listening and speaking, and also testing reading and writing properly.
- [ ] Other opinions: \_\_\_\_\_

**(14)** What kind of extra work or pressure if any do you think the revised exam has put on you in your teaching?

- [ ] Following a new syllabus
- [ ] Doing more lesson preparation
- [ ] Feeling more responsible for the success or failure of your students on the test
- [ ] Revising the existing teaching materials that you have
- [ ] Familiarising students with the content and format of the revised exam
- [ ] Employing new teaching methods
- [ ] Organizing more focused activities that reflect exam activities

(15) According to your educational beliefs, it is better to assess students in the following area(s) of language:

- ☐ Only oral and listening skills
- ☐ Only writing and spelling
- ☐ Only reading and vocabulary
- ☐ Only grammar
- ☐ An integration of all aspects mentioned above

(16) It is important for your students to study English in order:

- ☐ To pursue further education in or out of the country
- ☐ To pass examinations
- ☐ To satisfy parents' requirements
- ☐ To obtain jobs
- ☐ Because it is educationally and culturally desirable to have knowledge of English

### **PART III, SECTION I:**

(17) What are the most significant changes you have made in your teaching in the context of the new revised exam content?

- ☐ Teaching according to the test content
- ☐ Adopting new teaching methods
- ☐ Using a more communicative approach in teaching
- ☐ Emphasizing listening and speaking skills
- ☐ The amount and frequency of practicing activities similar to the final exam increased in my classroom as the exam approached.

(18) What do you find the most difficult or challenging aspects of teaching English in the classroom?

- ☐ Students' current level of English
- ☐ Class size
- ☐ Insufficiency-weekly time allotment for teaching English as a school subject
- ☐ Learning environment
- ☐ Limited teaching and learning aids and facilities
- ☐ Overly demanding workload (i.e. too many classes per week)
- ☐ Others, please specify \_\_\_\_\_

(19) What are the learning strategies you recommend to your students in the context of the new exam? I recommend students ...

- ☐ To take extra review hours in or out of school as test preparation lessons
- ☐ To train on previous exam papers similar to the revised exam
- ☐ To put more emphasis on speaking and listening skills and their components
- ☐ To focus more on reading specifically the reading passages in the textbook
- ☐ To put more emphasis on writing
- ☐ To concentrate on the grammatical structures provided in their textbooks
- ☐ To focus more on the tasks and exercises stated in the workbooks provided
- ☐ To use teaching materials other than their current textbooks
- ☐ Others, please specify: \_\_\_\_\_

(20) The factors that most influence your teaching are:

- ☐ Teaching experience and beliefs
- ☐ Teaching syllabus
- ☐ The need to obtain satisfaction in teaching
- ☐ Textbooks
- ☐ External exams
- ☐ Learners' expectations
- ☐ Inspectors' expectations

**In the brackets [ ], please mark the following on a five point scale as:**

**[1] = never    [2] = seldom    [3] = sometimes    [4] = often    [5] = always**

**(21)** How often do you do the following activities in class?

- ☐ Tell the students the aims of each lesson
- ☐ Demonstrate how to do particular language activities
- ☐ Make revisions to previous lessons
- ☐ Correct and explain homework
- ☐ Explain specific language items such as words or sentences
- ☐ Use previous years' tests to review the drills and format of the revised exam
- ☐ Organise group work or discussion
- ☐ Organise integrated language tasks
- ☐ Make students practice general information questions about events related directly to reading passages from the textbook

**(22)** How often do you do the following in class?

- ☐ Talk to the whole class
- ☐ Talk to groups of students
- ☐ Talk to individual students
- ☐ Keep silent

**(23)** How often do you use the following teaching and learning aids in your teaching?

- ☐ Course-book
- ☐ Workbook
- ☐ Teacher's book
- ☐ Language laboratory
- ☐ Pictures and cards
- ☐ The blackboard
- ☐ Television/radio

**Please TICK the appropriate answer or provide written answers**

**(24)** How often do you use English and Arabic in your classroom instruction?

- ☐ Mainly English
- ☐ English supplemented with occasional Arabic explanation
- ☐ Half English and half Arabic
- ☐ Mainly Arabic

(25) The content and format of the classroom tests you usually write to your grade 9 students and the content and format of the revised exam are:

- ☐ Inconsistent
- ☐ Slightly consistent
- ☐ To some extent consistent
- ☐ Highly consistent
- ☐ Others \_\_\_\_\_

(26) As you teach classes other than grade 9, the classroom tests you usually write for grade 9 classes and those you write for other grades/classes are:

- ☐ Inconsistent
- ☐ Slightly consistent
- ☐ To some extent consistent
- ☐ Highly consistent

(27) Your English teaching to grade 9 classes and other grades/classes are:

- ☐ Not similar
- ☐ To some extent similar
- ☐ Similar
- ☐ Very similar

### **PART III, SECTION II:**

(28) The people who are generally responsible for designing the teaching materials (i.e. textbooks) and the arrangement of lessons in your school are:

- ☐ Teachers of English in schools
- ☐ School principals
- ☐ The Authority of Education

(29) On what basis do you arrange your teaching in the classroom?

- ☐ According to the current textbook and its content
- ☐ According to the teaching materials you prepare by yourself
- ☐ Mixture of both
- ☐ Others, please specify: \_\_\_\_\_

**In the brackets [ ], please mark the following on a five point scale as:**

**[1] = never    [2] = seldom    [3] = sometimes    [4] = often    [5] = always**

(30) Which of the following language aspects from the content of the current teaching syllabus do you focus in your teaching?

- ☐ Reading comprehension lessons
- ☐ Writing aspects
- ☐ The knowledge of grammar and vocabulary
- ☐ Listening tasks and activities
- ☐ Speaking and pronunciation sections

**N.B.**

**\* Please check to see whether you have answered all the questions.**

**\* I need some volunteers to participate in the interview sessions. If you are willing to participate, please tick the box. ☐**

***Thank you very much for your help!***

## Appendix 5 Teacher Questionnaire for the Pilot Study

**Dear Teacher,**

I am conducting a small research survey for my doctorate at the University of Leicester in the United Kingdom. I would like to ask you for your opinions of the English examination revised recently, and administered to grade 9 students in your schools. To help me, please fill in this questionnaire based on your own experience. It should take no more than twenty minutes to complete. ***All information will be anonymous and treated in the strictest confidence.*** Thank you very much.

If you have any doubts about any of the questionnaire questions/items, please contact me as this is very important to make your answers as adequate and accurate as possible. (Abdulhamid Onaiba: Tel: 00 44 7833 693444; Email: [amo10@le.ac.uk](mailto:amo10@le.ac.uk) or [hamidmo71@yahoo.com](mailto:hamidmo71@yahoo.com) ).

### **PART I:**

***\*Please give yourself a code reference for future contact \_\_\_\_\_***

**Please tick one appropriate answer or provide written answers.**

(1) Your gender: ☐Female ☐Male School name and city: \_\_\_\_\_

(2) Your age: ☐22 - 30 ☐31 - 40 ☐41 - 50 ☐above 50

(3) Your academic qualification: ☐Bachelor ☐Teacher's diploma ☐others

(4) Number of years you have been teaching English:

☐0 - 2 ☐3 - 6 ☐7 - 10 ☐11 - 15 ☐above 15

(5) Grades/Classes you are currently teaching. Tick more than one box if necessary:

☐Grade five ☐Grade six ☐Grade seven ☐Grade eight ☐Grade nine

(6) Number of periods you teach English per week:

☐8 – 12 ☐14 – 18 ☐20 – 24

(7) The typical size of each class you teach in terms of student numbers

☐20 – 30 ☐31 – 40 ☐above 40

(8) The length of time between the announcement of the exam change and its first implementation was: ☐1 – 3months ☐4 – 9months ☐10 – 16months

(9) Have you been involved in workshops relating to the revised exam? ☐yes ☐no

(10) Have you taken courses specifically in testing and evaluation? ☐yes ☐no

**PART II:**

**In the brackets [ ], please mark the following on a five point scale as: [1] Strongly disagree [2] Disagree [3] Undecided [4] Agree [5] Strongly Agree**

**(11)** What do you see as the major reasons for the PCE (People's Committee of Education) to change the old English exam by administering a new revised one?

- [ ] To prepare students for their future education
- [ ] To motivate students to use integrated skills
- [ ] To cover all the components and contents of the English curriculum
- [ ] To eliminate the risk of cheating
- [ ] To score the answer sheets mechanically
- [ ] To disseminate the results quickly, adequately and as transparently as possible
- [ ] To provide students with a gauge of their language learning achievement regarding the English curriculum

“If you think there are other reasons, please add them here:

[ ] \_\_\_\_\_

**(12)** What are the key changes that you have perceived in the revised exam paper with comparison to the previous exam?

- [ ] More related to the principles of the current English curriculum
- [ ] More emphasis on reading passages stated in the textbooks
- [ ] More emphasis on grammatical usage
- [ ] More emphasis on oral activities
- [ ] More emphasis on essay writing
- [ ] Others, please specify: \_\_\_\_\_

**(13)** According to your own beliefs, the new revised exam is

- [ ] A good tool to gauge students' performance in classrooms
- [ ] A good indicator of students ability in using language in real life situation
- [ ] Not appropriate at all, and thus it needs to be recast and revised
- [ ] Being treated with scepticism by teachers and the public
- [ ] Appropriate but it needs to include other aspects such as: \_\_\_\_\_

**(14)** What kind of extra work or pressure or concern if any do you think the revised exam has put on you in your teaching?

- [ ] Following a new syllabus
- [ ] Doing more lesson preparation
- [ ] Feeling more responsible for the success or failure of your students on the test
- [ ] Revising the existing materials that you have
- [ ] Familiarising students with the content and format of the revised exam
- [ ] Employing new teaching methods
- [ ] Setting up new teaching objectives
- [ ] Meeting new challenges in teaching
- [ ] Organizing more focused activities that reflect exam activities

**(15)** According to your educational beliefs, it is better to assess students in the following area(s) of language:

- ☐ Oral and listening skills
- ☐ Writing and spelling
- ☐ Reading and vocabulary
- ☐ Grammar
- ☐ An integration of all aspects mentioned above

(16) Why do you think it is important for your students to study English?

- ☐ To pursue further education in or out of the country
- ☐ To pass examinations
- ☐ To satisfy parents' requirements
- ☐ To obtain jobs
- ☐ Because it is educationally and culturally desirable to have knowledge of English

(17) What types of activities do you believe are most effective in language learning?

- ☐ Task-oriented activities
- ☐ Language games
- ☐ Role play and group discussion
- ☐ Authentic materials
- ☐ Extracurricular

### **PART III: SECTION I**

(18) What are the most significant changes you have made in your teaching in the context of the new revised exam format?

- ☐ Teaching according to the test format
- ☐ Adopting new teaching methods
- ☐ Using a more communicative approach in teaching
- ☐ Emphasizing listening and speaking skills
- ☐ Focussing on language usage rather than language use in class
- ☐ the amount and frequency of practicing activities similar to the final exam increased in my classroom as the exam approached.

(19) What do you find the most difficult or challenging aspects of teaching English in the classroom?

- ☐ Students' current level of English
- ☐ Class size
- ☐ Insufficiency-weekly time allotment for teaching English as a school subject
- ☐ Inadequate time for students to practice English outside of the classroom
- ☐ Learning environment
- ☐ Limited teaching and learning aids and facilities
- ☐ Overly demanding workload

**Please tick the appropriate answer. Tick more than one if necessary.**

(20) What are the learning strategies you recommend to your students in the context of the new exam? I recommend students ...

- ☐ ] To take extra review hours in or out of school as test preparation lessons
- ☐ ] To train on mock exams
- ☐ ] To put more emphasis on speaking and listening skills and their components
- ☐ ] To focus more on reading specifically the reading passages in the textbook
- ☐ ] To put more emphasis on writing
- ☐ ] To concentrate on the grammatical structures provided in their textbooks
- ☐ ] To consult teaching materials other than their current textbooks
- ☐ ] To change from passive learning to active learning
- ☐ ] Others, please specify: \_\_\_\_\_

(21) The factors that most influence your teaching are:

- ☐ ] Professional training
- ☐ ] Teaching experience and beliefs
- ☐ ] Teaching syllabus
- ☐ ] Past experience as a language learner
- ☐ ] The need to obtain satisfaction in teaching
- ☐ ] Textbooks
- ☐ ] External exams
- ☐ ] Learners' expectations
- ☐ ] Inspectors' expectations
- ☐ ] Social expectations

(22) In what ways would you like to motivate your students to learn English?

- ☐ ] Encourage them to expose to various English media outside the classroom
- ☐ ] Advise them to create opportunities to use English in daily life
- ☐ ] Use more authentic materials
- ☐ ] Organize real life language activities
- ☐ ] Use more interesting language games
- ☐ ] Give students more encouragement to learn
- ☐ ] Create a positive attitude toward language learning
- ☐ ] Provide students with effective language learning strategies

**In the brackets [ ], please mark the following on a five point scale as:** [1] = never [2] = seldom [3] = sometimes [4] = often [5] = always

(23) How often do you do the following activities in class?

- ☐ ] Tell the students the aims of each lesson
- ☐ ] Demonstrate how to do particular language activities
- ☐ ] Explain the meaning of the text
- ☐ ] Make revisions to previous lessons
- ☐ ] Explain and correct homework
- ☐ ] Explain specific language items such as words or sentences
- ☐ ] Use previous years' tests to review the drills and format of the revised exam
- ☐ ] Organise group work or discussion
- ☐ ] Organise integrated language tasks
- ☐ ] Make students practice general information questions about events related directly to reading passages from the textbook

(24) How often do you do the following in class?

- ☐ ] Talk to the whole class
- ☐ ] Talk to groups of students



- ☐ Talk to individual students
- ☐ Keep silent

(25) How often do you use the following teaching and learning aids in your teaching?

- ☐ Textbooks
- ☐ Supplementary materials
- ☐ Language laboratory
- ☐ Pictures and cards
- ☐ The blackboard
- ☐ Television/radio
- ☐ scheme of work

**Please tick only on answer that you see most appropriate**

(26) How often do you use English and Arabic in your classroom instruction?

- ☐ Mainly English
- ☐ English supplemented with occasional Arabic explanation
- ☐ Half English and half Arabic
- ☐ Mainly Arabic

(27) How often are classroom tests you write for grade 9 consistent with the revised exam?

- ☐ Inconsistent
- ☐ Slightly consistent
- ☐ To some extent consistent
- ☐ highly consistent

(28) As you teach classes other than grade 9, how often are grade 9 on-going tests and other grades' tests you prepare during school year consistent?

- ☐ never
- ☐ sometimes
- ☐ often
- ☐ always

(29) How often similar is your class 9 English teaching with other classes you teach?

- ☐ never
- ☐ sometimes
- ☐ often
- ☐ always

### **PART III: SECTION II**

**Tick the appropriate answer. Tick more than one if necessary.**

(30) Who is generally accountable for designing teaching materials (textbooks) and the arrangement of lessons?

- ☐ Teachers of English in schools
- ☐ school principals
- ☐ PCE (People's Committee of Education)

(31) How do you arrange your teaching in your school?

- ☐ According to the current textbook and its content
- ☐ According to the school scheme of work
- ☐ According to other teaching materials you prepare by yourself

(32) What are the primary functions of textbooks in teaching?

- ☐ To provide practical activities
- ☐ To provide structured language programme to follow
- ☐ To provide information about the language

(33) Which of the following types of teaching materials do you use in class?

- ☐ Only the existing materials designed for schools by the PCE
- ☐ Only teaching materials other than the existing one
- ☐ Mixture of both

**In the brackets [ ], please grade the following questions/items on a 5-point scale as:**

**[1] = never    [2] = seldom    [3] = sometimes    [4] = often    [5] = always**

(34) According to your teaching agenda, which of the following language aspects from the content of the current teaching syllabus do you focus on more in class?

- ☐ Reading comprehension lessons
- ☐ Writing aspects
- ☐ The knowledge of grammar and vocabulary
- ☐ Listening tasks and activities
- ☐ Speaking and pronunciation sections

**N.B.**

***\* Please check to see whether you have answered all the questions.***

***\* I need some volunteers to participate in the interview sessions (either face-to-face or by phone). If you are willing to participate, please tick the box, and your participation would be highly appreciated. ☐***

***Thank you very much for your help!***

### **Appendix 6 Teachers' Interview Questions**

1. What do you think of the current English exam? What do you think are the positive and negative aspects of this exam? How, if at all, it could be improved?
2. To what extent is the new exam different from or similar to the previous one?
3. What language area do you think most important for students to learn? Why?
4. What other constructs do you think should be tested in the exam? Why?
5. Do you think that students can do well in the exam without taking coaching lesson or studying too much for it?
6. Regarding classroom instructional time, how often do you talk in class? Please give a percentage. In this regard, could you estimate how much time do you usually spend on practicing exam-related activities in class?
7. Have you observed any changes in your students' behaviour since the implementation of the revised exam such as asking exam-related questions?
8. Have you observed any changes in your own instructional behaviour as a teacher since the implementation of the revised exam?
9. Do you train your students on the tasks of the exam? (if yes) How?
10. Have you observed any (other) changes in your classroom, which you feel, were a result of exam change? (if so) could you please describe and explain them?
11. Has the implementation of the new exam led to your using any different teaching methods? (if yes) How are they different? Did you notice this with lower-level grades? Why did you think this change happened? Or (if no) why do you think your teaching methods did not change?
12. Has the implementation of the new exam led to your using any different teaching materials? (if yes) How are they different? Why did you think this change happened? Did you notice this with lower-level grades? Or (if no) why do you think you did not resort to any teaching materials other than the existing one in your classroom teaching?
13. (if yes to any of questions 7 – 12 above) Think back on the answers you have just given, do you think these changes are the direct result of the implementation of the revised exam
14. Would you like to add anything concerning the influence of the revised exam on EFL teaching and learning in Libyan schools?

***Thank you very much for your time!***

### **Appendix 7 Inspectors' Interview Questions**

- 1 How many schools and teachers do you visit during the school year?
- 2 What do you see as the major reasons for the Authority of Education to implement the new English exams? Are these reasons reasonable? Why?
- 3 Could you explain why and how the test was designed in the way it has been?
- 4 What are the key changes that you have perceived in the revised exam paper with comparison to the previous exam?
- 5 What are your views concerning test quality? In your opinion, what characterises should a good language test have? What criteria should good language tests meet? And how can you judge whether a language test is good or not?
- 6 As known that the final English examination is prepared and administered by the inspectorate in association with the examination board, have you ever been involved in the preparation of such examination before? If yes (or you will be given the opportunity to write such exam) what strategy will/have you use/d, or whether will you be or were obliged to follow set procedures?
- 7 What procedures are used for creating new forms of the test each year?
- 8 During your observation visits to grade 9 teachers in schools and as a result of exam change, have you observed any changes in teachers' teaching and testing practices? (if yes) what changes you observed and how and why? (if no) why?
- 9 Have you noticed teachers using mock exams as class activities?
- 10 Have you observed that teachers use teaching materials other than the existing one which has been designed by the PCE specifically for schools?
- 11 Regarding Q9 – 11, are the same answers applicable to grades other than grade 9?
- 12 (If yes to any of questions 9 – 12 above) Think back on the answers you have just given, do you think these changes are the direct result of the implementation of the revised exam?
- 13 Is there anything you would like to add?

***Thank you very much for your time!***

### Appendix 8 COLT Classroom Observation Scheme

Time	ACTIVITIES & EPISODES	PARTICIPANT ORGANISATION			CONTENT			STUDENT MODALITY							MATERIALS USED							
		Whole Class	group	Indiv.	Language										Type	Purpose						
					S ↔ S/C	T ↔ S/C	Different tasks	Same task	Different tasks	Same task	From-grammar	Form-vocabulary	Form-spelling	Function	Others	Writing	Reading	Speaking	Listening	Visual	Audio	Written

School: ..... Grade: ..... Teacher: ..... Visit No. ....

(adopted from Spada and Frohlich 1995)

### Appendix 9 Observation Note-taking Sheet

School: ..... Grade: ..... Teacher: ..... Visit No. ....

Unit and page No. .... Start at: ..... Ends at: .....

<i>Time/duration</i>	<i>Exam-related materials</i>	<i>Test-taking strategies</i>	<i>Info. about the exam</i>	<i>Others</i>
	<b>F1</b>	<b>F2</b>	<b>F3</b>	<b>F4</b>


➤ **F** = feature or category

## Appendix 10 Ethical Approval to Conduct and Pursue the Thesis Stage



**University of  
Leicester**

School of Education

21 January 2011

Mr Abdulhamid Mustafa Onaiba  
90 Bobbers Mill Road  
Nottingham  
NG7 5GZ

EdD Office  
162-166 Upper New Walk  
Leicester. LE1 7QA  
Tel: +44 (0) 116 229 7511  
Fax: +44 (0) 116 229 7527

Director  
Professor Janet Ainley  
21, University Road,  
Leicester, LE1 7RF, UK

Dear Abdulhamid,

### *Doctor of Education: Research Ethics Review*

Project Title: *Investigating the Washback Effect of a Revised EFL Public Examination on Teachers' Instructional Practices, Materials and Curriculum*

Further to the submission of your Research Ethics Review form, I can now confirm that your research project has been formally approved by the Research Ethics Review Committee.

We wish you every success in your continued studies.

Yours sincerely,

**Dr Alison Taysum**  
**Postgraduate Tutor (EdD)**

Cc Dr Glenn Fulcher

**Appendices in the accompanying CD**

Further appendices (appendix 11 – 34) are contained on the CD ROM attached to this thesis in PDF format.