University of Huddersfield Repository

Ahmed, Ahmed A. M.

Washback: Examining English Language Teaching and Learning in Libyan Secondary School Education

Original Citation

Ahmed, Ahmed A. M. (2018) Washback: Examining English Language Teaching and Learning in Libyan Secondary School Education. Doctoral thesis, University of Huddersfield.

This version is available at http://eprints.hud.ac.uk/id/eprint/34799/

Washback: Examining English Language Teaching and Learning in Libyan Secondary School Education

Ahmed Abubaker Massuod Ahmed

A thesis submitted to the University of Huddersfield

in partial fulfilment of the requirements for

the degree of Doctor of Philosophy

The School of Education and Professional Development

The University of Huddersfield

April 2018

Copyright Statement

1. The author of this thesis owns any copyright in it (the "Copyright") and he has given the University of Huddersfield the right to use such Copyright for any administrative, promotional, educational and/or teaching purpose.

2. Copies of this thesis, either in full or in extracts, may be made only in accordance with the regulations of the University Library. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.

3. The ownership of any patents, designs, trademarks and any and all other intellectual property rights except for the Copyright (the "Intellectual Property Rights") and any reproductions of copyright works, for example graphs and table ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.

Acknowledgements

Abstract

This thesis critically analysed the influence of the Libyan public examination on English Language Teaching (ELT) and on learning strategies and practice in secondary school classrooms. It investigated the washback of the Secondary Education Certificate Examination (SECE) on ELT teachers and explored practitioners' perceptions and practices. It also examined washback on students' perception, motivation, learning strategies and outcomes. Weir's socio-cognitive framework for test validity was chosen as a conceptual framework for its capability of conceptualising appropriate evidence on how testing constructs (policy & design) are operationalised and interpreted (use) in practice. This thesis describes an interpretative qualitative case study research conducted in the south west of Libya. Data were generated through interviews, classroom observation and document analysis. Semi-structured interviews were conducted with two ELT teachers and inspectors, a school headteacher and a representative of the examination office. Group interviews were also conducted with a number of the final year secondary school students within the research context. All the research data were analysed using a thematic data analysis.

The findings reveal that the lack of alignment between the focus of the English SECE and the objectives of the curriculum had a significant adverse effect on the Libyan ELT teachers, inspectors and students' perceptions about the aim and the value of ELT in the school education and their role within the policy as well as on Libyan school students' motivation. The study participants held the perception that developing language skills is not the aim of teaching English in Libyan secondary school since these skills have never been assessed in public examinations despite their integration in the curriculum. The Libyan ELT teachers and inspectors prioritised the aim of completing the curriculum through the use of traditional approaches of teacher-centred and Grammar Translation Method (GTM) and teaching to the test rather than meeting the pedagogical objectives of ELT or implementing the Communicative Language Teaching (CLT) proposed in the school curriculum. The Libyan on-going conflict, the public examination policy, and teachers' lack of assessment literacy were also significant on classroom testing. Students were largely passive in English classrooms as teaching focused on the SECE. Accordingly, students utilised different learning strategies to cope with the teaching such as prioritising the translation of textbook texts, relying on rote-learning, engaging in test-preparation activities and developing

test-taking strategies. Evidence accumulated through this study clearly indicates that Libyan students' experience of the public examination had a significant effect on their attitudes, perceptions and choice of learning strategies. This finding represented an important implication for developing the socio-cognitive framework for test validity.

The public examination strategies improved the Libyan secondary school students' examination performance in the SECE but not their English learning outcomes. The examination content and format as well as a social acceptance of cheating all have a significant effect on students' performance in the SECE and threaten its score validity.

Table of Contents

List of Figures

List of Tables

List of abbreviations

AFL ………………………………… Arabic as a Foreign Language
BEC ………………………………… Basic Education Certificate
BECE ……………………………… Basic Education Certificate Examination
BERA ……………………………… British Education of Research Association
BMET ……………………………… Beijing Matriculation English Test
CEE ………………………………… College Entrance Examination
CEFR ………………… Common European Framework of Reference for Languages
CET-4 ……………………………… College English Test Band 4
CLT ………………………………… Communicative Language Teaching
CSEPT …………………………… College Student English Proficiency Test
ECER……………………………… Exam for Certificate Exit Requirements
EFL ………………………………… English as a Foreign Language
ELT ………………………………… English Language Teaching
ENE ………………………………… English National Examination
ESL ………………………………… English as a Second Language
ESP ………………………………… English for Specific Purposes
FCE ………………………………… First Certificate in English
GEPT ……………………………… General English Proficiency Test
GTM ……………………………….. Grammar Translation Method
HKCEE …………………………… Hong Kong Certificate of Education Examination
HSSC …………………………..…. Higher Secondary School Certificate
IELTS ……………………………… International English Language Testing System
INUEE …………………………… Iranian National University Entrance Exam
ME …………………………………. Ministry of Education
MENA ……………………………… Middle East and North Africa
MHE ………………………………. Ministry of Higher Education
NMET ……………………………... National Matriculation English Test
SEC …………………………………. Secondary Education Certificate
SECE ……………………………… Secondary Education Certificate Examination
TEFL ………………………………… Teaching English as a Foreign Language
TELC ………………………………… Test of English Listening Comprehension
TESOL …………………………Teaching English to Speakers of other Languages
TOEFL …………………………….. Test of English as a Foreign Language
UE ……………………………………. Use of English

1 Chapter One: introduction

Introduction

This thesis examines the influence of the school leaving public examination on English language teaching (ELT) and learning in the Libyan secondary school context. This examination is a high-stakes test since its score is used to determine the future academic study of Libyan school-leavers. The phenomenon of testing influence or washback has been widely considered in the international context, especially in Europe, Asia and Canada, however, it has not been considered in the many developing countries such as the context of the Middle East and North Africa (MENA) and in the Libyan context in particular. The ultimate purpose of washback research investigates the role of testing in promoting positive teaching and learning outcomes and it can be used as a lever of change in an educational context. This thesis, therefore, gives an account of the effect of the Libyan Secondary Education Certificate Examination (SECE) on perceptions and teaching and learning strategies in English language classroom within the education system.

The thesis is of a particular significance for the area of educational assessment and washback research in particular. It discusses washback phenomenon from the under-investigated context of MENA in the literature which could also be applied to different developing countries.

This chapter presents the background of the research study, aims of the study, research questions, rationale of the study, reasons for choosing the topic and outlines the structure of the thesis.

1.1 Research background

Diplomatically isolated for over 30 years, the rapprochement between the international community and Libya kindled the development of the educational system during the 2000s. Libya implemented wide educational reforms in the last two decades, particularly in language teaching, in order to develop the traditional pedagogy practices, which characterised the education system. This development was interrupted by the so-called Arab Spring from 2011, and the on-going conflict in the country.

In 2000, Libya implemented a policy reform for developing English language education. For example, the Ministry of Education reformed the English language

curriculum with the aim to integrate the communicative teaching approach throughout the school educational system, focusing on developing the students' communicative abilities (Orafi & Brog, 2009). The English textbooks were changed to revamp classroom practices, in contrast to the traditional use of audio-lingual and grammar translation approaches in English language classrooms within Libyan schools for over half a century (ibid). Nevertheless, the language policy reform did not include the language assessment approach used in the school education system. The language assessment continued examining pupils' and students' linguistic knowledge of the studied textbooks during the academic year and did not focus on the communicative skills introduced in the ELT curriculum.

In 2009, the Libyan public examination approach was reformed using selected-response items across all school subjects. Accordingly, the final examination of the English language subject comes in the form of multiple-choice questions instead of traditional wh-questions (questions which begin with what, when, where, whom, which, whose, why, and how). The reformed examination approach was classified as a developed assessment practice in the Libyan education system. The reformed examination system aimed to minimise the risk of the prevalent phenomenon of cheating in Libyan public examinations, as reported by several research studies (e.g., Alhmali, 2007 and Onabia, 2013) and to score the answer sheets electronically and hence disseminate the results as quickly and as transparently as possible. This reform was described by the Ministry of Education as an implementation of the modern assessment processes (e.g. automated scoring) to improve the efficiency and reliability of scoring in the public examination.

However, many research studies revealed that the adoption of the communicative approach has not yet had the desired outcome. Since Orafi & Borg (2009) investigated the realities and intentions to implement the reformed English curriculum, most of those studies continue to reveal the failure of teachers to implement the CLT integrated into the English language curriculum. After studying English for more than six years, school-leaving students' English language proficiency has been frequently described as poor, even though they have passed the school and public examinations (Alhodiry, 2016; Altaieb, 2013; Dalala, 2014; Hammadi, 2013; Omar, 2014; Rahimi & Alavi, 2017; Soliman, 2013). However, research studies have not provided clear evidence of school-leaving students' English language ability or competence those in

their research. That was described by Rose (2015) as a continuing challenge for language education practitioners and researchers in the future.

Various factors related to cultural, teacher-related and contextual aspects have been identified as having a major impact on the deficiency of ELT teaching and learning. However, English language teachers have always been considered as the source of deficiency in ELT education within the Libyan school education system. They have been considered as having a lack of awareness about the function of English in its cultural context; lack of confidence in themselves; and inadequate knowledge and fluency of English (Abidin et al., 2012; Abusrewel, 2014; Ahmed, 2010; Aloreibi & Carey, 2017; Khalifa & Shabdin, 2016; Krayem, 2013; Mohsen, 2014). Nevertheless, the role of public examination policy and approach, and its washback influence on ELT curriculum, teaching and learning, and the related effective factors have not been considered in secondary school education. Most of those research studies built their arguments based on the ELT teachers' implementations of the proposed teaching approach away from considering the teachers' and students' perceived goals of the ELT classroom teaching and learning practices. It was important to investigate the teachers' and students' perceived goals of the classroom practices since most of the participant teachers in those research studies were teachers of the final year of the Basic and Secondary Education levels (More discussion in 2.2.2 and 2.1.2).

Therefore, this research study examines the washback effect of the Secondary Education Certificate Examination (SECE) public examination on the teaching and learning process and its outcomes in the Libyan context. It investigates how the public examination policy influences the ELT practitioners' practices in the classroom context. It also explores how those classroom practices influence the students learning strategies, practices and learning outcomes in the secondary school education. Although this research does not aim to provide a statistical measurement of students' language competences, however, it does aim to fill the gap mentioned by Rose (2015) by providing a description of secondary school students' learning outcomes through their reflection on their learning and performance on the SECE compared with their perception of their language knowledge and ability.

Moreover, there has not been empirical evidence available to demonstrate whether the reformed public examination approach achieved the intended washback by policy

makers or if it mediated different types of consequences. Therefore, this study focuses on these issues as that contributes to the originality and significance of this research in the Libyan context in particular and to washback literature in general.

1.2 The aim of the study

The aim of this study is to investigate the influence of the public examination policy and practice on the ELT teaching and learning process and outcomes in Libyan secondary school education. It aims to explore whether the recent reformed public examination approach has achieved the intended washback and whether any unintended consequences emerged affecting the ELT practitioners' and students' perceptions, motivations or practices. It also aims to raise awareness amongst those whose work may be affected by the test (e.g. students, teachers, inspectors, administrators, materials developers and policymakers) towards the importance of considering the role and effect of high-stakes examination policy and approach in the curriculum innovation process. So, the objectives of this research study are to:

- identify the influence of the public examination policy and approach on language classroom teaching and learning practices as well as the learning outcomes;

- record teachers' and learners' perceptions of the effectiveness/role of the Secondary Education Certificate Examination (SECE) approach in promoting Libyan students' communication abilities;

- gain a deeper understanding of EFL practitioners' perceptions and practices related to the public examination design;

- raise awareness of the importance and nature of the impact that assessment policy and practice have on the last curriculum innovation in Libya.

1.3 Research questions

In order to achieve the aims specified above, the study addresses the following questions:

1) To what extent does the Libyan public examination of the SECE have an impact on motivation, perceptions and attitudes towards English language teaching within secondary school education?

2) To what extent does the public examination of the SECE influence English language teaching, assessment and learning strategies in secondary school classrooms in the on-going conflict context of Libya?

3) How does the Libyan public examination of the SECE influence the students' English learning outcomes within secondary school education?

1.4 The rationale of the study

There are important areas where this study makes an original contribution to the context and to the literature. First, with the growing concerns and investigation of the barriers to developing the school students' communicative competences in the context of ELT in Libya, there is still a lack of understanding of the effect of the public examination policy and practice on English language teaching and learning. There has been a dearth of empirical research examining the potential role and influence of high-stakes testing of the SECE on Libyan ELT practitioners, students and classroom practices. This study is the first to investigate the washback effect of the secondary school public examination policy and approach on English language education in the Libyan context. It is also the first to examine the ELT classroom practices in regard to the perceived goals of school ELT classroom stakeholders (inspectors, teachers, head teachers and examiners) and participants (students) of their practices. It is hoped that the findings of this study will provide valuable information to the policymakers and educational practitioners at different educational levels to improve the quality of ELT education within the Libyan education system. The findings from this study could have important implications for developing or revising the language assessment practices, to enhance teaching and learning towards the objectives of the ELT within the Libyan education system and accordingly promote development in students' language learning and outcomes. In addition, this study could form a valuable source (baseline data) for future research studies about washback, either for the study context or for broader Libyan contexts.

For the literature, this study will contribute to the debates about the complexity of the washback phenomenon from a different under-researched context of the Middle East and North Africa (MENA). Sheehan & Munro (2017) stated that assessment is a situated practice, and it is important to investigate the nature of washback in different educational cultures and backgrounds in order to broaden our understanding and

cognition of the complexity of washback in a global context. This study is valuable since it adds new insights to the literature about ELT practitioners' perceptions and practices as a result of the curriculum innovation and public examination change. The need for such enquiries is raised by researchers (e.g. Alderson & Wall, 1993; Bailey, 1996; Shohamy, 2016; Tsagari & Cheng, 2017; Wall, 2005) in order to gain more understanding of the complex nature of washback in language education. This project provided an important opportunity to advance the understanding of the nature of the interaction between different factors that could mediate washback. By using a validity framework to guide this research, it could also add to the literature about washback in relation to the validity of the standardised examination, a top-down policy tool, and its socio-cultural and educational impact on ELT practitioners and test-takers' perceptions, attitudes and motivations in the broader study context. The findings should make an important contribution to understanding of the theoretical perspectives on the validity of language testing in contexts.

1.5 Reasons for selecting the topic

Mainly, I choose the thesis topic due to my interest in education and research, specifically in the field of English language teaching. I have always been fascinated by working in the field of ELT, as a school teacher, university lecturer and a teacher trainer. The research topic can be considered as a further development of my dissertation research project conducted during my study of the MA (TEFL) programme (Ahmed, 2010). The research was about the implementation of CLT in language classrooms in the Libyan secondary school context.

My experience of working within language education has also driven this research. My experience in taking different roles within the field of ELT in the Libyan context encouraged me to focus on the role of public examination in making different decisions regarding classroom teaching and learning practices. I have been intrigued by the variety of teachers' and students' attitudes and motivations towards English language teaching and assessment. I thought of how these perceptions and the motivation of teachers might be learnt and developed. These thoughts informed my first proposal of the current study.

Another reason for choosing this topic was its relevance to professional practices within the Faculty of Education where I work. As an ELT practitioner, I was also

motivated to conduct this study by what I saw as a clear need for such work in the field of EFL teacher education, especially in my context where such factors are investigated less. ELT education within school education is a controversial subject in Libya, among not only the practitioners but also the public. As a pre-service and in-service teacher trainer, investigating the importance of the role of public examination policy and practices in the educational context would help in evaluating and developing the practices of teacher training with the study context. The findings of the study could also help teacher trainers to build a comprehensive plan with other practitioners for developing ELT practices within the study context.

1.6   Composition and structure of the thesis

This thesis comprises nine chapters. Following this introduction, chapter two describes the research study context. It presents an overview of the Libyan education system from the perspective of its structure, the Ministry of Education, the assessment system, admission to higher education.

Chapter three looks critically at existing literature about washback research, including a detailed explanation of its definitions, nature, scope, mechanism and its relation to the test validity. Then the chapter discusses the washback effect in the education context, which is concerned with washback on teachers' perceptions and teaching practices, including teaching approach, classroom activities and classroom assessment, washback on students' perceptions, and attitudes, their learning motivation, and learning behaviour and strategy; factors influencing washback which deal with factors related to context, teachers and test design. A systematic search was applied to relevant literature, and this will be evident in the literature review chapter.

Chapter four discusses the research theoretical framework of situated washback within Weir's (2005) socio-cognitive framework for test validity. It starts with a justification of the relevance of the socio-cognitive framework to the current study and why it was chosen from different washback theories, models, and frameworks. The essential aspects of Weir's (2005) socio-cognitive framework for test validity, such as test-takers' characteristics, context validity, cognitive validity, scoring validity, consequential validity and criterion-related validity will be discussed.

Chapter five presents the research methodology, starting with a justification of the chosen philosophical paradigm and of the methodological strategy. It describes the sampling techniques and the participants; identifies ethical issues; discusses data generation methods and process; describes the thematic analysis techniques adopted in the study, and discusses the trustworthiness of the study.

Chapter six presents analysis and discussion of the findings in relation to the participants' perception of washback and the SECE. It discusses the participant perception towards the relationship between curriculum, assessment and pedagogy as well as assessment policy and students learning motivation. It also discusses the participants' perceptions towards the practices of the SECE and its efficiency in assessing ELT outcomes in the Libyan school context.

Chapter seven presents analysis and discussion of the findings in relation to the washback on classroom experiences. It discusses the washback on classroom teaching approach, classroom activities and classroom assessment. Then, it discusses the washback on classroom learning strategies and English learning outcomes.

Chapter eight draws upon the discussion of how the teaching and learning practices (discussed in chapters six and seven) affected the students' learning outcomes in relation to Weir's (2005) socio-cognitive framework for test validity.

The final chapter concludes the thesis, tying up the various theoretical and empirical strands in order to offer general conclusions and includes a discussion of the study's limitations and implication of the findings to the context and to future research into this area.

2 Chapter Two: the research context

This chapter provides an overview of the context of Libyan education. It describes the structure of the Libyan education system, with a reference to the objectives of school levels and the prescribed curriculum. Then, it explores the structure of the Ministry of Education. Finally, it discusses the examination system with the focus on public examinations in terms of objectives, design, content and its role in the Libyan context.

2.1 An overview of the Libyan education system

In recent decades, Libya has achieved remarkable results in its education system due to the high literacy rate of 99.9% within the youth of both genders (Rose, 2015). The Libyan public report for UNESCO stated that the rate of enrolment for compulsory education (from Year 1 to Year 9) was approximately 98% in 2008 (ME, 2008). Where education is not compulsory, the enrolment rate in secondary school placed Libya as the 11th highest level of enrolment in secondary school and university education in the world. These achievements contributed to Libya's presence amongst the top fifty highest-placed countries in the world according to the Human Development Index (HDI) (Rose, 2015). However, the reliability of these figures can be questioned after the conflict emerged in 2011 as since then the Libyan education system has been in turmoil. Since then, many children have experienced difficulties accessing school in different parts of Libya. Children are either displaced from their homes or their schools are occupied by those who have been displaced from their hometowns. The on-going conflict has also had an influence on the capacity of the education system to perform well, including the damage and destruction of many schools, closure of educational institutions due to violence and the dismantling of the educational infrastructure and equipment (Rhema & Miliszewska, 2012).

Education is free for Libyan citizens from school education until first-degree university education. School education is mandatory from the age of six up to the age of 15 (ME, 2008). Arabic is the language of instruction within the education system. The Libyan education system is divided into two main categories: Basic and Intermediate Education. The following figure (Figure 1) explains the structure of the Libyan education system.
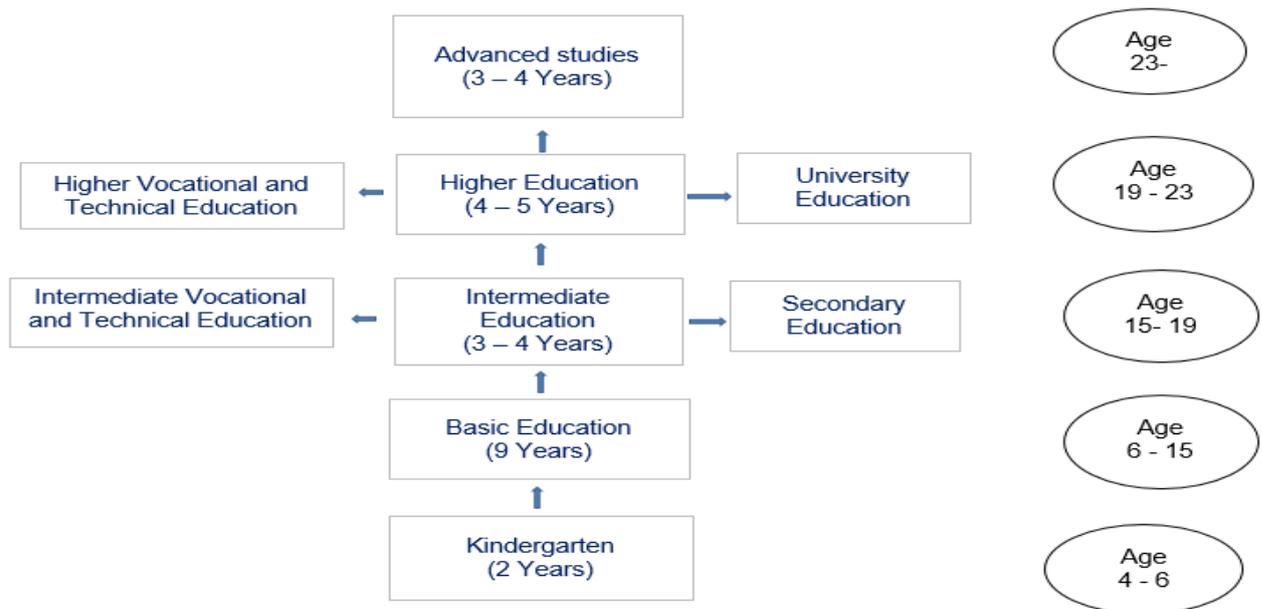
Figure 1: the structure of the Libyan education system

2.1.1   Basic Education

Starting at the age of six, Basic Education covers the first nine years of school education. During the first two years of school, pupils study Arabic, including reading, literature and writing (dictation, spelling and script); Islamic education; mathematics and arts. Pupils study more subjects as they move from one school level to the next. The inclusion of teaching English language starts from Year 5. In Year 9, students undertake public examinations to gain the Basic Education Certificate (BEC) before continuing to the Intermediate Education which includes Secondary Education to prepare for a university education or technical and vocational education route. The students' ability to continue on each route is mainly controlled by the overall grade on their Basic Education Certificate. As this study is conducted in the secondary school education, more details of the Secondary Education route are discussed in this section.

2.1.2   Secondary Education

Secondary education is the end of the cycle of school education in the Libyan education system. It starts in Year 10 and finishes in Year 12. For students to be able to join Secondary Education, they have to gain at least an overall grade of 65% (classified as "good") on their Basic Education Certificate. The first year of secondary school is general for all students. Moving to Year 11, students can choose either a

literary or a scientific specialisation. Students in the literary division mainly study history, geography, philosophy and sociology; whereas students in the scientific division mainly study physics, chemistry, biology and mathematics. English language is a general subject for both divisions, however, every division has its prescribed textbooks, sharing the same objectives with a different specialised content (See Appendix 1).

At the end of Year 12 (the final year of Secondary Education), students undertake a public examination known as the Secondary Education Certificate Examination (SECE) to gain the Secondary Education Certificate (SEC). The successful completion of this examination enables students to seek their chosen university discipline according to their specialisation and the average score on their certificate. Thus, the SECE has great value in the Libyan context – socially and professionally. This value and the importance of the Secondary Education Certificate demonstrated the significance of the phenomenon of washback of the public examination under study.

2.2 The Ministry of Education

The Libyan education system is centralised and run by the Ministry of Education, maintaining uniformity of schooling in the country. The Ministry sets the standards for a public curriculum so that all Libyan children study the same syllabus and the same course material. The Ministry determines funding and budgets, including building schools, employing teachers and supervising them, and preparing and distributing teaching and examination materials to schools. The Libyan schools have a basic infrastructure and most of them have been provided with advanced facilities like functional science, languages and computers, laboratories. To facilitate the administration and regulation, the country is divided into regional educational departments to supervise and maintain the educational process within those regions (Altaieb, 2013). These departments include offices of inspection, finance, assessment and administration.

Since the tension between the Libyan political parties in mid-2014, and the ongoing conflict, Libya has had two governments: the Interim Libyan Government is located in the city of Tobruk (situated in the east of Libya near the Egyptian border) and the Government of Public Reconciliation in Tripoli. The educational practices described in this research study are situated under the supervision of the government of Tripoli.

The Ministry of Education has several administrative departments which supervise the different educational levels and processes such as the departments of Secondary Education, Examination Office and Educational Inspection Department. The following figure (Figure 2) explains the construction of the administrative departments of the Libyan Ministry of Education.
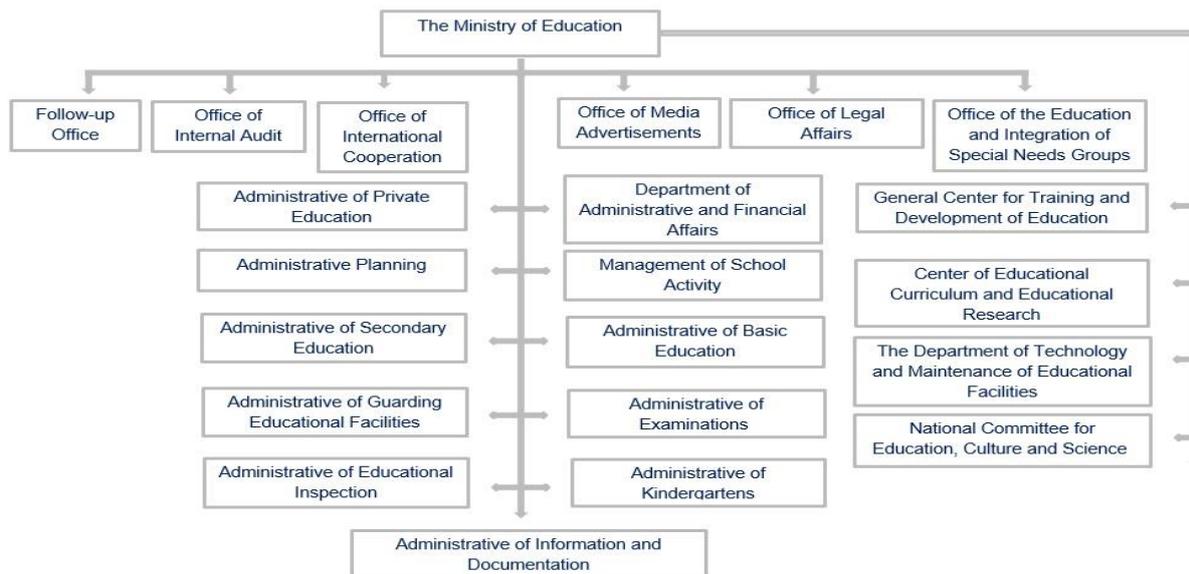
Figure 2: the construction of the administrative departments of the Libyan Ministry of Education

The Centre of Educational Curriculum and Educational Research is responsible for making decisions regarding curricular development, including goals, objectives and time allocation for each subject, based on international standards and experience. The educational specialists from the Higher Education (HE) sector supervise curriculum developments. Therefore, school education practitioners do not take part in these developing processes.

The Educational Inspection Department is the key department accountable for ensuring the implementation of the educational policies according to structures and regulations approved by the Ministry of Education (ME, 2012). Educational inspectors supervise the classroom teaching and assessment, evaluating teacher performance and competence with respect to the implementation of the annual teaching plan of the academic year (Shakuna et al., 2016). Practically, inspectors conduct periodic and random educational inspection tours to educational institutions in order to evaluate school teachers' practices. Therefore, the inspectors of English language teachers

were involved in the current research study as they are the only responsible body for evaluating teachers' practices in the Libyan education system.

The Examination Office is responsible for the management and the administration of mid-term or periodical examinations and of the final certificates examination (ME, 2012). It is the main controller that organises the personnel examination committees in charge at the public, provincial, municipal and school levels with the Ministry of Education. The regional examination offices supervise and monitor schools' examination through assigned teachers in each school. The following part of the chapter discusses the examination system, focusing on SECE in English.

2.3 Libyan education's assessment system

The goal of the examination system is to examine the outcomes of the educational process in which it aimed to identify the student understanding of the taught curriculum over the academic year (ME, 2008). The assessment criteria are determined by the examination office of the Ministry of Education. The examination scheme includes formative classroom assessment and summative written examinations at the end of each semester. These summative tests are completely constructed from the course textbooks content used in the classroom. All assessment practices are conducted by the classroom teachers, except in the final years of basic and secondary education. The total score of final year secondary students in the English language subject is the total of the classroom-based assessment and final public examinations as shown in the following table (Table 1).

Table 1: The division of students marks in the final score of the SECE

Type of assessment	Assessment criteria	Total marks
Classroom-based assessment	Classroom participation, reading and speaking skills and reading comprehension	14
	Notebook and workbook organisation	12
	Classroom tests	26
	Periodical examinations	28
Public examination(SECE)	final year examination	120
The total score on the SEC		200

2.3.1 Classroom-based assessment

 The classroom-based assessment is conducted by the classroom teachers under the supervision of the school, the inspectors and the regional examination office. The criteria of this assessment are assigned to the curriculum teaching and assessment plan issued by the Centre of Educational Curriculum. Classroom-based assessment represents 80/200 of the total final score of the subject. Those marks are divided between the formative evaluation of students' reading skills and comprehension, speaking skills in English, engagement in classroom activities, the organisation of the students' notebooks and their homework books of the ELT curriculum. This evaluation represents 26/200 of the total score. Besides, the final year students undertake monthly classroom tests and two school periodical examinations during the academic year, which represents represent 54/200 of the total score of the SEC. The remaining 120/200 of the total final score of the English language subject is devoted to the final public examination, which is discussed in the next subsection.

2.3.2 Public examinations

The Secondary Education Certificate Examination (SECE) is a large-scale public examination, administered at the end of the final year of secondary education in all school subjects. It is a test, quantifying students' knowledge in the school materials of the final year of secondary school education. The decision about students' graduations is based on their overall scores of the classroom-based assessment (discussed in 2.3.1) and on the public examination in each subject. It is a high-stakes test since it is used as a basis to make decisions about student graduations and their academic future at the next educational level in the Libyan education system.

Students to be successful must achieve at least: 40% of the public examination marks (SECE), i.e. 48/120; and 50% of the total final score on the SEC for the English subject which is100/200). Students who fail to pass on their first examination attempt have a chance to re-sit the examination. If a student fails the re-sit exam, then s/he must repeat the final year before proceeding to the next education stage. The SECE's questions are prepared under the supervision of the Examination Department of the Ministry of Education through a committee of experienced inspectors from the General Education Inspection.

The following subsection explains the nature, format, design and content of the reformed public examination of the English language subject.

2.3.2.1 Test format

Although the public examination has been described as an electronic examination by the Ministry of Education, students still use a pencil and paper test to undertake the examination. It should be highlighted that "electronic examinations" refer to grading method where the exam papers are marked electronically. The examination comes in two documents: the question sheet and the answer sheet (see Appendix 2). The cover sheet of the questions contains the student's name, registration number, city, school, subject matter, the allotted time and the examination test instructions. This information is written in Arabic so that students know what to do and to reduce the risk of their misunderstanding as the Ministry of Education highlighted. Even the instructions for each type of question are translated into Arabic. As students need to transfer their answers on a separate answer sheet, the means of transferring answers to the answer sheet are also explained on the cover page. The answer sheet is monitored by computer software to determine the total score in more reliable way and efficiency in the process. Although the most recent electronic marking simplified the marking process of the public assessment, it raises issues about the effect of the standardised multiple-choice question on the examination setting and educational process. Therefore, this research investigates the washback of use of the public examination approach in the Libyan education system.

2.3.2.2 Test design

The traditional examination approach of integrative construct-response items (WH-questions) used in the SECE was replaced with an exclusively based selected-response question, including multiple-choice, true-false and matching items. It is claimed by the Ministry of Education that the test design enabled examiners to cover all the contents of the prescribed textbook. The Ministry also aimed to facilitate the clarity of the students' responses to the public examination questions to ease the marking process. The reformed examination strategy and design were also intended to ensure the reliability of test scores by diminishing opportunities for cheating, which in turn would threaten the validity of the test. The examination questions are designed to be different in terms of question order from one student to another within the same

examination room. This strategy is used to diminish the risk of cheating among students according to the Ministry of Education. The possibility that students would take too much time to match their exam paper questions to another student's exam paper makes them less likely to cheat.

2.3.2.3 Test content

The Libyan public examinations are textbook-oriented or have a "syllabus-content approach" according to Hughes' (1989) classification. This means that the examinations are constructed from the verbatim linguistic features of the school textbooks of English language subject. The current SECE also neglects to assess the language skill integrated into the curriculum of English language subject. Therefore, the English language SECE is considered as a form of achievement test rather than a proficiency test. The examination questions are designed to assess the Libyan students' comprehension of the reading texts and grammar of the textbooks. The reading questions are usually based on specific pieces of information or on dates of some described in the reading texts of the school coursebook. These questions come in three different types: true/false, multiple-choice and matching. Since it appears that there is a mismatch between assessment and curriculum objectives, this study investigates whether the teaching and learning practices are test aligned (test-oriented teaching) or curriculum aligned (syllabus-oriented teaching). The study would examine the effective factors that influence the policy implementation in the Libyan education system, and whether the effect of the assessment practices has any type of washback on these practices and their learning outcomes.

2.4 Admission to Higher Education

According to the Ministry of Higher Education (MHE), there are two factors controlling the Libyan students' admissions to university education: specialisation and the overall score of the SEC (MHE, 2010). First, the secondary school graduate students can join faculties and specialisations compatible with their specialisation in secondary school education. For those who studied the path of literary division, they can study law, arts and media, languages and education (in literary departments). For those who studied the science division, they can study medicine, nursing, engineering, computing, science, medical technology, business and finance, astronomy, agriculture and education (in science departments). However, the students would need to fulfil the

requirements of each of these specialised fields in the category of the overall score of the SEC. These requirements differ from one specialised field to another.

Second, the overall score of the SEC also plays an important role in determining which faculties students can attend. The increasing preference for university education in the social context has created the demands of respectively high overall scores on the SEC. The minimum overall score required to enrol in a university programme is 65% (Clark, 2004); and it changes according to the academic discipline. For instance, students need at least 65% (classified as "good") to study education and arts, 75% (classified as "very good") to study engineering and low at least 85% (classified as "excellent") to study medicine or any other medical science. It is also possible that some departments within the field require certain levels of scores in specific subjects in the SEC. For example, some medical science programmes require an overall score of 74% in chemistry, biology and English. For students who obtain overall scores below 65% on their SEC, they can still attend the higher technical and vocational institutions. It is possible that such requirements could be a major factor influencing the social attitude towards technical and vocational education in the Libyan context. The majority of Libyan students choose secondary and university education as the technical and vocational education route is undervalued and not encouraged by most parents as in many Arab countries (Al Heeti & Brock, 1997; Rose, 2015).

2.5 Chapter summary

This chapter aimed to give the reader a description of the characteristics of the Libyan education context. It provided an overview of the Libyan education system. It presented a brief introduction about the rates of access to Libyan education system and its role in expanding the literacy rate and developing human resources in the country over the past twenty years and the current situation of these achievements. Then, the chapter described the division of the Libyan education system and key departments of the Ministry of Education related to this study. It also sheds the light on the assessment criteria in the Libyan education system and a description of the public examination approach. Finally, it discussed the admission to higher education.

3 Chapter Three: review of the research literature

Introduction

This chapter starts with introducing the concepts and the scope of testing influence in the educational research and the position of this study in the literature. Then, it focuses the concept of washback, the focus of this thesis in detail. It discusses the definitions and types of washback, and different models explaining the mechanism of washback in different educational contexts. It also highlights the nature of the washback on teaching and learning strategies and outcomes. This chapter ends with a discussion of the effective factors influencing the washback in context.

3.1 Washback, impact and consequences

The concept of test influence is central to the complex relationship between testing, teaching and learning. There are different terms used to describe the phenomenon of test influence in contexts such as test washback, test impact and the test consequences. Whereas test washback and impact has been described according to the scope of the test influence, the term consequences is used to refer to the nature of the test influence in context.

This research study is located within the scope of washback. The concept of washback has been attached to a wide range of areas directly linked the influence on the micro level (e.g., individuals, practices in classrooms, school and policies) within the educational system (more discussion in section 3.2). It is also argued that the test influence has an effect on two levels (Andrews et al., 2002). Washback can have an influence on the micro level within the educational process and context, and on the macro level throughout society. It also covers the effect of general opportunities for people within the society (Bachman & Palmer, 1996). However, there has been a growing debate about the scope of testing by differentiating between washback and impact in the field of language education research (e.g., Hughes, 2003; Taylor, 2005, 2006; Weir & Milanovic, 2003). Those researchers who support distinctions between test washback and impact claim that test impact is completely focused on the social effect of a test in the society whereas washback is related only to the educational process of teaching and learning or the educational context (Bachman & Palmer 1996; Milanovic & Saville, 1996; Saville, 2009). They claim that impact takes into consideration the cultural and social implications and consequences such as

behaviour, motivations, attitudes, qualifications and standards. Shohamy (2001, 2011, 2013) also argued that the test impact includes the ethical concerns of discrimination, fairness and equal opportunities such as the political use and abuse of language testing. For instance, high-stakes examinations are indicated as powerful tools that play a central role in creating social classes, immigration policy, determining knowledge, affecting the distribution of wealth, shaping language policies, teaching and learning, educational admissions policy and career employment opportunities (e.g. Cheng & Fox, 2013; Shohamy, 2001, 2013). Zhan & Andrews (2014) also support this distinction by referring to the need to broaden the view of the testing effect, as the term washback has been mostly connected to classroom behaviour and practice, and our understanding of the testing, and especially the high-stakes examinations' influences beyond classrooms and educational context is still restricted.

This study stands in congruence with other researchers who argued that the distinction is not necessary as the term washback at the micro and macro level includes the wider and narrower influence of the test (e.g., Andrews et al., 2002; Cheng & Curtis, 2004; Hamp-Lyons, 1997). Cheng and Curtis (2004) also declared that it is a matter of researchers' preferences for using the terms as both refer to different areas of the same phenomenon. Washback can be classified as one dimension of impact in which a test affects the educational context, including teaching and learning, and curriculum, while the broader test influences are considered as a test impact. It would also be difficult to separate them distinctly due to the complex relationships and connections between the micro and macro contexts. While introducing justifications of the terms "washback" and "impact" along with "consequences", Cheng & Fox (2013) and Taylor (2005) acknowledged the high complexity of impact and washback phenomenon. Due to the complicated relations between the features of the educational context mentioned by Saville (2009) and Shohamy (2001), it would be challenging to connect educational practices to a single feature such as tests. As the impact of a test on society is mediated by its interaction with dynamics beyond classrooms and educational systems, Saville (2009) argued for the usefulness of conceptualising the notion of impact to include the micro and macro contents in order to understand the complex connection between institutions, individuals and their relation to the systems in society. This is of essential importance as it affects the understanding of the notion of the nature and mechanism of the test influence in its context. This would help in

gaining a more comprehensive understanding of the phenomenon of the test influence through the emphasis on the crucial role of cultural and social perspectives, ethical issues and its in/directly effect on classroom practices by influencing the perceptions, attitudes and motivations of teachers and learners.

Test consequence is a term also used to include the concepts of both washback and impact. A test consequence is a common term used in general educational research, incorporating notions of both washback and impact from the field of language education and applied linguistics (Rea-Dickins & Scott, 2007). It refers to the planned and unplanned effects of the implementation or use of a test on educational policies and practices and on the society, whether of being negative or positive facets. (Cheng & Fox, 2013). This identification might be traced back to the test validity framework established by Messick (1996) and Weir (2005) where both concepts of washback and impact are located within the *consequential validity*. Consequences also take into consideration the influence of test uses which is connected with the interpretations of test score (Cheng et al., 2015; Weir, 2005). These two variables are situated and of significant influence in political, cultural, social and educational contexts. For that, it has been argued that washback researchers are required to work in partnership with other researchers in different fields such as innovation reforms, policy analysis and socio-cognitive studies in order to understand the scope and power of a test in context. This research adopted the socio-cognitive framework for test validation as a theoretical framework to achieve this goal (more discussion in chapter four).

For the purpose of this PhD research, however, the differentiation between the different terms used to refer to the test influence is not considered. The term washback is used to refer to the educational and social perspectives of the influence of the SECE in the study context. The terms test impact and consequence are used synonymously with washback to represent the focus of the study. It can also be argued that "test impact" in itself is too broad for doctoral research. Therefore, the three main types of influences dealt with in this study are the SECE effect on the students' and teachers' attitudes, perceptions and motivation; the influence on ELT teachers' classroom instructions and assessments, including the choice of the curriculum content and the selection of material resources; and the impact on the students' language learning behaviour, practice and learning outcomes. Thus, the following part of this chapter discusses washback from three perspectives. The first section discusses washback's

definition, nature, mechanism and its relation to test validity. The second section presents a discussion of the existing literature about the washback effect on the educational process on language teaching and learning. Finally, a discussion of the factors affecting washback in context is presented.

3.2 Washback

The introduction of the washback concept generated a discussion differentiating between "backwash" and "washback". The term "backwash" is used to demonstrate the influence of public examinations on the behaviours, attitudes and motivation of learners and teachers about their teaching and learning practices in general education (Cheng & Curtis, 2004). While Alderson & Wall (1993) argue that no semantic or pragmatic difference exists between them, Spolsky (1994) believed that backwash covers unintentional side effects of examinations, but not the intended effects when the main testing goal is controlling the curriculum. However, this can be seen as oversimplified descriptions of the phenomenon. Cheng, et al., (2004) revealed that the testing effect is highly complex and yet the backwash and washback phenomena can be manifested from one test.

The term "washback" is used in this research to refer to the phenomenon for two reasons. First, as it has been shown that the testing effect is highly complex and I am not convinced by the oversimplified descriptions differentiating between backwash and washback. Second, washback is the common term used in the literature of applied linguistic research where this research belongs. Alderson (2004) observed that washback is the term mostly used in literature and published research in the language assessment field, especially in the UK, where the current research is documented.

3.2.1 Definitions of washback

Washback has been defined differently according to which area a test can affect or upon whom it has an effect. The definitions cover a wide range of educational aspects and participants as the following definitions show:

- The effect of an examination on classroom practices including curriculum and teaching methods and students' learning strategies (Biggs, 1995).
- The effect of an examination on classroom behaviour and practice of learners and teachers (Alderson & Wall, 1993; Messick, 1996).

- The effect of an examination on curriculum and materials (Spolsky, 1994).

- The link between learning and testing (Shohamy et al., 1996).

- The effect of an examination on the attitudes, motivation and general behaviours of learners, parents and teachers (Pearson, 1988).

- The effect of an examination on teaching and learning (Bailey, 1996; Hughes, 2003).

- The unintentional or intentional function and direction of curriculum change on learning and teaching as a result of a change in public examinations (Cheng, 2005).

Although these definitions demonstrate a wide range of influences, it seems that they have been devised to suit individual cases but not a comprehensive conceptualisation of the related phenomenon. This study's working definition of washback is elicited from the three definitions that cover the macro and micro contexts mentioned by Alderson & Wall (1993) and Cheng (2005). That is because these definitions refer to washback as the test's effect on motivation, attitude, behaviours and practices of participants in educational and social contexts. Therefore, these practices can involve any changes in teaching, learning, curricula or contents. In this case study context, the term "washback" is used to refer, specifically, to the extent to which public examination of the SECE has an influence on education practitioners and EFL teachers' perceptions and their teaching and assessment testing practices as well as students' motivation, attitudes and learning behaviour and outcomes. The effects are considered as a washback only if these effects can be linked to the SECE in the context.

3.2.2 The nature of washback

In general education, the term washback used to be connected with unintended negative side effects of testing on educational processes. However, since the eminent publication of "*Does Washback Exist?*" by Alderson & Wall (1993) and "*Washback in Language Testing: Research Context and Methods*" by Cheng, Watanabe & Curtis (2004), it has been recognised that the washback phenomenon is highly complex. According to Andrews (2004), washback does exist and it has consequences, which could be negative or positive, intended or unintended, or could be multi-directional. Therefore, it is naive simply to suggest that tests always have a negative impact on education. Rather, the nature of washback is mediated by a number of variables in

educational and social contexts. These variables have mostly been linked to the type of feedback that is determined by "what the examination measures" (e.g. Cheng, 2005).

Positive washback occurs when testing practice promotes high-quality teaching practice (Alderson & Wall, 1993; Cheng, 1997; 2005; Ferman, 2004; Li, 1990; Shohamy et al., 1996). In the classroom basis, positive test washback encourages teachers to complete the curriculum more thoroughly within the assigned time, motivating and encouraging learners to enhance learning sufficiently. It can also manifest if tests are understood as valuable activities supporting teaching and learning, engendering constructive processes (Cheng, 2004). Hence, if the examination is relevant to the course content and purpose, the washback is classified as positive; otherwise, it is likely to be negative. For that reason, high-stakes examinations have been utilised by policymakers to accomplish teaching and learning goals and processes, or to introduce new curricula or textbooks. However, negative washback occurs when the examination system leads to constrictions on the language teaching and learning processes and outcomes. The literature generally affirms that testing encourages teachers to narrow the curriculum (Cheng, 2005; Cheng & Curtis, 2004; Qi, 2004). The common negative washback effect observed internationally is "cramming" (teaching/learning to the test) rather than teaching with the aim of developing learners' language acquisition (Alderson, 2004). Teachers and students have widely been observed focusing on mastering discrete knowledge that is tested rather than developing real-life language knowledge and skills in different contexts, whether on their own initiative or due to teachers' practices of teaching to the test. Consequently, that decreases their linguistic performance and negatively affects students' attitudes and motivation of language learning (Shohamy et al., 1996). Test anxiety among students and teachers has always been linked to this practice, whereas here the main motivation of learning English is to progress to the next education level rather than to acquire the language itself. On a broader level, the negative side of washback is that decision makers can employ tests to suit their political agenda to control educational systems (Cheng, 1997; Green, 2007; Pan, 2009).

Moreover, washback is described as neutral when the test does not show any specific effect. Some researchers (e.g. Bachman & Palmer, 1996; Davies, 1997; Hughes, 2003; Qi, 2005; Taylor, 2005) used washback as a neutral term in their work. They

acknowledged that the washback of a test could be multi-directional where a test could indicate as intended and have positive effects and at the same time as unintended which could be negative and differ for each individual learner. For instance, Bailey (1996) contended that any test can have either negative or positive washback to the extent that it promotes or impedes the accomplishment of educational goals held by learners. Similarly, Wu (2014) explored the notion of *washback intensity*, asserting that moderate levels of washback indicate a neutral direction. The intensity of washback was found to be in a linear relationship with the test's stakes: the higher the stakes of the test for participants results in stronger washback (e.g. Spratt, 2005; Watanabe, 2004). However, there may be other factors contributing to the existence, nature and intensity of washback in every specific context (Andrews, 2004). Therefore, it can be argued that the identification of the washback direction should be determined according to the comprehensive explanation of the factors that affect, or are affected by a test socially and educationally. Thus, the washback of the SECE is taken from a comprehensive perspective in this research that includes its stake and its effect on the course learning objectives and outcomes in the context. It would also be anticipated from how the SECE affected the participants' aims and objectives of their practices since those two goals could be different. Besides, the focus of this study agrees with Wang's (2010) argument of the need to redefine and conceptualise the phenomenon of washback rather than continue debating whether it is positive or negative, especially since every context has its particularity and specificity of characteristics, which shape the nature of a test washback. Hence, this research focuses on introducing comprehensive explanations and discussion of the washback effect on language teaching and learning practices rather than classifying washback as being either positive or negative, although it does highlight the adverse effect of washback on educational processes and learning outcomes.

3.2.3 Washback and test validity

Washback has been connected with test validity in the field educational research and of applied linguistics. The origins of the debate over the extent to which the notion of validation should take into account test consequences can be traced back to Messick (1989). Messick developed an argument associating washback on educational and social context with the concept of test validity. He placed washback in the theoretical concept of *construct validity* as a consequential aspect. Messick stressed that if the

washback effect can be directly linked to using or introducing a test, then it could be a part of its validity. Similarly, Weir (2005) built a socio-cognitive framework for test validity and placed washback in the concept of *consequential validity*. He argued that assessment should concern with validity on evidence-based principle. This necessarily involves providing data relating to *context-based* and *theory-based validities*, together with the various *reliabilities or "scoring validity"* (more discussion is presented in chapter four). However, Weir considered a test's validity as multi-faceted and residing in its score. Therefore, he argued for the need to obtain various types of evidence on validities in order to strengthen claims for a test score validity. These evidential bases are complementary aspects for test interpretation and not alternatives, in which any single validity aspect may not be looked upon as better or superior to another. Then, if there is a deficit in any validity, it might raise questions regarding the trustworthiness of the interpretation of the test scores. Accordingly, both Messick and Weir were concerned with the test's ability to produce accurate scores that reflect a particular area of candidates' ability. Then, cases can be made that different tests are valid if it shows similar results over time using various versions or administrated at different times. However, Bachman (2015) stated that these theoretical definitions of validation are too limited, and efforts to simplify the concept result in evidence that, in itself, is too simplistic and ignores context (including stakeholders). However, this research utilises Weir's socio-cognitive framework to examine the score validity from test-takers' perspectives as well as other validities which are concerned with how teaching and learning practices are managed by different stakeholders in order to prepare students for the test to cover Bachman's argument.

Others, such as Shohamy (2013), argue for taking into consideration different aspects of the test's use in context as part of its validity. She mentioned that the quality of tests is judged not only by how well tests measure the language knowledge from a measurement perspective, but rather by investigating the agenda of introducing these tests on policy, and ethics and their implications for individuals and groups at the educational, social and political levels.

It would be difficult to identify which teaching and learning practices relate to washback of contextual factors or test misuse. Hence, Messick (1996) advised that research practitioners should focus on seeking test validity through design rather than seeking

the nature of its washback as a sign of its validity. Despite the debates regarding the effect of test design and use on teaching and learning, no empirical research has produced evidence to distinguish the consequences of test design from those of test use be investigating such effect within one model simultaneously (Xie & Andrews, 2013). The effect of the test use and test design on language teaching and learning if it exists, it is likely to be different from each other. The adoption of Weir's socio-cognitive framework for test validity as a theoretical framework can be considered as an attempt to fill that gap in the literature. The current research considers the SECE design as a part of the contextual validity of Weir's socio-cognitive framework by investigating the effect of the task setting on teaching and learning. The SECE's validity will rely on the participants' reflections and interpretations of its score in an educational and social context to elicit the validity within the research's theoretical framework. Therefore, the discussion of the validity of the SECE can be considered as a further step to introduce comprehensive evidence for test validity in the literature from the perspectives of test use and design.

3.2.4 The mechanisms of washback

In exploring the complex mechanisms in which washback occurs in actual language teaching and learning environments, a growing body of research in the field has led to the emergence of different theoretical hypotheses and models to illustrate that mechanism. These research studies have developed methodological approaches and methods of washback research in the field (e.g. Alderson & Wall, 1993; Bailey, 1996; Burrows, 2004; Hamp-Lyons, 1998; Messick, 1996; Saville, 2009; Shih, 2007; Wall, 2000; Watanabe, 2004). This part of the thesis discusses some key theoretical developments to build an account of the current research scope and methodology in relation to the research trends in the field.

3.2.4.1 Washback hypothesis (Alderson & Wall, 1993)

Challenging the existence of washback in the field of language assessment in 1993, Alderson and Wall proposed the first hypothesis regarding washback influence on language teaching and learning in language education research. The hypothesis was based on a review of washback research in different contexts (the Netherlands – Wesdrop, 1982; Turkey – Hughes, 1988; Nepal – Khaniya, 1990; Sri Lanka – Alderson & Wall, 1993). They suggested that a test would influence teachers and teaching (e.g.

in terms of teaching content; rate and sequence; degree of depth; method of teaching), learners and learning (e.g. what and how learners learn; rate and sequence; degree of depth), and learners' and teachers' attitudes to language teaching and learning. Despite the intrinsic simplicity of their washback hypothesis, it gives an account of areas that could possibly affect or be affected by a test other than test design. The washback hypothesis can be considered as a re-conceptualisation of washback. It was a guide for further research in the field to explore the real existence of washback. Alderson and Wall's research can be considered as an innovation research in language assessment, which led to a significant development agenda for washback research and constructing washback studies in language assessment research. Moreover, this hypothesis gives an account of the importance of developing data collection methods to use ethnographic methods of research to understand classroom practices.

Criticising the simplicity of the washback hypothesis, Alderson & Hamp-Lyons (1996) argued that consideration should be given to the extent of washback. They argued that different types and amounts of test washback could occur on a number of learners and teachers but not on others. They added one more hypothesis in which the level of washback effect would vary according to the stake of the test, its relevance to practice and the extent to which ELT practitioners are willing and able to innovate in preparing appropriate tests for that. Therefore, Alderson & Hamp-Lyons (1996) admitted the need for further research to discover the way other aspects interact with a test in context, locating washback into innovation theory. Likewise, Hughes (1993) proposed a model consisting of participants, process and product. Hughes revealed that for valuable washback effects to occur, some conditions should come together: the importance of success for learners on the test, which should also be the teachers' aims; participants' familiarity with the test; the availability of test expertise (including materials, syllabus and teaching methodology); and the availability of crucial resources for successful testing preparations. Hughes' work was the basic theory behind constructing the first basic model of washback in language assessment by Bailey (1996) which is discussed below.

3.2.4.2 The basic model of washback (Bailey, 1996)

Bailey (1996) proposed the first cohesive model of washback in applied linguistic research. Combining reviews of the washback hypotheses and Hughes' model, Bailey

set out what is known as the basic model of washback to provide feedback on how washback functions. Her model was a summary of the previous research and established the foundation for subsequent washback studies in the field. The model theorises that a test influences participants' attitudes and perceptions of their educational practices (including test-takers, language teachers and practitioners such as advising and supervision, test development, and curriculum planning). Next, this leads to an impact on performance in the processes stage (materials development, syllabus design, teaching methodology, test strategy) and finally affects the products (learning outcomes: knowledge and developed skills and its quality; test scores). She also suggested that washback might operate in two ways: washback to learners and washback to programme providers.

However, the teaching method was not identified as having a direct effect on a test in Bailey's (1996) model. In test preparation classes, it is strongly believed that the aim could be improving scores by developing testing strategies rather than developing language knowledge and skills, which directly affect a test. Although, in her model, Bailey did not give a clear explanation of the process phase, she followed Alderson & Wall (1993) and Messick (1996), declaring that if classroom practices are caused by the use or introduction of a test, then this can be counted as washback. Bailey's (1996) model, as well as Alderson & Wall's (1993) hypothesis, ignored the macro context factors of social perceptions and use of a test.

In addition to that, Cheng (1997, 1998) developed the notion of "washback intensity" to refer to the extent of the washback on a number of areas of teaching and learning that influenced by a test. Each of the areas was studied in order to understand the mechanism of washback and its function on the participants, processes and products. This washback might have been mediated by the reform of a major high-stakes examination within a specific educational context. Subsequently, Watanabe (2004) theorised five dimensions of washback: intensity, specificity, intentionality, length and value. He highlighted aspects that facilitate the process of washback that may be influenced by testing, such as status-related, test-related and stakeholder-related factors, which could influence teaching and learning. Derived from the results of previous research, Watanabe also outlined different factors that affect the process of washback, including macro and micro context factors, personal factors, prestige factors and test factors.

It can be noticed that Watanabe (2004), with Alderson & Hamp-Lyons (1996), worked on specifying parts of the washback process. They demonstrated that washback varies in its nature and effect on participants due to contextual, personal and test factors. Their view supports Alderson & Wall's (1993) view of the importance of considering classroom behaviour and practices when investigating washback. Watanabe's dimensions can be seen as the general characteristics of washback in context. Likewise, Burrows (2004) highlighted that the beliefs, assumptions, knowledge and reactions of participants and teachers caused by assessment practices are of great importance and concern. Burrows assumes that behavioural models proposed for educational change in different educational fields (e.g. Markee, 1997; McCallum et al., 1995) might be applicable to washback. That is because the use of these models in curriculum innovation studies examining the impact of educational change on teaching and learning. Elaborating on the work of Burrows (2004), by describing further factors that play indispensable roles in the washback mechanism, Shih (2007) proposed a further model. The model was advised to investigate test washback on language teaching and learning in the school education context.

3.2.4.3 Washback model of students' learning (Shih, 2007)

Shih's model was built on his research exploring the General English Proficiency Test (GEPT) influence on teaching English in Taiwan. Based on the available understandings of washback, Shih integrated factors reported in different empirical research studies (test, contextual and teacher factors) influencing the quantity and mechanism of washback on class teaching systematically together. The model aimed to enable practitioners to recognise potential washback effects. In contrast with washback models and hypothesis mentioned above, Shih's model was the only one attempts to illustrate washback on language learning. It was also different as it takes into consideration the socio-economic factors besides educational, contextual and personal factors, which illustrate the subjectivity of the nature of washback in context. It shows the connections between different facets to enable researchers to observe the dynamics of washback. Further to that work, Shih (2010) developed another model to reveal the GEPT's washback complexity on school and departmental policies, explaining the different factors that teachers had to consider before implementing the

educational requirements, such as school, social and educational factors, and students' and parents' factors.

In both models, Shih tried to concentrate and gather the aspects affecting students' psychologies and learning through extrinsic factors (educational and school, socio-economic, friends, colleagues and family, and personal), intrinsic factors (differences of individuals, characteristics and perceptions) and test aspects (content, difficulty, stakes etc.). His models have conveyed a different view of what constructs and mediate washback in the field. Shih deems that these three factors are interrelated and act simultaneously to make a difference to the washback of a test on students' language learning and psychologies. The washback of a test on psychology was argued to affect the student language learning content, time, strategies, motivation and anxiety. This washback continues to affect the results of the test and subsequent language learning. Accordingly, the results of the test can, directly and indirectly, counter-influence the three primary factors, particularly those which are intrinsic. Since washback on student language learning is one of the main focuses of the current research, Shih's tentative washback model of student learning is a major reference point in this study.

Since Messick (1996) advised the importance of considering *impact by design* into the development of language assessments research, Green (2007) linked together test design with the participant, process and product dimensions to explore a language test influence throughout language teaching processes to test outcomes. He devised a comprehensive washback model that integrated washback intensity, variability and direction. He built his model based on the work of Hughes (1993) and Bailey (1996) by adding a number of new features that were unspecified by the latter, particularly focusing on the process aspects as discussed below.

3.2.4.4 The comprehensive model of washback (Green, 2007)

Green (2007) expanded his model to outline the relationships between test design, participants' values and motivations, test resources and the importance and the difficulty of the test. He declared that test design is a key determinant of washback direction mediated by participants' values, motivations and resources. It is also a major determinant of washback variability. Moreover, Green considered the perceived difficulty and the importance of the test as a significant determinant of washback

intensity. This argument forms a further contribution to washback models to developing a deeper understanding of the process category that has been left unspecified in the previous models. As the social and individual differences that govern participants' values and motivations, as well as the test importance and difficulty, participants may be affected by a test in different ways in the same general context. Green (2007) underlined the importance of detailed analysis of the test instrument and an evaluation of its congruence or overlap with the planned curriculum. He argued that the overlap between the test's main construct and its characteristics could make it difficult to differentiate between the two concepts. Although this model was linked to the test design as an effective factor to influence the participants, process, product and the intensity, the social aspect is neglected as the model does not focus on how perceptions and value within the educational context are influenced or affected by test design. The neglect of the social aspects could be attributed to the fact that he built on the work of Hughes (1993) and Bailey (1996) by adding a number of new features, which were unspecified by the latter, particularly focusing on the process aspects. Such a critique made it difficult for the model to act as this research's theoretical framework since the social status of the SECE is considered as a major focus in the current research study. However, the model gives an account of the test design as a key part of researching washback in the study context.

Recently, Cheng et al. (2015) argued that investigations into washback are never straightforward because of the complexity of the involvement of multiple factors that affect, and are affected by, the test policy and practice in context. That complexity leads practitioners to integrate further theoretical developments to explain the key factors, theories and conceptions that contribute to the understanding of washback and how this knowledge could be further examined. Therefore, washback research has been integrated into other theories in the fields of social science, learning and teaching, and psychology. That integration aimed to explain the different factors that mediate, or are affected by, testing in order to uncover their effect on washback. For instance, Xie & Andrews (2013) employed the Expectancy-Value Motivation Theory (by Jacobs & Eccles, 2000) to develop a statistical model tracing how test-takers' perceptions of the design and use of a test influence their preparation practices. They found that learners' perceptions of both test use and test design have an impact on preparation practices to a certain extent. In addition, Weir (2005) considered the social

and cognitive aspects and social psychology to explain human behaviour in social contexts and mental interactions. He followed recent theories in the social sciences, as well as measurement in applied linguistics. He constructed a socio-cognitive framework for test validity, which has been adopted as the research's theoretical framework (more discussion of the reasons behind its adoption is presented in chapter four). This step has raised the researchers' and other educational policymakers' awareness of the need for collaborative work to decide when to link a test's consequences to its validity.

Although the development of washback's hypotheses and models has formed perspectives regarding educational, social and test design and construct aspects that should be taken into consideration, the need for further research has been identified by almost all researchers in the field. From educational perspectives, Wall (2000) declared that the main concern of teaching and testing which include, learners and learning, requires greater research attention, as the direct impact of testing is on them. Similarly, Cheng & Fox (2013) stated that investigation is needed to explore the influence of testing on students' cognitive (e.g. strategy use), psychological (e.g. anxiety and motivation) and emotional status (e.g. perception). Alongside focusing on the impact of test design on washback, test construct, test method, test use and the test function on students and on learning processes and outcomes should also be considered in further research (Green, 2007, 2013). Therefore, this study incorporated learners' perspectives and practices. From the social perspectives, the need to examine tests exists as not only rarefied academic exercises, rather they are influential societal instruments that have far-reaching mutual influences with ideology, policy agendas, diversity, politics, economics and education. Tests need to be examined in relation to the strong impact they have on various societal dimensions beyond the declared intentions of just measuring knowledge (Shohamy, 2013). These dimensions will also be considered in my research. The information, gained from different practitioners in the research context, will generate recommendations that take collaborative account of different practitioners' perspectives.

The following two sections discuss different observed washback effects and the effective factors that have been observed in the global context, using the previous washback hypothesis, models and frameworks. That is to highlight the general

knowledge of the high-stakes examinations impact in order to conclude how the current research will add to this knowledge.

3.3 Washback influence on the educational process

There is no doubt that high-stakes tests have a powerful impact on teaching and learning strategies in classroom contexts (Cheng, 2017; Cheng & Watanabe, 2004; Shohamy, 1993; Tsagari, 2017). Understanding washback, to identify the quality of the high-stakes examinations and the effectiveness of its implementation can contribute significantly to the success of educational authorities in improving the quality of education, especially in public education systems. Hence, a growing body of evidence in the recent research studies has been conducted, investigating the washback of the high-stakes examinations, especially public examinations of the local language teaching and learning in different contexts. Table 3 in Appendix 3 summarises the recent washback research in the field. Examples of these studies are: Austria (Mewald, 2016), China (Wang, 2017), Iran (Damankesh & Babaii, 2015; Nikoopour & Farsani, 2012; Pour-Mohammadi & Zainol Abidin, 2011), Pakistan (Adnan & Mahmood, 2014; Aftab et al., 2014), Bangladesh (Hoque, 2011); Turkey (Toksoz & Kilickaya, 2018; A review) and Indonesia (Furaidah et al, 2015; Hartaty, 2017; Sukyadi & Mardiani, 2011).. Nevertheless, only a few research studies have investigated washback in the context of MENA, such as Egypt (Gebril, 2017; Gebril & Brown, 2014; Gebril & Eid, 2017), Tunisia (Hidri, 2016), Yamen (Tayeb et al., 2014) and Libya (Onaiba, 2014). It should be highlighted that Onabia's (2014) study was concerned with the effect of the public examination moving from basic school education to secondary school education. This examination is different from the public examination, which is the subject of this study.

The topic of this research is concerned with the washback of the school-leaving public examination on teaching and learning English in the Libyan secondary school context. This study aimed to contribute to the development of the literature discussion of washback in general and in the MENA context in particular. The following section presents a discussion of the washback effect of high-stakes examinations on teaching (including teachers' attitudes, teaching and assessment practices) and learning (including students' motivation, learning strategies and learning outcomes) within

different educational contexts. It also discusses the effective factors affecting washback in context.

3.3.1 Washback on educational practitioners' and ELT teaching practices

A significant body of literature in washback studies gives much attention to the effect of language testing on teachers' perceptions and different aspects of their classroom practices. That includes the teaching approach, content of teaching and assessment methods (e.g. Alderson & Hamp-Lyons, 1996; Burrows, 2004; Cheng, 2014; Kılıçkaya, 2016; Pan, 2011; Smyth & Banks, 2012; Wall & Horak, 2006; Wang, 2017; Watanabe, 2004). These research studies indicated different washback types and intensity to which language testing prompts changes in teachers' attitudes and perspectives, and changes in their classroom instructions and assessments. Hence, this section is divided into three subsections to discuss washback on teachers' attitudes and perceptions, washback on teaching approach and washback on classroom assessment.

3.3.1.1 Washback on language teachers' perceptions and attitudes

Many of the research studies in the field have investigated teachers' perceptions of the washback of high-stakes examinations on their teaching and assessment strategies (Al-Issa et al., 2017; Cheng, 2004; Kennedy & Lui, 2013). Only a few research studies have investigated how the teachers' perceptions of high-stakes examinations mediate the washback or how the use of the high-stakes examinations affects the teachers' perceptions and attitudes towards their role in language teaching in context. As discussed earlier, Hughes (1993) stated that a test could affect the perceptions and attitudes of participants, such as teachers and other practitioners, which may then influence their practices (processes) and shape the language learning outcomes (products). For instance, Tsagari (2017) reported that Greek teachers were strongly in favour of the First Certificate in English (FCE), was recognised by the Greek state as a language qualification. By interviewing teachers who prepare students for the exam, Tsagari mentioned that the test affected the language teachers' perceptions regarding their role, as well as the students' perceptions. Tsagari (2017) reported that the teachers felt that their students expected them not only to teach the language use but also to prepare them successfully for the exam. In another study about the washback of the Beijing Matriculation English Test (BMET) using questionnaires and

interviews, Kennedy & Lui (2013) showed that both students and teachers perceive the aim of the English class of the final year was to prepare students for the BMET; and students' improvements in English skills are unimportant. Therefore, it is important in this research to address teachers' perceptions and attitudes towards the use of the SECE in the context to determine any effect of these perceptions on their processes and products as Hughes mentioned. It also considers those who their practices could influence the teachers' perceptions.

Moreover, research in language innovation has shown that teachers' perceptions are an important aspect of policy implementation and practice. The teachers' willingness to implement language policies was identified as being influenced by the teachers' beliefs and their situated goals, as well as the social and personal dimensions of classroom teaching (Borg, 2003). The teachers' perspectives are important for this research since the reformed assessment approach of the SECE was considered as innovative in the Libyan educational system. Wall (2000) mentioned that the rate of an innovation adoption is determined by many factors, including the characteristics of the innovation (such as the assessment approach in the study context), the context (classroom, school, the educational and cultural system), models for introducing change (top-down or bottom-up policy implementation) and the potential users. For instance, Spratt (2005) outlined that teachers' attitudes and perceptions towards examinations can either motivate them to willingly work towards integrating the educational innovation in their classroom practice or place stress on teachers because of the tension between ethical decisions and the pedagogical demands. In her earlier research about teaching Arabic as a second language, Shohamy (1993) and Shohamy et al. (1996) found that the teachers' negative feelings towards the Arabic test, considering it as insignificant in their context, affected policy implementation. Likewise, Hartaty (2017) investigating the impact of the 2015 public assessment policy implemented in Indonesia, found that the teachers were willing to make the change as they knew that the results of the national examination would not be used for graduation judgement. Using observations and interviews, Barnes (2017) also found that the Vietnamese teachers' beliefs and attitudes towards teaching practices in general English classes and TOEFL iBT preparation courses about what constitutes "good" teaching practices were different. Their attitude and beliefs were specifically

depending on the role and purpose of the course (test preparation versus general English).

Other studies have revealed that teachers' perceptions and professional stances towards high-stakes examinations determine their motivation and effort to improve pedagogy and the integration of changes into their teaching or assessment practice (Aftab et al., 2014; Thaidan, 2015; Toksoz & Kilickaya, 2018; Turner, 2006; Wang, 2008). In a survey with secondary school ESL teachers in Canada, Turner (2006) concluded that the teachers' perceptions were of a significant impact in transferring teaching and assessment practices of the curriculum, to be aligned with the high-stakes assessment system. In the same vein, teachers' perceptions towards the use of high-stakes examination scores were identified to affect teachers' professional values in different contexts. For instance, Gebril (2017) argued that scoring on school public examinations affects teachers' professional values in different MENA countries. She stated that the teachers believed that the failure of students to pass public examinations could be attributed to the lack of teachers' professionalism. Wang (2008) also mentioned that monetary reward and academic promotions affected the Chinese tertiary teachers' perceptions of language policy implementation using classroom observations and follow-up interviews in her research study. She found that the teachers followed the strict regulations from their own universities to prepare students for the College English Test Band 4 (CET-4) rather than teaching according to the structure of prescribed textbooks of the College English Teaching Syllabus. She highlighted that disobeying the university rules might adversely affect practical benefits such as monetary reward or academic promotions. Similarly, Cheng & Qi (2006) reported that Chinese English language teachers in the last year of senior secondary school felt that their students' scores on the National Matriculation English Test (NMET) affected every aspect of their professional lives. Tsagari (2017) also reported that the FCE test increased the teachers' accountability towards parents and employers as their professional status was judged by the yardstick of students' successes in the exam. Using a questionnaire and interviews, Shohamy et al. (1996), explored an introduced EFL and AFL policy in Israel and found that the teachers feel that the students' failure or success has its impact on them. They also highlighted the pressure to cover all the materials in the textbook for the exam.

Wang (2017) argued that fundamental changes in teaching practice can be maintained through essential changes in teachers' characteristic (e.g., beliefs, knowledge, attitudes and thinking) that inform their teaching strategies. To achieve this purpose, teachers need to be trained to enhance their understanding of the benefit and the purpose of tests in order to promote positive perceptions about their role in the educational system, especially when high-stakes examinations lead them to feel accountable to a number of stakeholders.

Consequently, these studies concluded that the teachers' attitudes towards different aspects of general education, including teaching and learning processes, are strongly affected by high-stakes examinations. For that, this research would have a similar focus to Tsagari's (2017), which takes into consideration the understanding of how teachers' perceptions towards the SECE promote washback in their context. Since the focus on the public examination is different from the focus on the English language curriculum objectives, the teachers' perceptions towards public assessment and the characteristics of its approach are important to investigate washback in context. Their attitude could be the starting point for identifying the existence of washback in the study context.

3.3.1.2 Washback on teaching approach

The relationship between high-stakes examination and classroom teaching practice has been extensively investigated in language education research (Damankesh & Babaii, 2015; Nikoopour & AminiFarsani, 2012; Safa & Jafari, 2016; Zhan & Andrews, 2014). A common observation through research in the field has reported that public examination usually leads to the phenomenon of "*teaching to the test*". Teachers were observed choosing teaching approaches and methodologies compatible with the aim of the test rather than implementing the curriculum's goal where developing language learners' communicative competence is emphasised. This phenomenon affected different aspects of daily classroom teaching methodology, content, time management and classroom assessment.

Literature on washback and teaching has shown the varied extent to which language tests induce changes in teachers' teaching approaches. For instance, Spratt (2005) believed that there has been an assumption that assessment affects teaching content and material but not teaching methodology and strategies. Such effect was observed

in the earlier research findings of Alderson & Wall (1993). They found that the revised examination of the O-level English exam in Sri Lanka had no practical impact on the teachers' teaching methodology towards integrating CLT as planned by the test designers. Likewise, Cheng (2005) found that the introduction of the Hong Kong Certificate of Education Examination (HKCEE) did not promote changes in the classroom teaching approach and methods. Although the test intended to integrate task-based approaches in classroom activities and practices, the test remained to dominate the classroom teaching approach and activities. Nevertheless, a number of research studies have shown that revised high-stakes examinations lead teachers to transform their classroom teaching practice. For example, Barnes (2017) approved that Vietnamese teachers dedicated their practices to teaching the examination's content and skills rather than developing language skills through communicative language teaching. For example, Shohamy (1996), discussed the Israel's EFL oral test, and Watanabe (1996) in Japan, have found positive washback effects on teaching methodology towards communicative language teaching with various degrees of intensity, depending on different factors. Likewise, Andrews et al. (2002) point out that the revised high-stakes tests of Advanced Supplementary (AS) "Use of English" (UE) oral examination led teachers to use more explanation techniques when engaging in certain exam tasks based on their empirical research in Hong Kong. Another example was seen in Ferman's (2004) research about a newly introduced oral matriculation test in Israeli high schools. Ferman declared that when students are required to perform speaking activities by a high-stakes test, teachers tend to shift their teaching focus from grammar-based teaching to more communication-oriented teaching in order to promote and upgrade the students learning of oral skills. Ferman did not use observation but utilised questionnaires, interviews (structured and non-structured) and document analyses but in this research. Such research design could challenge the certainty of the research claim. Hartaty (2017) investigated the effect of the 2015 public assessment policy on language teaching and learning in Indonesia using also questionnaires. Hartaty found that the new policy has changed the status of the national examination from high-stakes testing to lower stakes testing, however, it brought positive washback to English language teaching in the context. The teachers in Hartaty's study mentioned that the new policy released them from teaching students to pass the test as the main goal. The policy instead enabled the teachers to apply

most of the prescribed communicative language teaching practices and gave them the freedom to engage in different communicative activities and use different materials.

From the perspectives of washback on teaching content, there is an agreement between the research studies in the field that high-stakes testing affects classroom teaching content. Since teachers are seen as teaching to the test, they tend to adopt more test preparation activities in their classroom practice. Evidence for this emerged in the results of many research studies in the field (e.g. Aftab et al., 2014; Alderson & Hamp-Lyons, 1996; Barnes, 2017; Cheng, 2005; Green, 2007; Hamp-Lyons, 1997; Hayes & Read, 2004; Shohamy, 2001; Wall & Horak, 2006). For instance, in investigating the washback of the National Matriculation English Test (NMET) in China, Qi (2005) found that teachers tended to focus on linguistic knowledge of the test instead of focusing on language use. Using questionnaires and interviews, Qi found that the teachers work towards the immediate goal of raising scores contrary to the test designers' intentions. Teachers tended to spend a lot of time explicitly teaching test content and focus on test preparation activities and administering mock tests. In another research study conducted via a questionnaire, Chou (2015) also reported that drilling the grammar and memorising vocabulary are greatly emphasised in Taiwanese high school English classrooms because the English test content of the College Entrance Examination (CEE) purely based on discrete-point of these aspects and Chinese-English translation. According to Chou, these practices are popular in many Asian countries (e.g. Vietnam, Bangladesh, Indonesia and Hong Kong) with a similar public examination focus and assessment system. Shohamy (2001) has also discussed the introduction of a new national test on reading comprehension in Israel. She reported that teachers felt forced to used test-related materials in the classroom, rather than teaching reading in more integrated ways. Similarly, she also found that the introduction of an English oral test in the 12 th grade of school education caused the teaching of test preparation activities in classrooms. Therefore, the language teaching became limited to these very tasks, which were included on the test.

High-stakes examinations were also found to exclude the focus on developing the language skills of the learners in the classroom. This effect is compelling teachers to use lecture-based and teacher-centred approaches. Using interviews in their research, Aftab et al. (2014) found that Pakistani language teachers were ignoring the teaching of speaking and listening, in favour of reading and writing skills. The teachers

justify the negligence as they are mainly focused on preparing their students for the public intermediate examination to obtain the Higher Secondary School Certificate (HSSC). In Iran, Safa & Jafari (2016) investigated the effect of the public examination on leaving high school in the general educational system using a validated questionnaire. They found that EFL teachers spend most of their class time practising the material that is likely to be included in the third-grade final exam. They also neglected the communicative skills of the language that were excluded from the exam. Watanabe (2000) also reported that the teachers of the pre-college level used various test related materials to explain English words and their organisation patterns and cultural background. They also made their own materials from past exam papers in order to prepare their students for the Japanese university entrance examinations. Mogapi (2016) argued that teachers considered test-related materials as being a relevant source of content, which increased its use in schools. However, such materials potentially limit teaching and learning activities and the content in the syllabus. Saukah (2015) in his investigation of the washback of the Indonesian public exam of high school education found that the public examination has adversely affected the curriculum implementation, leading to its reduction of test coaching or teaching to the test. Au (2007) argued that the nature of washback induced on curricular control is highly dependent on the structure of the tests. This argument was based on his results from a study using qualitative meta-synthesis (template analysis) to analyse 49 qualitative studies to interrogate how high-stakes testing affects the curriculum. Similarly, Toksoz & Kilickaya (2018) found that tests lead to a narrowing of the curriculum in schools and make teachers and students master the format of these tests instead of acquiring the necessary skills for language learning in Turkey. They found that teachers tend only to deal with multiple-choice questions and they ignore the productive skills (writing and speaking) and the receptive skill of listening. These findings were based on the analysis of the studies (e.g., Külekçi, 2016; Özmen, 2011; Sayın & Aslan, 2016; Yıldırım, 2010) conducted on the washback effect of national language examinations in Turkey between 2010 and 2017, such as the Foreign Language Exam (YDS), Bachelor Placement Exam (LYS). Such teaching approaches could also be related to the linkage between the students score and the quality of teaching in context as identified in Middle East and North Africa (MENA) context. Gebril & Brown (2014) argued that the public examination scores tended to be linked to the quality of schools and educational improvement within an examination-

driven society such as MENA context and many Asian countries, where teachers considers a high quality school teaching is judged by student exam results.

The above discussion reflects Barnes' argument that effective practices in a test preparation course could help learners to pass the test although it may not espouse a communicative language approach or represent the best practices in language teaching. In the research of TOFEL and IELTS test preparation courses, Read & Hayes (2004) and Wall & Horak (2006) also reported that the teachers' reliance on materials of exam-preparation in their studies on IELTS and TOEFL were the main base of the courses investigated in their study in order to meet the expectations of their students. The current research study is among the few that have considered students' expectations as a factor of washback. Although the use of such material in TOFEL and IELTS test preparation courses could be considered as common due to the objectives of the courses, it can be a negative washback on language teaching classrooms where the focus is to develop the learners' language knowledge and abilities such as in the education sector. It represents the neglect of the language teaching course objectives of the focus on developing the students' language use. Weir (2005) mentioned that test preparation courses may have a positive effect on the candidates of being prepared for the linguistic and metalinguistic demands of the test, however, if the test promote the use of test-taking strategies that enhance performance without the association of developing the students' ability included in the test, then, there must be some concern. Thus, this study aims to find the relationship between the SECE design and content, the teachers' ways of mastering the prescribed English language curriculum and the classroom teaching approach. It is important in washback research to investigate the motive behind what and how the teachers teach, whether it is the test itself or other factors in the educational and social context.

The refocus on the teaching approach and content in favour of high-stakes examinations was also affected by the language classroom's time management. Devoting more time to test-related activities and instruction of test-taking strategies than activities to develop students' language abilities and competences was reported as a type of negative feedback of a test in different research. In a more recent washback study on the English National Examination (or in Bahasa Malay *Ujian Nasional*) in Indonesia, Furaidah et al. (2015) explored how the test influenced the amount of class time dedicated to preparing students for the test. Using in-depth

interviews and classroom observations, they argued that the only identified washback was the increase in time allocated for the tested subjects. They observed that teachers spent a large amount of time on explicit test preparation activities, such as teaching to the test's content and administering mock tests. However, they found that these practices were likely to be beneficial as they can potentially support the development of communicative competence as stated in the standards of competence of the English subject. Weir (2005) also mentioned that teachers tend to provide students with information regarding the test rubric to make students familiar with the nature of the test and clearly train them on the amount of time needed to spend on each item or part of the examination. According to this variable, students would consider tasks that involve an extensive planning as less difficult.

The literature review has shown that high-stakes testing may affect the teaching approach, activities and time management in different ways. However, still a number of research studies in different contexts (e.g., Aftab et al., 2014; Hartaty, 2017; Qi, 2005; Safa & Jafari, 2016), and some research in Toksoz & Kilickaya (2018), rely on interviews and questionnaires to build their arguments about the washback effect in the classroom context. Thus, more evidence that is empirical needs to be collected to demonstrate the nature of washback in contexts where washback has not yet been empirically researched as MENA context and in particular the Libyan context.

3.3.1.3 Washback on classroom assessment

Generally, the review of washback literature in relation to classroom assessment has shown that relatively little research has taken into consideration the influence of high-stakes exams on classroom assessment. Despite the role of classroom assessment in the education process, few research studies have yielded a negative impact of washback on the teachers' classroom assessments and evaluation procedures (e.g. Froetscher, 2013; Ghorbani, 2008; Hamp-Lyons, 1997; Nikoopour & AminiFarsani, 2012; Safa & Jafari, 2016). These research studies suggest that the effect on what and how teachers test is more significant than the effect on the teachers' methodologies. For instance, classroom assessment was usually considered as a tool for training students for the examination, and to evaluate the effectiveness of teaching practice towards enabling students to pass the examination. Using a questionnaire, Safa & Jafari (2016) found that EFL teachers' classroom assessment procedures and

evaluation formats were designed with similarity to the content and format of the high-stakes tests of the high school third-grade final exam in the general educational system of Iran. That was to maximise preparing their students for the test in advance. They declared that the teachers are not only "*teaching to the test*" but also "*testing to the test*" (Safa & Jafari, 2016). In her research of the washback of the Austrian school-leaving exams (Matura) on classroom testing, Froetscher (2013) found that teachers are restricted to the Matura examination approach; designing their own tests according to Matura using the previous exam papers.

Besides high-stakes large-scale assessment, there are other key factors that were identified as affecting classroom testing in different contexts. For instance, the significant effect of teachers' assessment literacy and professional development on their classroom assessment practices has been extensively researched (e.g., Alderson et al, 2017; Brown & Bailey, 2008; Fulcher, 2012, 2013; Sheehan & Munro, 2017). Bachman (2013) stated that teachers are widely found that they neither well trained in language assessment nor knowledgeable about it. Consequently, Bachman argued that it is necessary to understand the teachers' knowledge and needs about assessment literacy in the field of language assessment and how that influence their classroom assessment strategies and practices.

Another factor was identified by Gunn et al. (2016) who mentioned that school or education policies often instruct teachers to administer a certain type of test with little or no chance to give feedback regarding whether they feel that these tests are useful, valid or should even be given. A further factor observed is the teachers' experiences of assessment, which could not be altered through the professional education they have been exposed to (Alderson et al., 2017). Similarly, Sheehan & Munro (2017) captured the experiences and attitudes of language teachers regarding language assessment's literacy practices across Europe. They stated that assessment practices are rooted in past learning experiences, and through knowledge-sharing with colleagues.

Hence, it can be argued that there is a necessity for detailed analysis of the reliability and validity of classroom tests in washback research. This research focuses on the washback on classroom assessment to fill the gap in the literature regarding this issue to give a more insightful overview from the perspectives of assessment policy, ELT practitioners and students.

3.3.2 Washback on learners and learning

Washback was identified to have its impact on the perceptions and attitudes of language learners as well as teachers (Cheng et al, 2004; Wall & Alderson, 1993). These perceptions then influence the learners' practices and form their learning outcomes. In spite of the comprehensive amount of washback research, there is still little research investigating the washback effects on students' learning processes in relation to the work considering effects on teachers, teaching or programmes. Watanabe (2004) argued that more emphasis should be given to washback to learners in order to understand how teaching and assessment practices affect language learners' perceptions, motivations and their learning and outcomes. Hence, the current research considers the effect of the SECE assessment policy and approach on the students' learning perceptions, motivation and practices. The following sections are discussing the literature relevant to washback on learning in language assessment research.

3.3.2.1 Washback on students' attitudes and perceptions

Although Alderson & Wall (1993) hypothesised that a test will influence learner attitude, still there is a shortage in the focus on this area in washback research, according to Booth (2012). Few research studies in this area indicated a strong relationship between students' perceptions of a test and their goal of the educational context and learning outcomes (e.g., Cheng et al., 2011; Manjarrés, 2005; Rahimi & Nazhand, 2010; Shih, 2007). For instance, high-stakes examinations tend to induce a negative effect on learners' attitudes towards the language learning objectives and practices. Different research studies identified factors related to high-stakes tests such as their value, nature and content mediating such effect. While describing the washback effect of the English test on a public high school classroom in Colombia, Manjarrés (2005) mentioned the undesirable washback on language learners' perceptions and attitudes towards the value of a course. Such washback could result from the under-representation of communicative competence, especially when only linguistic competence (grammatical and textual competences and textual coherence) is included in a mandatory test. He based his argument on his research using interviews with teachers and students, observed classrooms, analysis of policy documents and used tests. Similar results were found by Spratt (2005) in her review of the empirical studies of washback, demonstrating that negative attitudes and

feelings of language learners towards their learning practices and motivation are usually generated by exams.

The test's value, nature and construct play a significant role in its impact on the students' perceptions and motivations. For instance, Aftab et al. (2014) found that the public examination format and content negatively affected the Pakistani school students' beliefs about the examination requirements and the objective of the language education within the school system. The students' attitudes towards English examinations suggested that they treat English as a subject and are not concerned with acquiring skills to become proficient in the language. In earlier studies, Cheng (1998) was the first to investigate the students' perceptions and their practices in relation to the HKCEE examination reform. She found a contradictory feeling among students towards the exam. On the one hand, they recognised the positive effect that the examination increased their effort in the study to achieve good scores, but, on the other hand, it does not accurately reflect their language ability. The study identified washback in the refocus of students' learning activities but not in students' motivation and learning strategies. Likewise, Weili (2010) highlighted that the reformed College English Test 4 Listening Comprehension Subtest (CET-4 LCS) showed positive effects on learners' attitudes towards the listening test, such as test format, design, scoring criterion, and reliability. Shih (2007) explored the impact of the General English Proficiency Test (GEPT) on learning in Taiwan, interviewing teachers, students, departmental heads and family members and conducted classroom observations to collect the research data, determined that the test content has a remarkable influence on the students' perceptions and their learning practices and strategies. Shih argued that the CET led students to frustration and demotivate them since the test content is different from the teaching curricula. Rahimi & Nazhand (2010) concluded from a questionnaire that IELTS preparation courses generated a positive washback on the Iranian students' perceptions of the expected learning outcomes in these courses. Since their aim was to pass the IELTS test, this effect was powered by the fact that these courses narrowed the range of skills that students need in order to develop their proficiency in English required for various academic and professional purposes. However, still, the use of the questionnaire as a single source of data can be considered as a limitation of the claimed significance of such effect.

Nevertheless, there is still a wide range of interrelated factors significantly contribute to learners' attitudes and practices in learning, such as learner identity, motivation, anxiety and self-confidence, which need to be taken into consideration while investigating washback. For instance, Murray et al. (2012) associated the attitudes towards a test to the candidates' experiential and demographic backgrounds. They found that personal experiences, such as failure in the test and the impact of other people yielded negative attitudes towards the test: particularly teachers and other test-takers.

This research project has taken into consideration exploration of the impact of the SECE on the students' attitudes and perceptions towards their language education in general and towards their learning motivation and practices. It will elicit the effective factors relating to this area of washback to fill the gap in the literature.

3.3.2.2 Washback on students' motivation

Due to the significant impact of motivation on language learners' achievements, it is crucial to investigate how assessment affects learners' motivation, especially in a test-oriented education system (Zhao, 2016).

Tsagari (2007) argued that the relationship between motivation, language learning and high-stakes exams is highly complex. Research has identified varieties of washback effects on language learners' motivation with different intensities. For example, Madaus & Clarke (2001), examining the use of high-stakes testing in the American education system on students, found that high-stakes standardised tests do not motivate students in classroom learning in a positive way. They found that the examination could not promote the learning interest for the unmotivated or less-motivated students. Their data was from research conducted at Boston College over 30 years. Along the same lines, Pan & Newfields (2012) examined the washback of the College Student English Proficiency Test (CSEPT) and the General English Proficiency Test (GEPT) as graduation exams on students' learning motivation in and outside of the classroom in Taiwan. They found that GEPT and CSEPT had not affected the students' motivation significantly. However, students who had English graduation exams at schools were more motivated to study English than those who did not have graduation exams at school. The positive relationship between the students' attitudes towards the test and their motivation was also observed in some

other research. For instance, Li (1990) also detected positive attitudes towards the Matriculation English Test and its content, which increased their motivation to learn English in the context through teachers' and students' questionnaires in China. Using interviews, Watanabe (2001) found that Japanese university students' test preparation for university entrance exams were linked to their motivation in which the test had a positive effect, encouraging students to undertake more test preparation activities if it was of appropriate difficulty. Using students' questionnaires, Read & Hayes (2003) found that learners have positive feelings about the IELTS exam and were motivated to study. However, such data depend on one source and which is difficult to provide an in-depth picture of the phenomenon. It cannot provide how that positive perception and motivation affected the practices.

A negative impact on motivation was noted in more research studies. Harlen & Deakin-Crick (2003) and Pan (2014) noted that language learning motivation could be discouraged unwittingly by assessment practices or as a result of cramming in their reviews of research on the impact of testing on students' motivation. This conviction is mostly directed towards high-stakes summative tests (Huang, 2012). For instance, several research studies in different contexts identified the use of high-stakes scores within the education system as a significant negative influence on students' motivation. High-stakes examination grades tend to diminish students' interest in learning to develop their language knowledge and ability (e.g. Al-Yaqoobi et al., 2010: Bahrain; Cheng & Curtis, 2004: many Asian countries; Guilloteaux & Dörnyei, 2008: South Korea; Hadri, 2015: Tunisia; Pan, 2014: Taiwan; Stan, 2012: Romania; Tayeb et al., 2014: Yamen; Zhao, 2016: China). In this regard, Kohan (1993) wrote:

> *The carrot-and-stick approach is generally unsuccessful; grades, in particular, undermine intrinsic motivation and learning, which only serves to increase our reliance on them. The significance of these effects is underscored by the fact that, in practice, grades are routinely used not merely to evaluate but note also to motivate. In fact, they are powerful de-motivators regardless of the reason given for their use*, (p. 201)*.

Despite the fact that Kohn's statement can be supported by several research studies as cited above, Hartaty (2017) introduced different results in an investigation of the effect of the Indonesian 2015 public assessment policy. Based on the fact that the

national examination was considered as a mean of graduation under the new policy, Hartaty found that this policy had a negative effect on the students' motivation to develop their language knowledge and ability. Besides, test anxiety and stress to achieve higher test scores were also considered as a type of negative washback of high-stakes testing (e.g. Cheng et al., 2014; Zhao et al., 2015). In a cross-public study across Canada, China and Taiwan, Cheng et al. (2014) analysed data elicited from test-takers, found that there is a complex relationship between learning motivation and test anxiety as they are closely related to social variables (i.e. perception of test purposes and importance) and personal variables (i.e. age and gender). Despite these results, however, one should be cautious when arguing that testing may be harmful for learning motivation and lifelong learning, especially when most of these research studies do not obtain in-depth data regarding the students' perceptions and motivation. These studies mostly relied on questionnaires to collect their data and neglected ethnographic data collection such as in-depth interview and observation.

The test effect on the students' perceptions and motivation was found to vary among students in the same level or between different levels. For example, Scott (2007) found that the intensity of a test influence could increase depending on the required grades by students in a research about stakeholder perceptions of test washback on learners in primary schools in the UK. Similarly, Ferman (2004), examined the washback of an oral examination on language education, found that students learn for the test if they are with low ability, believing that cramming can help them achieve a better score. The nature and intensity of washback on motivation were also found to be affected by different socio-contextual factors related to testing in the educational and social context. For instance, the extensive social and educational value of the examination grade could also affect students' learning motivation and practices in different Asian and MENA countries. Zhao (2016) concluded that learning' motivation of Chinese learner of English language was affected by the intergenerational motivation of their parents and teachers and the adverse adults' attitudes towards their test results. Zhao argues that the students' development of multiple language competencies was hindered by the excessive focus on increasing test scores within the test-driven education system contributes to the complexity and multi-dimensionality of language learning motivation. Another factor identified by Dörnyei & Ushioda (2013) was that the general perceptions and attitudes towards language learning are tightly linked to

motivation. It has been acknowledged that learning motivation is a complex concept, which constitutes various psychosocial factors related to learners' personalities and learners' social environments, test factors and students' extrinsic and intrinsic factors (e.g. Ainley, 2006; Harlen & Deakin-Crick, 2003). For instance, Gosa (2004) noted Romanian students' expectations, attitudes, perceptions and learning styles intervened with the washback of the English component of the Romanian school-leaving examination on the students' motivation and other areas of learning behaviour and practices.

As a result, the field of washback would benefit from more research investigating the effect of high-stakes examination on students learning motivation and before and after the examination (Shih, 2007; Tsagari, 2007). It is also important to shed light on how motivational behaviour is shaped over time and how it is influenced by the wider socio-cultural and contextual environment impinging on language learners, which this research study intends to consider.

3.3.2.3 Washback on learning strategy and practice

In spite of the extensive investigation of washback on teaching and teachers, empirical research studies into washback on learning and learners has just begun to gain more attention in the field. Research has noted that language learners' learning strategies could change according to their assessment practice, including test design and test use (e.g. Damankesh & Babaii, 2015; Mogapi, 2016; Phakiti, 2003).

Following washback on language classroom teaching, the high-stakes assessment was noted to affect students to adopt a measurement-driven learning approach. For instance, in different Asian and MENA countries where success in tests relates to life-long decisions about the test-takers' future, cramming tends to be adopted by most students in order to perform well in tests, rather than engaging in a real learning process. Sukyadi & Mardiani (2011), investigating the washback of the Indonesian English National Examination (ENE), using empirical qualitative multiple case studies, found that the test affects the students' learning in the classroom. The students' classroom practices were mainly mastering the classroom material for the test and developing test-taking strategies. Similarly, Damankesh & Babaii (2015), investigating the effect of the Iranian high school final examinations on students' learning strategies, found that high school final examinations direct student towards a measurement-

driven learning approach of learning through the extensive focus on test preparation material and activities. They identified different test preparation activities which were used by the high school students such as studying new words in the vocabulary part of reading comprehension texts, revising grammar and exercises, utilising memorisation, studying test-related textbooks, reviewing and practising previous test papers while neglecting parts not deemed for the examination. The researchers observed that high school students' learning practices were linked to the objectives of the course and the assessment, focusing on the tested language aspects rather than developing communicative language ability.

Zhang (2016), as well as Weili (2010), investigating the effect of CET-4 on Chinese college students, found that the examination encouraged the students to take an instrumental attitude towards their learning. They considered getting high academic scores as the only goal for their efforts in learning. Therefore, they usually sought to master the materials only in order to maximise their grades. The students adopted traditional test-preparation learning behaviours, such as practising test-related questions in and outside of school, to prepare for the test. Another example of how examination could narrow the students' focus on developing language skills was explored by Chou (2015), who investigated the effect of the introduced Test of English Listening Comprehension (TELC) in Taiwan. Chou found that students valued English skills, such as reading, grammar and vocabulary more than Listening because traditionally they have been emphasised in the College Entrance Examination (CEE). However, they started to focus on developing listening skills after the introduction of the TELC. In the same way, Cheng et al. (2011), studying the washback of Hong Kong school-based assessment, found that students' focus in their learning on test-related activities and the intensity of this learning practices increased closer to the exam time. Similar results were also reported in washback research of the CET on students learning in the tertiary education in China (Xie, 2011; Xie & Andrews, 2013). Gosa (2004) also indicated that Romanian high school students' intensively practised public examination tasks habitually utilised cramming as a learning behaviour.

Moreover, standardised tests were found to be responsible for the restriction of student learning as they focus only on cognitive dimensions of learning-tested discrete points of knowledge, and they ignore learning real-life knowledge and other qualities that are essential for successful language acquisition and use (Bailey, 1996; Chan, 2008; Pan,

2009). In the same vein, Green (2007) stated that test design has a strong relationship with learning behaviour in which it encourages language learners to focus on developing test-taking strategies rather than engaging in lifelong learning practice. Pour-Mohammadi & Zainol Abidin (2011) also highlighted that Iranian undergraduates held positive attitudes towards developing test-taking strategies in order to answer multiple-choice English reading comprehension tests. These test-taking strategies were identified as using world knowledge, translating directly into L1, employing reasoning, eliminating bad alternatives, utilising lexical cohesion, using collocation, considering global semantic context, guessing intelligently, heeding orthographical clues and alternatives to fill in the blanks. Aftab et al. (2014) also indicated that Pakistani high school students depend heavily on rote memorisation from the test-guide textbooks since they consider that the select-response public examination itself promotes these practices to pass the exam. There are other factors influencing students' emphasis on the test practice. An example observed by Dörnyei (2005) was that students with an insufficient level of interest in learning English as a second or foreign language usually tended to adapt learning to the test behaviour. In the same way, Bailey & Masuhara (2012), stated that the intensity of washback is linked to the importance placed on both examinations and coursebooks by the teachers' and learners' attitudes and needs, and desires to teach and learn this material. Stoneman (2006) argued that test-taking experiences and the students' past learning significantly influence their choice of learning strategies and practices. For example, Cheng (2005) attributed the students' use of the traditional methods of memorising grammatical, vocabulary and phrases to the teachers' use of teacher-centred instruction and text explanations in the Taiwanese language education context.

It can be noticed that different research studies have highlighted a direct washback impact on students from different perspectives (e.g. attitude, perceptions, learning strategy, motivation, learning focus and content, and test-taking strategy). Nevertheless, researchers highlighted that it is necessary to investigate washback on students, their learning and on the broader socio-cultural context where the test score is interpreted and used (e.g. Cheng et al., 2015; Shohamy, 2001; Weir, 2005). These socio-cultural factors could have an indirect impact on students' language learning. Therefore, this research aims to add to the literature by investigating how the test

design, the socio-cultural and contextual factors may influence language learning practices in the research context.

3.3.2.4 Washback on learning outcomes

It has been noted that much washback research tends to focus on investigating the test effects on educational processes and neglects the test effects on the quality of the students' learning outcomes (Wall, 2000; Wall & Horak, 2006). Some researchers have attempted to investigate the relationship between teaching practices and score improvement as an indicator of language learning outcomes. For instance, Andrews et al. (2002) investigated washback on learning outcomes through analysing test scores of the high-stakes examination of Use of English (UE) from 1993 to 1995 in Hong Kong. They compared the test scores of three groups of students in the final year of schooling (Secondary 7) over three years. The first and third groups were taught through a measurement-driven approach in their first and second years. It was found that no statistically or significant improvement in the students' scores in the first and third groups. Similar to several studies in the field (e.g. Aftab et al., 2014; Andrews et al., 2002) argued that such type of washback seemed to represent a superficial learning outcome. They attributed the students' improved scores to their acquaintance with the test format, formulaic phrases and the amount of time spent on language learning and the rote-learning of exam-specific content and strategies. In contexts where a prescribed curriculum is used, the possibility of such practices would be highly increased. It has also been shown in the case of the students in schools with English proficiency test of Exam for Certificate Exit Requirements (ECER) that the time spent practising the listening skills of the examination had a positive effect on increasing their score in listening (Furaidah et al., 2015; Pan, 2016; Pan & Newfield, 2012).

Further evidence was provided by Elder & O'Loughlin (2003), examining the connection between band score gains on the IELTS and intensive English language study in Australia and New Zealand. They found that during the three-month period students made progression in English with an average of about half a band overall. Elder and O'Loughlin found that improving scores on tests were linked to factors such as personality, motivation, confidence and exposure to examination content and skills. These findings indicated that teaching to the test does not automatically improve

students' scores, however, other factors such as students' original proficiency and motivation does have a significant effect on enhancement of their test scores.

Nevertheless, some researchers have stated that it would be simplistic and naïve and to claim that score gain mainly reflects a development in language learning or to teaching practices (e.g. Damankesh & Babaii, 2015; Pan, 2016). The scores' improvements may simply reflect the development of effective test-taking strategies. For instance, Green (2007) investigated the effect of test preparation classes on learners writing test scores development in the UK, using supplementary questionnaire and test instruments. He found that test preparation classes did not help students improve their IELTS scores. Other researchers concluded that the increase of the students' scores on the standardised examination could be a result of using test-taking strategies rather than to the development of the students' language learning outcomes (e.g. Aftab et al., 2014). For instance, Pour-Mohammadi & Zainol Abidin (2011) revealed via a questionnaire that the use of test-taking strategies by Iranian EFL undergraduates was of a significant effect on their performance on a multiple-choice reading comprehension test. In the same context, Damankesh & Babaii (2015) found that guessing had a significant effect on the students' scores in the multiple-choice examination. Such an impact has been widely observed in the standardised examination in the global context (Hsieh, 2017: Taiwan; Toksöz & Ertunç, 2017: Turkey; Yu & Jie, 2013: China). These findings are important for the current research since the SECE is a select- response examination.

Teaching to the test and test-taking strategies might increase students' scores, but the score gains are not always statistically significant. Moreover, classroom instruction of exam-specific strategy factors, such as exposure to the test content, could contribute to a score gain. Due to the drawback of research in this area, the extent of the effect of testing on learning outcomes should be further investigated since measurement-driven instruction and practice, or factors such as their motivation, original proficiency, personality, and the exposure to the target language all seen as elements of language learning outcomes (Pan, 2016). More empirical evidence is required to link the classroom education practices and test design to test scores and language learning outcomes (Green, 2013). Thus, it is important to indicate from the learners' points of view whether the improvement of scores reflects the fact that their knowledge was developed or reflect the rote-learning ability or developed test-taking strategies.

Hence, the current research tries to fill this gap in the research field by presenting such evidence through seeking to understand the students' perceptions of the reflection of their SECE and their language learning outcomes and the relationship between their scores and learning outcomes.

3.4 Factors influencing washback

Wall (2012) argued that it is difficult to separate the test impact from the other influences caused by other variables affecting language teaching and learning in the educational contexts. These factors could be connected in a way to determine the nature and the degree of washback on language teaching and learning in contexts. These factors could be test-related factors (e.g. the stake, content, design and the purpose of it) or teacher-related factors (e.g. beliefs, attitude, knowledge, training and experience). This section briefly discusses teachers and test-related factors, which could affect washback in context.

3.4.1 Test-related factors

Alderson & Wall (1993) stress that the type of washback may not always depend on the quality of the test since any type of tests, good or bad, could cause constructive or destructive washback. However, it is essential to design valid tests, which produce accurate scores and to create positive washback. Green (2007, 2013) revealed that a number of test design features may affect classroom language learning and teaching behaviour and can guide washback in a positive or negative direction. These include test description categories such as test item format (e.g. extended response, selected-respond items, short answer questions), purpose (learner performance, teacher or curriculum evaluation). difficulty (easy or challenging), content (topic and skills), level of knowledge required (understanding, retention or use), complexity (the number of content areas and interrelationship), and type of items (or proficiency). The following section discusses how these test design features promote washback.

3.4.1.1 Test purpose

It has been argued that when test results become the major determinant of future education or career opportunity, societies tend to regard success in the test, rather than the achievement of knowledge and developing academic skills, as the principal goal of school education (Green, 2007; Hyland, 2011). As a result of the insufficient

emphasis on the application of knowledge in the global educational systems, the role of testing could alert learners' perceptions and motivation towards language learning and teaching. This makes individual participants to value success in a test more than construct knowledge and understanding. Many participants were reported as being encouraged to focus on rote learning of the test content over deep learning with the objective of maximising understanding and interest in the subject matter through the purpose and the nature of testing (see 3.3.2 and 3.3.3 for examples). Ajideh & Mahmoudi (2017) found that the Iranian National University Entrance Exam (INUEE) due to its norm-referenced nature contributes to the negative washback effect of the test on language teaching and learning practices. Green (2007) also concluded, from his study of the effect of test preparation on IELTS score improvement, that the potential drawback of IELTS would be the need to succeed in the test by whatever means rather than to develop the range of skills required in the target domain.

3.4.1.2 Test format

Large-scale standardised tests were reported to have negative influences on the quality of language teaching and learning. It has been described as limiting students' language learning of high-level cognitive skills as it focuses only on cognitive dimensions and ignores different essential aspects of student success in developing language use and competence. For instance, Rahimi et al. (2016) reported that standardised tests restrict test content and atomise knowledge and learning. Fulcher (2000) and Patrick et al. (2016) also believed that standardised tests only assess knowledge-based memorisation and factual recall, rather than high-level or complex cognitive skills. Hughes (2003) stated that the multiple-choice questing leads to discrete points teaching and focus on developing test-taking strategies; and that does not enhance effective learning strategies developing their language knowledge and ability. Such characteristics could not only lead to negative washback on language teaching and learning but it may also be considered as deficient in terms of theory-based validity of a test according to Weir (2005). An example of such washback was observed within the format of the NMET, which adopted multiple-choice questions in China. The NMET did not mediate the test designers' intentions of promoting communicative teaching and learning in order for students to be able to use English in social and educational contexts (Li et al, 2012; Qi, 2004). Although the test format is conducive to high reliability, the disconnection between test tasks and the intended

classroom activities produced undesirable washback on language teaching and learning (Cheng & Qi, 2006).

Moreover, Bailey (1999) and Taylor (2005) claim that the test alignment with either traditional teaching approach or the CLT can direct the test's washback to be negative or positive. In support of this claim, Berns (2017) recommended that researchers should attempt to understand the extent that language defined washback in relation to the presence or absence of CLT through examining several prominent washback studies. However, Bachman & Palmer (2010) argued that the skills required to perform and succeed in a test can never fully represent those needed for completing real tasks in a target language-use context in spite of the best efforts of test developers. To some extent, that could be true but it can still be argued that the construct-response item format, such as writing in the IELTS, could yield more positive washback than selected-response items in terms of developing students' language knowledge and ability.

3.4.1.3 Test content

The effects of high-stakes testing depend on the quality of the test items to a large extent as it may affect the implementation of the curriculum positively: "curricular content expansion, the integration of knowledge, and more student-centred, cooperative pedagogies" (Au, 2007: 258). Thus, it is important that the examination and curriculum content match, in order to achieve test content validity (Hughes, 1993; Weir, 2005). The alignment will assist language learners in obtaining and developing the knowledge and skills introduced in the curriculum to be able to perform well in the exam and be able to use the language.

It is generally noted that the gap between the test content and the objectives of the course always has the consequence of undesirable washback. The available evidence from research studies (e.g. Alderson & Hamp-Lyons, 1996; Qi, 2005; Wall, 2005; Watanabe, 1996) shows that teachers tend to limit the content of instruction to material covered in the test and use tasks in the classroom that reflects test tasks. It also appears from studies that teachers are selective in their focus on different aspects of a test which have more weighting than on others (Green, 2007; Saville & Hawkey, 2004; Tsagari, 2009; Watanabe, 2004).

3.4.1.4 Test administration

Primary considerations affecting washback on the teaching and learning towards the preparation of the test are the circumstances under which the test takes place. If the test is not well administered, Weir, (2005) stated that unreliable results may occur. Cheating in public examinations is an international issue in general, and in the MENA countries particularly. It has been observed by many research studies in the field. Cheating behaviour has been found to be an educational and social behaviour in many countries. For instance, parents are concerned to assist their children to pass public examinations in order to be able to get a job and help them in India. In this scenario, if cheating is seen as acceptable, it could affect students' motivation in learning since they know that they will get help to pass their examinations. Detailed process and arrangement should be prepared to ensure the efficiency and the uniformity of a test administration for the candidates anywhere and at any time. This requires providing the candidates with a precise and clear set of instructions about test conditions, especially rooms and test materials and equipment should be carefully checked to overcomes any problems before the test administration. Procedures for dealing with candidates' cheating should be determined in advance with the observer. This involves controlling access to the precise test content by candidates, test evaluation and test scoring. In other words, candidates should not be able to access to unauthorised copying of the test before the test sitting. If some candidates get access to the test content then they would know the answers in advance and they will proceed that through different skills required by the test, i.e. solely reliant on memory.

3.4.2 Teacher-related factors

Teacher-related factors have been found as affecting the level and the nature of washback on teaching practices. These factors include teachers' beliefs and attitudes towards their professional stance, teachers' education and knowledge and teachers' experiences. An example of the effect of these factors was highlighted by Watanabe (1996), investigating the effect of the university entrance examination on the prevalent use of the grammar-translation method in Japan. Using interviews, classroom observations and analysing the previous English examinations. Watanabe found that there was no significant connection between the test content and the teachers use of GTM. Rather, teacher-related factors such as past education and academic

background, and personal beliefs had an important role in their choice of the teaching methodology they employed in their classrooms.

3.4.2.1 Teachers' beliefs, attitudes and perceptions

Several research studies confirmed that teachers' perceptions of a test determine their motivation and their interpretation of the educational policy and practice. The teachers' perceptions affect their alignment of the curriculum objectives and enhancing improvements in their pedagogical practice and develop students' strategic cognitive skills. For instance, Pan (2014) found that teachers' beliefs in the importance and the value of the high-stakes public test for students' futures encouraged them to assign students for test-preparation practice out of school classrooms. Similar washback was observed where the English National Examination (ENE) in Indonesian secondary education affected the 12th-level English teachers' classrooms, but not for the 10th and the 11th levels. This indicates that the washback effect of the ENE only takes place when the teachers feel that the ENE preparation is crucial. Similarly, Alderson & Hamp-Lyons (1996) and Qi (2005) argued that the perceived function of the test and the interpretation of the test construct compel teachers to work for improving scores through the focus on the language linguistic knowledge instead of focusing on language use. Hence, Gunn et al. (2016) mentioned that it is vital to know what impression teachers and other practitioners have of these tests and their impact on themselves, their language teaching and their students.

Such practices can also be encouraged by factors other than the examination. In their research projects, Hamp-Lyons (1998) and Green (2007) confirmed that associated poor teaching practices might be caused by a traditional English language teaching profession rather than the test design and content. For a researcher to build such an argument, he would need to ensure that such practice is not influenced by other teacher factors.

3.4.2.2 Teachers' knowledge, education and training

The importance of the meaning of assessment literacy in terms of the message systems of curriculum, pedagogy and assessment has been acknowledged in different contexts. Berry & O'Sullivan (2016) defined assessment literacy as the level of the required knowledge of assessment concepts to interact effectively with assessment in their situation. They identified six areas in which language teachers need to develop

their assessment knowledge in order to be able to administer effective assessment practice for learning:

- Skills in identifying and evaluating appropriate assessments for specific purposes within specific contexts.
- Skills in analysing empirical data in order to improve one's own instructional and assessment practices.
- The knowledge required to assess learners effectively and maximise learning.
- The knowledge and understanding to interpret and apply assessment results in appropriate ways.
- An understanding of the principles and practice of sound assessment.
- The wisdom to be able to integrate assessment and its outcomes into the overall pedagogic process (Berry & O'Sullivan, 2016: 18).

Despite the vital role of assessment literacy in the quality of teaching, there is evidence that many teachers have an inadequate level of assessment literacy (Bachman, 2013; Baird, 2010). Several research studies have identified different contextual factors that influence teacher assessment literacy. For instance, Gebril (2017) found that the Egyptian teachers held the perception that their teacher training programme was not effective in providing sufficient focus on language assessment literacy. Gunn et al. (2016) mentioned that teachers are often instructed to administer a certain type of test and have little or no chance to give feedback regarding whether or not, they feel that these tests are useful, valid or should even be given. Such practices could constrain the creativity and spontaneity of teachers and demean teachers' professional judgements. Sheehan & Munro (2017) in their research about the experiences and attitudes of language teachers regarding language assessment literacy practices in classrooms across Europe, suggested that assessment practices are rooted in past learning experiences, and through knowledge-sharing with colleagues.

Teachers' original assessment knowledge, however, was barely considered as a factor of significant influence in washback research. Researchers tend to consider the teachers' assessment practices as a reflection of their assessment literacy. What is missing is the understanding of how these factors interrelate with each other and with the original knowledge of teachers' assessments, especially since Brown & Bailey (2008) argued that we need to understand more about how language testing is taught

in language teachers' training programmes universally. The connections between the language teachers' knowledge and beliefs about assessment and their assessment practices in the language classroom have not been extensively researched. Particularly when these factors are put together, it is unclear how assessment literacy is developed and enacted in day-to-day classroom practice. This study will focus on how these public examination practices and policy affected the teachers' beliefs about assessment in the study context.

3.4.2.3 Teacher experience

Teacher experience plays a significant role in their beliefs, perspectives and decisions about their classroom teaching and assessment practices. Teachers' translations of curricular goals into test specifications can provide a negative link between the curriculum, teacher experience and testing. Huang (2012) argued that it is possible that teachers could continue to be affected by the roots of the traditional language teaching approach or method, such as GTM, since this approach has been successfully compatible with the assessment focus and practice rather than integrating the communicative approaches of teaching languages since language ability is not tested. In washback research, teachers' teaching experience was identified as a key reason explaining the level of washback on teachers (Cheng, 2005; Shohamy et al., 1996; Watanabe, 1996). For instance, Shohamy et al. (1996) showed that experienced teachers were affected more by public examinations than the inexperienced teachers. They showed more commitment to align their practices with the test requirement as they consider it as the key guidance for classroom instructions. Sheehan & Munro (2017) mentioned practices could lead teachers to develop their assessment practices through their own experiences. They claimed that it could have the potential to have a positive impact on classroom practice, as teachers can learn strategies, which have been successfully deployed.

Yang (2013) stated that more research is required in language education to describe explicitly how different teacher factors, such as their experience, education and pedagogical beliefs, about washback within their unique contexts, affect the implementation of the curriculum and examination. Thus, this research took in consideration how different teachers, test and contextual-related factors influence

teachers' decisions about teaching and students' learning practice to deal with the externally initiated test score policy within the Libyan educational context.

3.5   Chapter summary

Washback has a complex association with social and educational contexts. Empirical research on washback of high-stakes examinations revealed the existence of washback in which it can directly influence upon different aspects of classroom teaching and learning in various forms and intensities. It also could affect stakeholders and test-takers differently (Cheng, 2005). However, washback influence is not always straightforward since it can be indirectly intervened by different variables such as test-takers and teacher characteristics, test design and use, and educational policy. The complex interaction between these variables determines the existence and nature and intensity of washback in different contexts. To date, research studies do not show agreement on the nature of washback that each factor facilitates. Thus, washback can be considered as an unpredictable phenomenon that its nature and intensity influenced by different factors besides the test design and use, such as contexts and stakeholder characteristics. Hence, more research on washback is required in different educational contexts, especially once the exam's format and contents had become familiar to teachers (Spratt, 2005), such as in the Libyan context. It is possible that there is still different interaction between factors influencing washback in different teaching and learning contexts, which has not been described yet. For that, these issues are worthy of investigation in this research and in the study context.

This study would examine the way in which variables (e.g. motivation, beliefs, knowledge, experience and contexts) facilitate or hinder various forms and degrees of washback. It aimed to provide thick descriptions, contextualised from an ELT practitioner including students, in order to present an authentic picture of the washback through in the research context. In the case of teaching English as a foreign language, practitioners could encounter challenges and difficulties in accommodating the SECE testing systems with their cognitive beliefs, experience, practices and the educational policy.

4 Chapter Four: the conceptual framework

4.1 Introduction

Weir's socio-cognitive framework of test validation (2005) is the conceptual theoretical framework that informed the data generation and analysis of this study. The socio-cognitive framework was chosen due to the complex nature of washback and the different related factors affecting construction and the process of the educational system in the study context. It promotes interrogation of the participants' motivations, perceptions, perspectives and experiences during their preparation and performance in the SECE examination in the on-going conflict context of Libya. It is also a practical model in order to achieve and generate test validation evidence at different stages, i.e. the *a priori* and *a posteriori* stages, about a test efficiency and usefulness in context (Geranpayeh & Taylor, 2013).

Over the last decade, the extensive research studies on washback have used different theoretical models to explore the complex phenomenon of washback, highlighting the different relationship between language teaching, learning and assessment. An example of the theoretical models used is the washback hypothesis by Alderson & Wall (1993), the basic model of washback by Bailey (1996) and the Washback Model of Students' Learning by Shih (2007, 2009) discussed in the literature review (section 3.2.4). These models have not been chosen for the current research due to their limitation to fulfil the aim of the research study. Whereas the washback hypothesis and Washback Model of Students' Learning are limited to the effect of assessment in the micro context, the basic model of washback has not presented a detailed explanation of the process in which washback operates.

Other theoretical models that have growing influence on language policy and education within Europe and worldwide, are Bachman's (2005) and the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001). Although Bachman's model of communicative language ability is considered to be "the most promising" of the language models, it has been criticised for having a psychological basis and undermined social and cognitive aspects. This limitation critically reduces its suitability for use in language tests (Weir & O'Sullivan, 2011). The CEFR, on the other hand, included a set of common reference levels for language assessment as well as teaching and learning from basic levels to advanced. The

CEFR seeks to determine the key features of advanced language proficiency, however, it is a practical model where a number of theorists, such as Fulcher (2004), highlighted that the model was not supported by language theories. Clearly, it has no direct relevance for the public examination of the SECE as it was designed for testing academic language skills, which is not included in the test under investigation.

The socio-cognitive framework is the only viable model that considered the understanding of psychological and social perspectives as a central aspect of language test development. It offers an explanation of how these perceptions are connected to validation and it can help in developing, validating and critiquing tests. It is also had the capability to enable researchers to provide adequate evidence on how testing design and content is operationalised and used or interpreted in practice. Therefore, an evaluation of the effect of the SECE on language teaching and learning: washback is important in order to build an argument about its efficiency and usefulness (validity) in the Libyan education system. As the practicality of Weir's validation framework has been demonstrated by a number of studies (e.g. Khalifa & Weir, 2009; Weir & Shaw, 2006) the current study will test the practicality of the socio-cognitive framework in investigating the validity and washback of the SECE in the Libyan context.

4.2 The socio-cognitive framework of test validation (Weir, 2005)

The socio-cognitive framework of test validation (Figure 3) was first developed by Weir (2005) and subsequently further developed by Shaw & Weir (2007), Khalifa & Weir (2009), O'Sullivan & Weir (2011). The framework is a precise, practical and systematic model that provides stakeholders of language testing, such as test developers and test score users with a comprehensible methodology for designing or developing a test or to validate one. The model is also a coherent and accessible methodology to describe and evaluate a range of English language tests used for different purposes. According to O'Sullivan & Weir (2011), that is important so that tests can be localised, so that the test specifically suits the candidates and the knowledge and skills being assessed. Hence, the framework is highly adaptable to suit any context. It can be used to examine or evaluate different facets of English language competence and to assess subject-specific material. The developed version of the socio-cognitive framework by O'Sullivan (2011) has now been widely implemented to develop tests, teaching and

assessment, especially by the British Council. O'Sullivan (2012) encourages the socio-cognitive framework's application in innovation about developing a new testing system. An example of such development was the testing system of APTIS, which is emerged by the support of the socio-cognitive framework.

Figure 3: The socio-cognitive framework of test validation
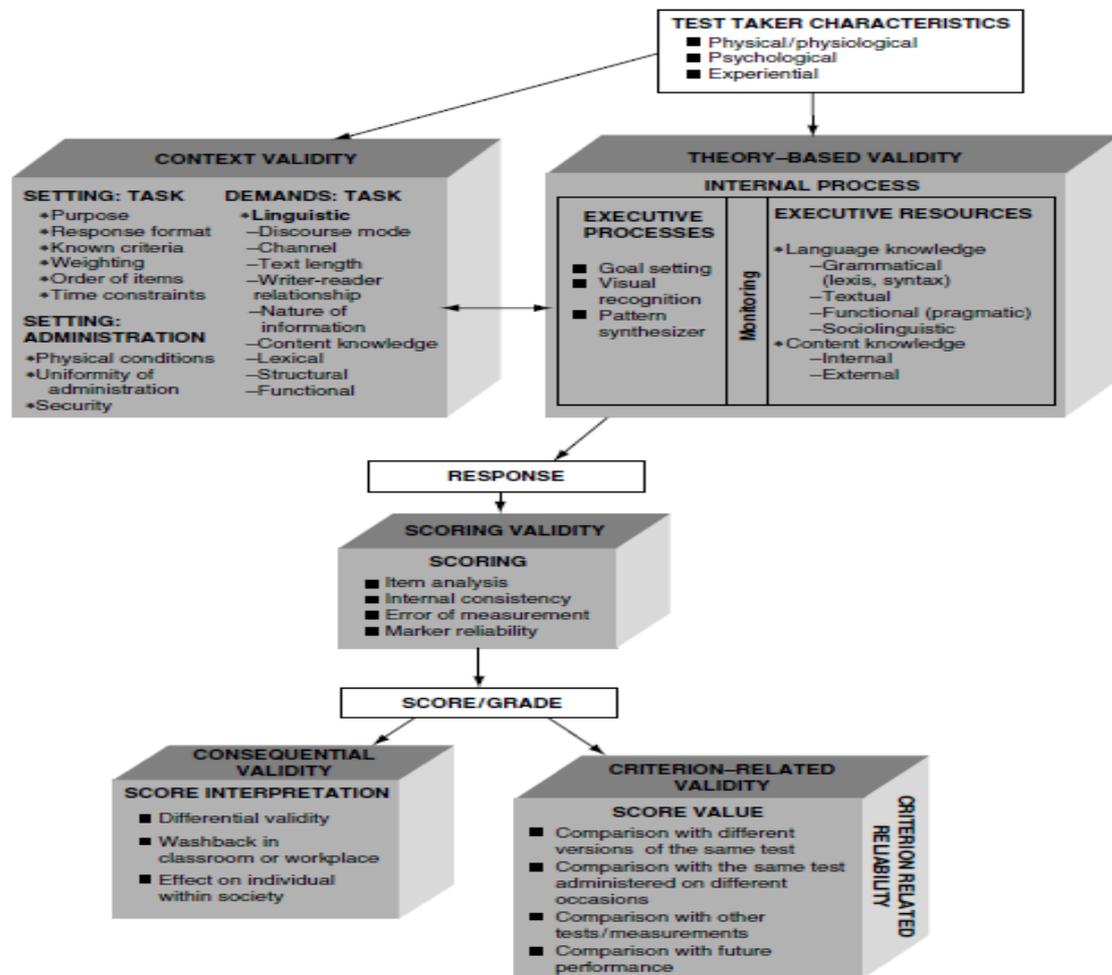
The source: Weir (2005)

The origins of the framework can be attributed to the early academic works of Weir, such as "*Communicative Language Testing*" (1990) and "*Understanding and Developing Language Tests*" (1993). However, the framework was based on his work on developing the public College English Test, the Test for English in China, and the Cambridge English examinations in the UK. The framework was developed to specifically highlight the conditions and operations underlying language performance. That was achieved through presenting a systematic and coherent approach to these

activities, combining the cognitive, social, and evaluation (grading) dimensions of language use with the context and consequences of test use.

The approach identifies the evidence required to develop a transparent and coherent validity argument, while at the same time addressing the interaction between different types of validity evidence. It should be emphasised that the socio-cognitive framework considers the abilities to engage with language users' mental processing (the cognitive dimension) and their abilities to engage with language use in performing tasks as a social practice but not as a pure linguistic practice. The dimensions of individual characteristics, internal cognitive processing and external contextual factors establish the model components for understanding and describing second language ability and skills. The outlines three key elements: the candidates' linguistic and cognitive resources; the test system including the administration of the test's tasks and its sensitivity to the social and performance variables of the context; and the scoring system, which is theoretically linked to the former mentioned components of the framework (O'Sullivan, 2008).

Beside test-taker characteristics, the framework includes context, theory-based (or cognitive), scoring, consequential and criterion-related validity to provide a comprehensive evidence for validity. The socio-cognitive framework was detailed differently to suit the four language skills (listening, speaking, reading and writing). As the SECE test content emphases reading and grammar, the socio-cognitive framework for validating reading tests was adopted to this study. The framework reflects that different components are complementary rather than alternatives, where none of the described validities has priority over others to represent the basis of a test. The study applied Wire's (2005) socio-cognitive framework to examine various aspects of validity of the SECE in term of contextual parameters, cognitive processing and skills, test results and the effective factors. To establish situational and interactional authenticity, this study relied on the perspectives and practices of stakeholders; including students, ELT teachers and inspectors, and other practitioners (more details in section 5.4); to empirically examine the degree of the correspondence between the characteristics of the English language test of the SECE and give explanations of test performance to gather evidence of interactional authenticity. Therefore, the data of this study was generated vertically according to the socio-cognitive framework. Even the design of the primary data generation tool was

designed to generated data in the same way. The following sub-sections present the socio-cognitive framework for validating tests in detail and its influence on the process of data generation and analysis.

4.2.1 Test-taker characteristics

The socio-cognitive framework emphasises the importance of considering the test-taker as the cornerstone of the process of cognitive validity. The test-taker characteristics are directly connected to cognitive validity as they have a direct impact on processing the test task by candidates in the set-up of the context validity. The test-taker characteristics and cognitive validity concerned with the candidate in the test event. They represent the individual cognitive or mental processing abilities of the language user since their characteristics influence the way they process the test task. The test-taker characteristics include: physical/physiological (such as problems with speaking, i.e. dyslexia, hearing and vision), psychological (cognitive style, personality, concentration, motivation, memory, and emotional state), and experiential characteristics (education, examination experience and preparedness, communication experience, target language country residence). O'Sullivan (2000) introduced a useful synthesis of literature on characteristics of test-takers, which have a potential effect on test performance. For instance, the candidate's motivation could influence is the strategy of managing the task. Factors such as the type of preferred learning styles and personality, experience in preparing for examinations, educational and cultural background, and familiarity a test design and content could influence performance in the test. In this study, the test-takers of the SECE were interviewed prior to the examination to enquiring about their perceptions, attitudes towards the test design and its use in the Libyan context, and language education. It also enquires about their experience with public examinations. The observation was also utilised to examine if these characteristics have a relationship with their preparation for the SECE in their classroom practices which could influence the cognitive and context validities.

4.2.2 Context validity

Context validity examines the task's social and external contextual parameters, which were represented in terms of the task demands, setting, (the input) and its expected output. In other words, it refers to the appropriate domain of knowledge or skills important to performance in the test. The context validity for reading tests is

categorised into task setting, task demands and setting of administration. Task setting contains test purpose, known criteria, response format, weighting, the time constraints and order of items. Task demands concerns with discourse mode, the channel of communication or presentation, writer-reader relationship, text length, nature of information, the content knowledge required and other linguistic variables including structural, lexical, and functional knowledge. Test administration comprises physical conditions of the uniformity and security of the test itself and its administration. It was recommended to investigate the participants' interpretation of test demands and the way participants respond to the skills needed for promoting beneficial washback since the perceptions of different participants could inform the development of assessment procedures and enhance its effects on the education system.

Tests should ensure the authenticity of interactional (cognitive) and situational (contextual) aspects (Douglas, 2000; O'Sullivan, 2006). Thus, the context must be acceptable to testers and test-candidates as an applicable situation to assess a domain language ability. Thus, Weir (2005) highlighted that the degree of candidates' familiarity with the content knowledge required for achieving a certain task affects their performance on that task. Therefore, he emphasised that candidates should be sufficiently familiar with the linguistic genre of the content of a test in order to obtain a sufficient existing schema enable them to use different skills and strategies to understand the text. However, such matching with examination could threaten the content validity in a context where schools use prescribed textbooks for the curriculum. The use of the prescribed textbook to design the national examination item could lead to the repetition of the exam question. Saukah (2015) observed that such practices have created mass production and, more seriously, a proliferation of the answer books to do the public exam, increasing the opportunity to cheat during the test administration. It was also noted by Froetscher (2013) that cheating can be effective if candidates obtained exact or close copies of the test questions before the test. Such context would have a significant influence on the test score representation and validity.

Another example is that the use of authentic texts and materials for second language learners is usually frustrating for both students and teachers due to the various type of competence needed to process the test input, such as pragmatic competence. Although ensuring complete authenticity may not reasonable in a language test, however, selecting realistic settings for teaching and testing in terms of various critical

contextual features can be possible (Khalifa & Weir, 2009; Weir, 2005). For that, teachers and testers should modify the test content to enhance students' interaction with the included knowledge in the text (Baleghizadeh, 2011). Since the socio-cognitive model considers the ability in language as a part of the language users in context, this emphasis on context makes the framework as highly adaptable for different aspects of English language competence and for assessing different subject-specific material. In this research, interviews, observation and document analysis were utilised to examine the SECE demands and design in relation to the curriculum to seek their relation to the students' cognitive process and skills as well as teaching and learning strategies and practices in the English language classroom. This also includes the other stakeholders' perceptions and practices (including teachers, inspectors, head teachers and examiners) and its effect on the students' perceptions and practices. In addition, the study examined through the interviews how the test administration in the on-going conflict of Libya influenced the stakeholders' perception and practices.

4.2.3 Cognitive validity

Cognitive validity is the updated term used in relation to theory-based validity by Shaw & Weir (2007), and Khalifa & Weir (2009). Cognitive validity is associated with the internal mental processes concerned with linguistic knowledge acquisition. It involves the executive processing and resources of the test. Executive processing consists of monitoring, goal setting, visual recognition and pattern synthesiser. Executive resources include the knowledge of language and the test content. The language knowledge takes account of grammatical knowledge (lexical and syntax), textual knowledge, functional (pragmatic) knowledge and socio-linguistic knowledge. The content knowledge comprises the task-specific knowledge (external) and the candidate background knowledge (internal). Douglas (2000) raised the importance of considering the relationship between the content of the examination texts and the general and cultural knowledge of candidates (background knowledge) and specific subject knowledge such as text issue and content) when designing a test content.

Although context and cognitive validity were treated separately for descriptive purposes, they complete each other. They are interrelated in the way they interact with scoring validity and form the construct validity. The way that the test task knowledge influences the candidate's internal mental process (cognitive) to varying degrees as

that knowledge interacts with their linguistic and content knowledge of internal and external resources of the test.

In test validation research, the cognitive validity evidence of a test can be established before the test performance through verbal reports from test-takers since the candidates activate the cognitive processing needed for the test task prior to the test event. It can also be established through statistical analysis of scores following test administration. In this study, data were designed to be generated about the students' cognitive process and skills used to prepare for the SECE and while undertaking the test from students' perspectives through prior and subsequent interviews to the SECE as well as classroom observation. In the data analysis, it was also important to investigate the factors behind the students' chosen of the cognitive process and skills to prepare for the SECE and their language learning strategies in general.

4.2.4 Scoring validity

Scoring validity presents how a test item (performance) is scored (evaluated). It accounts for the degree to which the test score is measurement-error free. It concerns with the reliability of a test score to be used to make decisions about the test-takers. The lack of scoring validity of a test seriously threatens its validity as it can provide sufficient evidence about validity in relation to the other aspects of validity. However, scoring validity is not a necessary reflect the quality of test since it should be balanced to the other aspects of validity.

The framework consists of four elements representing scoring validity: internal consistency, item analysis, the error of measurement and marker reliability. Internal consistency is used in homogeneous tests where statistics can be used to offer data about the reliability of the test. Item analysis discusses the easiness or difficulty of the test items using statistics in order to offer more information about a test's candidates' abilities. The error of measurement illustrates differences between the test score and the tested language proficiency. Finally, marker reliability greatly influences the general reliability of a test. Through its symbiotic relationship with the cognitive and context validity, scoring validity constitutes the overall construct validity of any test. Weir (2005, p.20) refers to this relationship as follows:

"*There is a symbiotic relationship between context- and theory-based validity and both are influenced by, and in turn influence, the criteria used for marking which are dealt with as part of scoring validity*".

The test scoring process is usually affected by the test type of the test (objective or subjective), the number of raters, methods of scoring being manual or mechanical (Khalifa & Weir, 2009; Weir, 2005). This study designed to elicit the evidence for the scoring validity through the students' reflection on their SECE's score and performance in comparison with their English language knowledge and ability. It was also designed to elicit more evidence from the educational practitioners through their reflection on the secondary school students learning outcomes. These reflections meant to be compared to the objectives of the English language curriculum in the data analysis to identify the relationship between the identified ability and the target of English education.

4.2.5 Consequential validity

Consequential validity is where this study is located. It represents the consequences of assessment on participants involved in teaching and assessment (most notably teachers and learners) or on the whole society (Khalifa & Weir, 2009; Shaw & Weir, 2007). The test can be classified as having a positive washback on language teaching if it mediated beneficial learning and teaching strategies and practices (Hughes, 2003). However, such positive influence should be long-term effect when it should not be exerted over language teaching and learning only because candidates valued success on the test more than learning. Green (2013) stated that the short-term goal of success on the test could inhibit the development of tested language ability. This can be true, particularly when the focus of the teaching programme is orientated to test preparation. The elements of consequential validity concern with the test consequences in three areas: differential validity, washback in classrooms and its impact on individuals in society. This study is designed to examine the washback of the SECE on perceptions and attitudes through the data generated about the test-takers characteristics and the related factors influencing these characteristics. Evidence regarding washback on teaching and learning practices elicited from the analysis of the data in regard to the contextual and cognitive validities. Finally, the washback on learning outcomes had been placed in the scoring validity of the SECE.

4.2.6 Criterion-related validity

Criterion-related validity describes the connections between test scores and other assessment tools that assess or examine the same knowledge or ability (Khalifa & Weir, 2009; Weir, 2005). The criterion-related validity reflects the corresponding between reliability and validity. The socio-cognitive framework considered the comparison with different versions of the same test and administering the same test on different occasions as external measurements tool which can be used to scrutinise criterion-related validity. Since the English language school curriculum and the SECE has not been attached to any external measurement criteria, it would be difficult to make such comparison in this study. However, I examined the objectives of the final year students' textbook in relation to my understanding and study of the CEFR to identify the relationship between the identified language ability by the students in relation to the textbook objectives. It was not possible to include the SECE score in this comparison since it has not got any outlined for the desired level of knowledge and ability on the test.

4.3 Chapter summary

This study used Weir's socio-cognitive framework for test validation as the conceptual framework for data generation and analysis. The framework was built on the basis that test validity comprises the social condition where a test is used as well as the cognitive processing of the language knowledge required for the test. The framework focus on the test candidates as the central aspects of the theory for designing, developing and validating a test. It a practical tool for researchers and teachers to examine washback or to evaluate teaching and learning practices in a context (O'Sullivan, 2011). It comprises three elements of the assessment process: the test candidates' characteristics and their cognitive and linguistic resources, the test system (including variables of the tasks demands and performance in the wider context) and test scoring which influenced by the previously mentioned elements. Thus, the framework is highly sensitive to context. Thus, Weir's socio-cognitive framework, with its core components, offers researchers and test designers a coherent and systematic methodology that considers different variables concerning with the development and validation of tests and investigating test washback.

Moreover, the framework enables exploration of language tests' consequences, focusing on washback at macro and micro levels (cognitive, social dimensions and test design) and in the broader notion. Specifically, Weir's socio-cognitive framework was applied to critically evaluate the Libyan SECE, examining how the test operationalises the learners' language abilities and constructs their motivation. It was used to analyse systematically features that affect the SECE's validity and reliability, and identify how the test can be improved in perspective of its fairness and purposes in the Libyan context. Besides, the framework enables the researcher to generate a description of the SECE candidates' performance in relation to the curriculum objectives, teaching and learning strategies and outcomes.

The next chapter gives a detailed description of the research methodology and explores the data collection and analysis through the philosophical and methodological lens.

5 Chapter Five: research methodology

Introduction

This chapter presents the research methodology employed to answer the research questions of this study:

1) To what extent does the Libyan public examination of the SECE have an impact on motivation, perceptions and attitudes towards English language teaching within secondary school education?

2) To what extent does the public examination of the SECE influence English language teaching, assessment and learning strategies in secondary school classrooms in the on-going conflict context of Libya?

3) How does the Libyan public examination of the SECE influence the students' English learning outcomes within secondary school education?

The chapter starts with a discussion of the researcher's ontological and epistemological perspectives, which informed the interpretivist research paradigm and the qualitative research approach to examine the Libyan secondary school teachers' and students' attitudes, perceptions and motivations towards the SECE and its effect on their educational behaviour and practices. Next, it presents the case study methodology used in this research study. The chapter also covers the sampling technique and the participants, as well as the data generation techniques including semi-structured and group interviews, observation and documentary research. Moreover, the stage and the processes of the piloting study were also described. Further, the qualitative data analysis of thematic analysis is also presented as the research data analysis technique. Furthermore, it discusses the associated ethical consideration and issues of ensuring the trustworthiness of the study such as credibility and dependability. Finally, my role as a researcher is also presented in this chapter.

5.1 Research paradigm and approach

The choice of the research approach is determined by the nature of the research questions in order to provide a comprehensive interpretation of the phenomenon under study (Creswell, 2007). Before that, it is crucial for researchers to identify their own

view and philosophical stance that they would adopt for any kind of research, whether it is literature evaluation or investigating institutional, personal or group practices (Denzin & Lincoln, 2005). The philosophical stance contains important assumptions about the researchers' views of the world, which in turn underpin and strengthen their research strategy and the methods they choose as part of that strategy (Cohen et al., 2011).

In educational and social research, three key principles should be taken into consideration: ontology, epistemology and methodology. Ontology concerns theories of reality construction (what constructs reality?); epistemology is the knowing consideration (what is knowledge?, how do we know things?); and methodology deals with how we gain knowledge about the world (Matsuda & Silva, 2005).

Ontology is a system of beliefs that reflect an interpretation of an individual about what constitutes a fact (Gray, 2004). In other words, ontology is associated with a central question of whether social entities need to be perceived as objective or subjective. The ontological position taken by the researcher perceives that social phenomena and their meaning are shaped and accomplished by consequent actions and perceptions of those social actors concerned with their existence (Bryman, 2008). I consider reality as personal and variable, depending on an individual's interpretation of a phenomenon in different ways (Guba & Lincoln, 1994). This ontological position is in line with the nature of the research phenomenon (washback). The washback phenomenon could exist or not according to the researcher's interpretation of factors related to whether the test design or use affects language teaching and learning in a context. If it exists, it can then be positive or negative according to participants' responses and contextual factors in social and educational settings (Wall, 1996). The understanding of the socio-cultural context and the concerns of the stakeholders and other participants' beliefs, perceptions and attitudes play a key role in the full account of antecedent conditions that should be taken into consideration whilst investigating the washback. Based on that, washback is determined through the interpretation of different participants' perspectives and practices in the educational and social context. The understanding of whether washback exists or not, and what its nature is, mainly depends on participants' perceptions and beliefs of what derive language teaching and learning in the study context.

Epistemology is the consideration of knowing what is related to our understanding of the nature and scope of knowledge, assumptions and bases about the status of knowledge (Henn et al., 2006; Schwandt, 2003). Epistemology concerns the explanations of *what is knowledge* and *how we know things*. It also seeks explanations of the different kind of knowledge that enables us to act productively. My epistemological philosophy demonstrates that truth or knowledge is based on social interaction. In other words, knowledge is gained through a strategy that "respects the differences between people, and social scientists are required to grasp the subjective meaning of social action" (Grix, 2004, p.64). Knowledge is gained inductively through personal experiences to create a theory, arises from particular situations and is not reducible to simplistic interpretation. Since washback is a situated phenomenon, this epistemological philosophy enables me to take into consideration the individual characteristics of the research context in general and the participants in particular.

Based on the ontological and epistemological positions, I locate my research within an interpretivist paradigm. The interpretivist paradigm originates with the principle that access to reality is value-bound and largely dependent on the people experiences. This includes their interpretations, meanings and understandings of a phenomenon in real-life contexts. According to interpretivism, a reality cannot be separate from our recognition of it. Thus, the interpretivist paradigm suggests that researchers' values are a part of all stages of the research process. Interpretivists consider behavioural uniformities as an illusion – a social construction but not as an evidence of underlying or essential uniformity among entities. Therefore, people could react differently to the same situation according to their interpretations of their world (Hammersley, 2007). The distinctive point of qualitative research acknowledges and recognises the significance of the role of the researcher in the research process.

Collins (2010) described interpretivism as meaning-focused research that can potentially integrate various methods to reflect different features of the research phenomenon. It emphasises the importance of beliefs, intentions and perceptions in actual contexts. Thus, the participants' beliefs, perceptions and attitudes, and the interpretation of the full account of antecedent conditions should be taken into consideration whilst investigating the washback phenomenon. As washback has been identified as a complex concept involving various dimensions and institutional factors (Shohamy et al., 1996), the complexity of the research subject necessitates more than one data collection method. I used the interpretivist paradigm

as it adopts a naturalistic research approach. One of the strengths of qualitative research is its capability to examine complex phenomenon and offers thick textual descriptions of people's experiences of a research phenomenon. Thus, this study can benefit from this approach by including a broad array of constituent stakeholders related to the phenomenon.

I adopted a qualitative research approach. Qualitative research is interested in understanding meanings that people have constructed, using different tools to generate data that does not indicate ordinal values (Ritchie & Lewis, 2003). Therefore, I adopted the qualitative approach in order to employ different data generation methods, such as observation, interviewing and the analysis of the textual documents related to the topic to answer the research questions. That would also enable me to take into account different factors affecting or affected by washback, such as: the macro context (e.g. society) to the micro-context (the school setting); the test status within the whole educational system; test factors (test purpose, design, contents, decisions made on the basis of test results); and personal factors (participants' beliefs regarding the teaching/learning process). The meaningful discussion between me and participants via these methods facilitates the construction of a meaningful reality in a collaborative format. Watanabe (1996) has argued that incorporating a qualitative research approach in investigating washback and contextual factors is important to get deep insight of the phenomenon.

Subjectivity is one of the criticisms often directed towards interpretivism. It is sometimes claimed that interpretivist research is affected by researcher bias. Thus, findings of such research cannot be generalised since data is heavily influenced by the researcher's personal viewpoint and values. Thus, the reliability and representativeness of the qualitative data are contested by positivists. Nevertheless, the in-depth primary data generated through interpretivism studies could represent a high level of credibility and dependability in which they aim to achieve research trustworthiness. Therefore, this study adopted the qualitative research approach with the application of qualitative data collection and analysis.

The final key principle that should be taken into consideration in social and educational research is methodology, which is concerned with methodological strategy. This will be discussed in the following section.

5.2 Case study research methodology

In educational research, a case study is often regarded as a prime research strategy both to develop educational theories to evaluate and enhance educational practices. The case study as a research methodology enables the researcher to obtain a thorough understanding of the perspectives and experiences of the research participants in a natural context (Hammersley, et al, 2000; Yin, 2009). The case study is my research strategy. The characteristics of the case study encouraged the researcher to adopt it as a research methodology in this research in particular.

Firstly, Yin (2009) clarifies that a case study is in line with the interpretive paradigm to construct meaningful and real knowledge from the participants. As I locate my work within the interpretivism paradigm, the experience of the enquiry is a process of interpretation and making sense of the phenomenon under study. The theoretical framework of my research is grounded in the interpretations of the ELT practitioners' and students' perceptions of public assessment and its connection with their position and practices in the case study's context. Secondly, the case study is used to generate in-depth understanding by providing a holistic view of a situation as case study research is mostly connected to research seeking in-depth understanding by asking "how" and "why" questions. This characteristic serves the nature of the focus of the research questions about how the SECE influences practitioners' and students' attitudes, perceptions and motivation as well as their behaviour and practices in language classrooms. Seeking interpretations of such perceptions and practices represents the question "why?". Merriam (1998, 2009) mentioned that case study is the most appropriate strategy to gain an understanding of a situation, where the interest is more in the process of enquiry. Thus, case study strategy enables a rich description of the complex phenomena, such as washback in this study. It provides a humanistic and holistic understanding of the phenomenon by utilising multiple data sources, such as documentation, archival records, diaries, interviews and observations (Baxter & Jack, 2008; Denzin & Lincoln, 2005; Denscombe, 2010; McNeill & Chapman, 2005; Yin, 2009). A lot of data can be collected that enable researchers to establish systematic correlations between attitudes, perspectives and behaviours, and practices and experiences. It also enables investigation of potential changes in regard to the policy and contextual factors. In this study, the qualitative data was elicited from different language education practitioners and students through

individuals and group interviews, observations and policy documents. The generation of data from different sources is likely not only to give a fuller picture but also to enhance the test sufficiency and validity. Besides, the case study is a flexible approach which is "neither time-dependent nor constrained by methods of generating data" (Simons, 2009, p.23). For instance, interviews and observation were conducted at different times in this study. It also provides an opportunity to examine an aspect of a phenomenon in depth in a specific time limit.

Thirdly, the contextualisation of a case study makes it trustworthy since a phenomenon is studied in their natural, unique and dynamic contexts. This characteristic is compatible with the epistemological perspectives of the researchers which demonstrate that each context has special characteristics, which have direct effects on the construction of beliefs, experiences and perceptions of its participants (Merriam, 2009). The case study strategy was suitable to investigate the research phenomenon in depth from different perspectives with considering details to uniqueness and complexity of a particular policy, project, institution and programme within a bounded system in a real-life context (Creswell, 2007; Nunan, 1992; Simons, 2009; Thomas, 2011). Therefore, it can focus on the details of individuals, a group, a single social setting and an organisation or the whole country. The context plays a vital role in establishing the relationships that contribute to formulating the experiences, perspectives and beliefs of teachers and students towards language teaching, learning and assessment. All these aspects are studied in their natural context (schools and classrooms). In the present study, the context includes the schools and related aspects of individuals and the social context.

From the previous discussion, it is obvious that the most important aspect of case study research is the reference to the specific context in which it is situated. In her discussion of what constitutes a case, Merriam (1998) referred to the phenomenon unit or entity that defined boundaries or "fence in" (p.27) that a researcher can establish to determine what will be excluded from the study. The case is defined as "a thing, a single entity, a unit around which there are boundaries" (p.27). It could be limited to a certain number of people, concerns or hypothesis, an instance of some issues or a finite time frame for observations. So, the case study could be a single case or multiple cases within the same boundaries (ibid). In washback research, Watanabe (2004) mentioned that tests are used for particular contexts and specific

purposes, and it is important to identify complications that have been recognised by test users in that particular context. Accordingly, this study is a single case study and the basic unit of my case is the school context and the related issues that affect the participants' practices within that context. My case study involves English language classes of final year students in a secondary school in the south-west of Libya. That will identify the main and subordinate units of analysis provisionally in my research. The research participants would be those whose practices are connected with the case study directly or indirectly.

There are some criticisms usually directed to case study research such as the difficulty to generalise from a specific case, the amount of time taken and the volume of documentation generated. However, Stake (2008) argued that case studies in social science research facilitate the understanding the research population and issue and this involves naturalistic generalisation, which is different from generalisation in science research. Gomm et al. (2000) also stated that "case studies present a microcosm of a larger system of the whole society: that is what is found there is in some sense symptomatic of what is going on more generally" (p.99). Other researchers believe that a case study is possible to be generalisable to theoretical propositions but not to make a statistical generalisation to populations or universes (Creswell, 2003; Yin, 2009). In other words, knowledge could be based on one case study. Similar to interpretivism, subjectivity is also considered as another major weakness of case study research. However, Simons (2009) argues that subjectivity is not something we can avoid whatever methods we adopt in a qualitative enquiry where people, including the researcher, are an inherent part of the case. Research subjectivity can be decreased by acknowledging its inherent subjectivity and clearly demonstrating how a researcher's values, predispositions and feelings, and their relationship with the research site, had influenced the research process (ibid).

5.3   The study context

The case study has been described as a detailed examination of a single setting, or one particular event (Yin, 2013). Thus, the qualitative case study can be defined as a holistic, intensive, analysis and description of a single unit, phenomenon or social issue. The current case study is conducted in a secondary school in a town in the south-west of Libya. The school is a coeducation school similar to the majority of

secondary schools in the region. For that, the school can be described as a typical state secondary school in the region of the south-west of Libya. The school capacity is between 500 to 600 students; there are about 30 teachers working in the school teaching different school subjects. There were only two permanent English language teachers in the school. Those teachers teach the first (Year 10) and final year (Year 12) secondary school students. The second year (Year 11) students are mainly taught by assistant teachers who work based on annual contracts. Robson (2011) maintains that the research site should be chosen with regards to the willingness of participants to cooperate, ease of access, the logistics required for completing the research and, if possible, where contacts are already established so that research aims can be fulfilled. I chose the research site, "the school", for different reasons: safety in the first instance and willingness of participation, and convenience of accessibility. As the town where I originally planned to conduct my research had witnessed military action during the time of data collection, I changed the research site to another place where it was safe and secure for the participants and me to conduct the research interviews. I was not able to wait long since I had limited time of absence on my study visa. The familiarity with the context where I have connections with some school head teachers and inspectors of English language at the secondary school level made it easy to find an alternative. As the participant inspectors supervise more than one school, they had no problem with changing the study site as long as it was within their supervision area. Therefore, I had connections with the other involved participants in this study site, mainly the teachers and the headmaster, to get their agreement to involvement in my research study. Dörnyei (2007) stated that participants have to be willing to participate voluntarily to give in-depth information, which could maximise our knowledge of the research phenomenon. Marshall & Rossman (2006) also highlight more interest shown in conducting the study on the level of processes, people, programmes and interactions; the more likely to establish a trusting rapport with the participants and to assure data quality and credibility. That was even important in the Libyan context due to the on-going conflict and the socio-political tension in the society (more details in 4.5). Therefore, the practitioners and students' agreement to be interviewed and observed in their classrooms as a part of their participation in the research was a key factor in the choice of the school.

5.4 Sampling

There is two main methods of sampling: random or purposive sample. Qualitative research and case study research have a tendency to utilise the technique of purposive sampling (Maxwell, 2012). Purposive sampling is also a reliable technique when the research emphasis is on in-depth understanding rather than descriptive research (Patton, 2002; Tongco, 2007). So, the selection of the purposive sample employed in this study was based on its purpose. It is used to represent a small sample of participants who are uniquely able to be informative as they are privileged witnesses to an event (Maxwell, 2012). Thus, the sample of this study consisted of three groups of participants whose work is related to EFL teaching, learning and assessment in context. The first group included practitioners other than teachers who had a direct or indirect connection with language teaching and learning in the school classroom context, such as the school headmaster, teachers' inspectors and representatives of the Examination Office in the region of study; the second group included teachers (see Table 2); the last group included students.

It should be highlighted that finding alternative participant teachers to agree to participate in the research was a real task in such an on-going conflict context as Libya. In spite of my connections with many teachers in the context, I had difficulties finding teachers who would agree to be interviewed and observed. Thus, the individual's trust was of significant importance for the participants' decisions to participate in this research study. Therefore, the alternative participants were those who felt more confident about expressing their own views and perceptions with the researcher and the confidentiality of the given information. It has been argued that including people with known characteristics in interviewers that seeks attitudinal perspectives in the sample of a certain group, should be ensured since the interview is dynamic social moment based on sharing, negotiating knowledge and opinions

(Robson, 2011). Besides, choosing known participants also helped me to have contact with and access to most of them beyond the working time for the data-generation process. Other criteria were used to select each group of participants and their characteristics are going to be discussed in detail according to each group.

Table 2: Participant teachers and practitioners' profiles

Participant code	The title	The recent qualification	Work experience	Training courses
P1	The head of English language inspection unit	Bachelor degree in Education - English Teaching obtained in 1984	38 years' work as an inspector	Teaching methodology - 1986 Curriculum development - 2000
P2	The representative of the examination office	High Diploma in Education obtained in 1981	13 years' work as a school teacher, 6 years' work as school head teacher, and 15 years' work in the examination office	Teaching methodology
P3	Educational inspector	Bachelor degree in Arts – English language obtained in 1984	5-6 years' works as a school teacher and over 30 years' work as an inspector	Curriculum development - 2000
P4	English language teachers' inspector	Bachelor degree in Arts – English language obtained in 1984	15 years' work as a school teachers and 20 years' work as an inspector	Curriculum development - 2000
P5	The school head teacher	High Diploma in Education obtained in 1995	13 years' work as a school teacher, 10 years' work as school head teacher	Teaching methodology
T1	English language teacher	High Diploma in Education obtained in 1999	7 years' work as an English teacher in basic education and 8 years in Secondary Education	None
T2	English language teacher	Bachelor degree in Education - English Teaching obtained in 2004	11 years' work as an English teacher in Secondary Education	None

5.4.1 Teachers

Two English language teachers participated in the current study. The teachers were chosen on the basis that they had to be currently teaching final year secondary school students, and they had to be willing to be interviewed and observed. This is because observation was an important aspect of the research as it was employed to triangulate data-generating methods to ensure the credibility of the research findings and the strength of the research trustworthiness. Along with the second category used to

select inspectors, teachers were also chosen according to their experience of the public examinations system. They had to have experience of teaching English language to final year students for at least seven years. Moreover, they were chosen on the basis of their work under the supervision of the participant inspectors since the inspectors were chosen first to participate in the research study. Their reflection on their English language teaching strategies and practices, and the students' learning behaviour and outcomes in regard to the public examination system was used to identify any washback effect. Through that, I intended to achieve the research aim by exploring the teachers' knowledge of language assessment, how it affects their perceptions and practices of their classroom teaching and testing strategies.

5.4.2 Inspectors

Two inspectors participated in this study. The inspectors were chosen according to their experience and their work with the research site in the region. The participant inspectors were the supervisors of the participant English language teachers in their classroom context. This was to ensure that the data collected from the participants described the study context (the school) as a case study but no other in the educational context. That also enabled me to observe the inspectors' practices and its effect on the participant teachers' perceptions and their teaching practices in the study context. In addition, inspectors had to have experience of English language education for at least seven years so that they had witnessed the use of the previous and current public examination systems to be able to reflect on the perceived effect of the public examination on ELT teaching and learning practices and outcomes during their experience. Their perceptions about the role of the SECE within the educational context and its effectiveness were also used to answer the first and third research questions. Their reflection on teachers' classroom teaching and assessment practices and students' learning outcomes, was also used to answer the final research question by eliciting any changes in final year secondary school students' motivation and learning outcomes over time since the change of the public examination system.

The head of English language teaching units within the inspection office was also included in the study since he is responsible for inspection policy and practices in the study context. He was interviewed to investigate the broader policy effect on the inspection practices of overall English language teaching especially with regards to the SECE. Such information was used to answer the first and third research questions

investigating how English language teaching practitioners perceive the role of testing within the social and language education contexts, as well as its effect on language teaching and assessment practices and students' learning outcomes.

5.4.3 The representative of the regional Examination Office

The representative of the regional Examination Office was also included to represent the view of the examination department, which is mainly responsible for the assessment practices in the context. His perceptions regarding the public examination policy, system and design were helpful in exploring the relationships or weight of these practices on ELT classroom practices. He was also included to reflect on students' achievements and learning outcomes within the context, which was used to answer the third research question about changes in final year secondary school students' score improvement and their language learning outcomes over time since the change of the public examination system.

5.4.4 The school head teacher

The head teacher of the research site context was also interviewed in this study. The head teacher's role regarding the successful implementation of teaching and assessment policy in the school context is important as it shapes the teachers' teaching and assessment practices. The school head teachers are also responsible in the first instance for the execution of the school periodical tests for final year secondary school students, which are credited towards the final score of the SECE and on their Secondary Education Certificates (SEC). For that, his perceptions about the examination policy, system and design were helpful in exploring the SECE effects on classroom teaching and learning practices as well as learning outcomes in the school context.

5.4.5 Students

A total of 15 students (seven males and eight females), distributed in three groups of five students in each group, participated in this study (see Table 3). They were final year students of secondary students, aged between 17 – 18 years old. Those students were chosen from different classes of the final year taught by the two teachers who participated in the study. It should be highlighted that the participant students had different experiences and backgrounds of secondary school education since they came from different villages in the region. Most of them moved from their previous

secondary schools due to the shortage of teaching staff in different subject areas, especially in the final year. Since a group interview technique used to generate data from students relies on group interactions and dynamics, it is possible that during interviews in such context for conflicts to occur within the group due to personalities and for a power struggle or disputes of status (Robson, 2002). As I was working in an on-going conflict environment, careful consideration was given to the group dynamic, ensuring that students would be willing to participate in a discussion with each other in the same groups away from their socio-political perceptions and attitudes. Therefore, individual students within the groups were selected by the headmaster, as he knew the students' personalities and their characteristics and socio-political perspectives. To ensure that this type of sampling does not affect the representation of the sample, no other classification was used. For instance, I ensured that students should not be excluded because of their socio-political position and participants were not chosen according to their academic level or specific educational or social background.

Students were interviewed to explore their knowledge of the English exam of the SECE, and its score use. They were also interviewed to explore their perceptions, beliefs of characteristics, advantages and disadvantages of the SECE, as well as the English language classroom tests and the effect on their language learning motivation and strategies. These types of data were used to answer the research question regarding teachers' teaching and assessment strategies, and the question regarding the effect of SECE on students' motivation, learning strategies and outcomes.

Table 3: the participant students' profiles

The participant code	Gender	Specialisation	Desired score on the SECE	Desired university major
S1	Male	Science	Excellent	Engineering
S2	Female	=	-	Medicine
S3	=	=	High score	-
S4	Male	Arts	High score	-
S5	Female	Science	Very good	Education - English
S6	=	=		Medicine
S7	=	=	High score	-
S8	=	=	-	Medical Laboratory
S9	=	=	-	Education - English
S10	=	=	Very good	-
S11	Male	Arts	High score	-
S12	=	Science	Very good	-
S13	=	Arts	=	-
S14	=	Science	Excellent	-
S15	=	Arts	Excellent	Law

5.5 Data-generating methods

According to the nature and complexity of washback, different data-generation instruments were used in this research study: semi-structured and group interview, observation and documentary research. These methods of data generation are intensively used in washback research studies as being suitable to generate in-depth data while researching washback in context (Tsagari, 2009; Wall, 2005; Watanabe, 2004). The data-generation instruments were used collaboratively to answer each

research question and gain a comprehensible view of the washback effect (see Table 4). Hyland (2002) notes that using a multi-faceted data-generating approach helps to overcome the deficiencies of any particular method and offers a way of cross-checking data. Further detailed explanations and the justification for the selection and action processes of the instruments individually are presented and discussed in the following sections.

Table 4: Data generation instrument used to answer the research question

Research questions	Data generation instrument used to answer the research question
To what extent does the Libyan public examination of the SECE have an impact on motivation, perceptions and attitudes towards English language teaching within secondary school education?	Interviews Observation An analysis of policy documents related to testing practices and teachers' exams
To what extent does the public examination of the SECE influence English language teaching, assessment and learning strategies in secondary school classrooms in the on-going conflict context of Libya?	Interviews Observation An analysis of a copy of teachers' exams
How does the Libyan public examination of the SECE influence the students' English learning outcomes within secondary school education?	Interviews Score analysis

These techniques were used with different groups of participants as showed in the following figure (Figure 4) that illustrate the research design.

Figure 4: the research design

Research Paradigm	Interpretivism
Research Approach	Qualitative

Case study

Participants	Data generation methods	Data analysis
Students	Observation · Group interviews · One-to-one Semi-structure interviews · Documentary research	Thematic analysis
Two English language teachers		
Two educational inspectors		
A representative of the examination board		
The school head-teacher		

5.5.1 Interviews

Rubin & Rubin (2005) stated that interviews are usually used to establish people's perspectives about things. This study relies on interviews as the main data-generating technique to get rich information from the participants about their perceptions of experiences of washback for different reasons. Epistemologically, interviews can be considered as the heart of human interaction for producing knowledge and emphasising the social context of the research data (Fontana & Frey, 2005; Kvale, 2008). Hence, interview data were considered as participants' interpretations of the world in which they live. That epistemological point of view is in line with my research's theoretical perspectives. For that, interviews are the most common data-generating instrument in empirical research of washback both in general and language education (e.g. Andrews et al, 2002; Qi, 2005; Wall, 2005; Wu, 2014). Wall & Horak (2006: 26) declared that interviews are the best way "to gather data that would provide deeper insights than questionnaires, even if it meant working with fewer participants" in

washback research. In addition, interviews are a good opportunity for participants to reflect on events where meanings can be immediately clarified and explained. Interviews were conducted with teachers, students, inspectors and other participants exploring their perceptions, attitudes and experiences of language teaching, testing and learning in secondary school education. They were interviewed to reflect on their concern about educational operations, processes and outcomes, and the perceived changes as a result of their involvement in the reforms of public examination system in the research context.

There are three types of interviews: structured, semi-structured and unstructured interviews (Robson, 2011). The semi-structured interview and unstructured interview are considered as in-depth interviews since they adopt a flexible design and utilise open-ended questions, unlike the structured interview. This flexibility enables researchers to be flexible and responsive to the interviewee conversation to obtain additional information and clarification to obtain a deep and in-depth understanding of the research phenomenon. These interview strategies are often used to unfold complex processes through providing a complete picture of what happened in the programme and why in the context (Boyce & Neale, 2006). Thus, I used semi-structured interview techniques to achieve the objective of the research of gaining a deep and in-depth understanding of washback in the study context.

5.5.1.1 Semi-structured interviews

A semi-structured interview technique was appropriate for this study because it is flexible and interactive in nature (Arksey & Knight, 1999; Bryman, 2008). The interview is more a guided conversation than a staccato questions and answers session. Such kind of interaction was an important aspect when interviewing participants, who have a lot of experience in teaching or hold a high position, such as school head teachers, inspectors and the representative of the regional Examination Office. The flexibility, in which the questions' sequences, forms and wording can be changed according to the participants' responses and focus, eased the flow of the interaction. It made them feel as they were leading the conversation, especially when there was a degree of informality between the interviewer and the interviewees. It helped me to get a thorough considering all the factors that promote participants' answers: opinions, reasons, feelings and beliefs. Hammersley (2008) argued that such established rapport encourages participants to talk freely in ways that can reveal the

distinctiveness and complexity of their perspectives that the researcher seeks. The interviews were conducted in the mode of face-to-face and one-to-one semi-structured interviews. Listening to the interviewees' voices and observing their physical responses is important in interviews. Face-to-face interviews enabled researchers to notice the participants' unobservable aspects such as feelings, thoughts and intentions, behaviours, how people have organised meanings with respect to the world and their practices within it (Patton, 2002). One-to-one interviews were also important in such an ongoing conflict context as they enabled the interviewee to: feel free and comfortable, be more willing to discuss any personal factor relevance to the research and to discuss their perceptions and attitude of the educational policy and practices, including the public examinations.

The interviews consisted of core and supporting questions to encourage depth and detail. These supporting questions were based on the general conceptualisation of the existing theoretical knowledge of washback studies to address the research questions and the aims of the research. I identified prior' themes from reviewing the literature (e.g. Burrows, 2004; Weir, 2005; Wu, 2014; Yang, 2013) with the key themes of my research aim and questions to build the interview guide themes (see Appendix 3). The interview guide questions and prompts made it even easier to encourage extensive responses in depth and for the data to be thematically transcribed, categorised, and then interpreted. There were different areas of focus with teachers as a group and school head teachers, inspectors, the representative of the regional Examination Office and the head of the English inspection office as another group of participants (see Appendix 4). However, all of the interview questions were based on five main areas: perceptions of the assessment practices in an education context; the SECE of the English language subject (characteristics, advantages and disadvantages); perceptions and attitudes towards the effect of the SECE on teachers, students and educational practices; English learning outcomes; and finally, the general score use of the SECE and the English language subject in particular.

Legard et al. (2003) recommend that researchers should have potential contact with the intended sample in two stages: at the initial point of contact asking them to participate in the interview; and another time to contact the respondents to determine a convenient time and place for them to conduct the interview. I had contact with some participants as inspectors and the representative of the Examination Office to organise

the interview time and place as well as to discuss the process of signing the consent form. However, the interviews with the other participants of the study were organised after I met them in the study context after I changed the original plan. I had met the participants to explain the nature of the research project and my intention to interview them. A copy of the research prior themes was given to participants to get an overview of the kind of topics which would be discussed in the interview. The choice of the place and time to conduct the interviews was given to the participant to ensure that the interviews could run smoothly. The interviews were introduced by "a briefing" (Kvale, 2008, p.56) in which each participant was informed about the aims of the research, research procedures, the use of an audio recorder and participants' rights. The interviews lasted approximately forty-five minutes and all interviews were audio-recorded. The interviews were conducted in a safe, quiet and comfortable place to encourage informal interaction. It was long enough to make them feel that their discussion was being taken seriously. Due to my relationship with some participants, some interviews lasted up to one hour with teachers and inspectors as participants shared a lot of stories and memories. The Arabic language was used to conduct all interviews so that I ensured that the interviewees could express their feelings freely without any language barrier and to ensure that there was no misunderstanding between myself and the interviewees. Arabic was also used to avoid the hindrance of having to speak English in front of me. Such a decision was made to build a rapport with the focus of the research on washback on teaching and learning rather than to judge their English ability. That was important in this context since the interview language was one of the basic enquiries when I sought participation in the research. The discussions involving the personal experiences of the participants in this study were facilitated by using Arabic as the medium of communication.

5.5.1.2 Group interviews

At the preparation stage of the data-generation instruments, I planned a focus group with the participant students. I had used the focus group technique when piloting the data generating instruments (more details in 5.6). However, the participant students had difficulties sustaining the discussion in the focus group during the research's actual data generation process. Since it was their first time experiencing such a technique, they were confused regarding my role as a researcher in a focus group. Although I had explained the techniques several times so that they could sustain the

discussion, they had difficulties in doing so. They told me that I needed to contribute to the discussion with them to be able to sustain the discussion. Therefore, I decided to switch the technique to a group interview. The group interview is a type of interview in which a researcher simultaneously discusses one phenomenon with two people or more (Patton, 2002; Robson, 2002). It was mentioned that group interview could be used if there was not much time to collect information individually when interviewees are cooperative with each other, as in my case (Creswell, 2007). The participant students knew each other as they were taught by the same teachers. Therefore, a participant can fill in details that the other has omitted and elaborate on each other (Gray, 2004; Patton, 2002; Punch, 2013). The researcher can gain a better understanding of the situation through the interpretation of the group interaction and elicitation (Berg, 2009).

I conducted three pre- examination group interviews and one post-examination group interview with final year secondary school students. The group interviews were conducted with the students before undertaking the public examination and one group interview conducted with some students after that. The pre-examination group interviews were conducted in the school. The interview was based on the knowledge of the SECE (characteristics), the perceived advantages, disadvantages and score use, the focus of English language teaching and learning in the classroom, classroom assessment focus and practice, and their relation to the SECE and learning outcomes. All pre-examination interviews lasted 45 minutes as planned. With the aim that participant students would not need to miss their classes for the interviews, we organised the time of the interview at the time when the students did not have classes. Since the students had not got teachers in some subjects and because of some teachers' absences, students were available to conduct the pre-examination group interviews on the same day. At that time, although the town where I conducted my study was safe and away from the conflict, several teachers who come from other towns were absent for their classes.   Those teachers had had difficulties with transportation to reach their classes, however, the post-examination interview was conducted only with some students in another place as the students finished the school and joined different colleges. The post-examination interview was basically a reflection on their performance in the SECE with regard to their experience of classroom English language teaching and learning strategies and practices. In

addition, students were asked to reflect on the difference between their perceptions in the pre-examination and post-examination 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.

5.5.2 Observation

Observation was utilised in this research to collect data regarding the nature of the classroom teaching and learning strategies and practice. The observation is considered as a rich source of data associated with ethnographic methodology and case study design in particular since it studies people in their natural context or settings. Robson (2002) described observation as the most appropriate technique for getting directly at real-world life in natural settings to record people's actions, behaviours and interactions in order to obtain a detailed narrative of social or educational settings or events. Through analysing and interpreting generated data, the researcher can situate people's behaviour within their own socio-cultural context (Gary, 2004; Hennink et al., 2011). This gives observation a high level of authenticity that is rare in other techniques. In washback research, the inclusion of observation as a data-generating technique is of great importance since examinations are described as situated practice. The use of classroom observations to generate detailed descriptive data is important to understand the complexity of teaching and learning practices and related factors. It is also important to support the findings from other data-generating methods for all the research questions (Bailey, 2001; Watanabe, 2004). Wall & Alderson (1993: 65) argued that direct observation of the classroom practices is required in order to contextualise incomprehensible responses and inform interviews. In this study, observation is used with semi-structured individual and group interviews.

Another important reason for using observation is that it enabled me not only to establish the presence of washback but also the absence of predicted types of test effects. The observation enabled consideration of unintended as well as intended washback through identifying the intention of the teachers and/or learners underlying their behaviours in their classrooms. Observation gave me the opportunity to discover issues that participants were not willing to mention in the interview. It also enabled me to note routine actions among people in the context to inform certain effects and their impact on the classroom context. (Robson, 2011) stated that data obtained from

observation should serve as supplementary to substantiate information generated by means of interviews.

The observation schedule used in this study was adopted from Shih (2009). It was used to observe aspects of interest in the research questions, such as interaction and talk about testing and assessment, type of activity and exercises, and the use of textbooks (see Appendix 5). This adoption of the instrument from a previous study can ensure credibility through the adoption of well-established research methods (more discussion in 5.9.1.1). Field notes, along with observation and visits to the school, were used to interpret the generated data. Bailey (1996) described field notes as the prop of generating and analysing field data as it develops an analytical process and gives an opportunity to record any emerging variable of the phenomenon in the research context.

From the different role of the observer mentioned in the literature (participant-as-observer, observer-as-participant and complete observer), I took the role of a complete observer. In the complete observer role, the observer takes no part in the session or activity; only observes and records field notes when it is possible. I took this role to reduce the reactivity aspect of the observation by minimising intrusion and avoiding interference with the interaction patterns of the observees in the setting of the study. Although the teachers agreed to audio-record their classes, they were very anxious about doing it. Their agreement signalled appreciation for the researcher and his work as well as their trust in him. However, I did not audio-record the classes to avoid disrupting the naturalness of the setting and the participants. After the observations, I tended to open up an informal discussion about points that had come to my attention to gain clarification and explanations from the observees regarding the classroom event and practices. That information from the discussions was also recorded as part of the field notes.

During the first time of data collection, I could not conduct any observation as schools ended up closing due to the conflict, because of transportation difficulties due to the shortage of motor vehicle fuel. Therefore, I travelled again to conduct a direct physical presence observation. I observed six classes for the teachers for one week (three classroom periods each). The observation included the inspector's visit to the teachers' classes. That was to observe the type of comments and advice that inspectors gave to teachers and how that could affect teachers' practices. This data

helped to examine the connections between inspectors' and teachers' perceptions and attitudes towards assessment policy and its impact on their classroom language teaching practices to answer the first research question. Observation data about the students was also used to answer the second research question of the public examination policy and practice on the students' learning strategies and practices.

5.5.3 Documentary research

Documents play an explicit role in any data collection in doing case study research (Yin, 2009). The overall value of documents enables them to provide substantial support for the interpretation of interviews and observational data that enhance the study with evidence. Creswell (2005) indicates that documents can be used as an independent method for generating data in research. However, Denscombe (2003), as well as Cohen et al. (2011), argue that with careful analysis and interpretation, it could also be treated as a supplementary instrument to help find accurate answers to the research questions. Thus, I used this technique both as an independent technique to achieve triangulation and as a supplementary technique to clarify responses and practices. Wall & Alderson (1993) argued that the analysis of documents, such as textbooks, teacher-devised materials, examination records and student portfolios, can provide further evidence of classroom teaching and learning practices. It also helps to contextualise any incomprehensible responses from interviews in the analysis. The analysis of classroom-based tests and public examinations and other policy documents is conducted to minimise the research bias and get in-depth insight/explanation of the phenomenon and to ensure trustworthiness of the generated data through triangulation.

During data generating, I was able to gain different documents that have a direct or indirect relationship to the classroom assessment and the SECE from different participants. For instance, I obtained samples of the teachers' classroom tests and the school periodical tests. I also obtained a sample of the English language exam of the SECE from participants. These samples were used to examine the relationship and the effect of the public examination on the teachers' classroom testing, which could be considered as a sign of washback. Other documents collected were the teachers' evaluation forms, textbook teaching plans, students' assessment criteria and a copy of the teacher development courses. These documents were used to give a clear explanation of different practices of the practitioners and students. Such explanations

give more description of the contextual data, which characterise the social aspects of the research framework which helps in the analysis, understanding and interpretation of the generated data.

Moreover, documents about the rate of success in public examination in recent years were also collected. This helped me to perceive the effects of implementing the recent public examination system on the overall school students' results in the study context. These statistics were compared with the data generated from the teachers' and inspectors' perceptions about the achievements of the school-leavers students. These two steps can be used to identify any impact of assessment on students' results and tracing perceptions about their language learning outcomes as shown in the following table.

Table 5: The type of documents collected to generate data

Document collected	Data analysed
A sample of the English language exam of the SECE	The content of the exam The focus of the exam
Samples of the teachers' classroom tests and the school periodical tests	The content of the exam The focus of the exam
Teachers' evaluation forms	The focus of the policy in the evaluation
School English language textbook content and teaching plans	The content and its relationship to the SECE The focus of the plan The aimed level of English language ability
Students' assessment criteria	The credit of the final exam towards the overall score The credit of language skills towards the overall score The credit of classroom work towards the overall score The credit of classroom-based assessment towards the overall score
A copy of the teacher development courses	Areas of concern Areas of development Given solutions for practice
Documents about the rate of success in public examination in recent years	The statistical record of the overall school students' achievements in public examination

5.5.4 The researcher's diary

The researcher's diary represents the researcher's personal experience of the research process. It offers data about the methodological decisions made on a daily basis and the reasons for making them or as needed. It also records a variety of information about self and method. The importance of such diaries was emphasised by different researcher methodology writers (e.g. Borg, 2001; Finlay, 2003; Finlay & Gough, 2003). In the current study, the research diary was used for reflection, which helped to deepen my understanding of different aspects of the research processes.

I kept my own diary to record the difficulties and issues which occurred during the process of data generation. I also recorded the methodological decisions and the motive behind them during the research process. The diary contained a detailed chronological record of days and it also contained a description of the classes observed, and the research site and field notes of non-verbal communication, such as the tone of voice and facial expression (Gray, 2014). I recorded notes about participants' reactions and my impressions of their responses. Keeping a research diary assisted me to record my experiences, thoughts and feelings, opinions visible in order to be acknowledged in this research how it influences the data generation, analysis and interpretation process (see Ortlipp, 2008). These reflections had a significant influence on the research findings, which is evident throughout the data.

5.6 The piloting of the interview questions

A pilot study is defined as a micro study conducted prior to a planned project (Yin, 2011). It is used to test aspects of a research design (including refining research questions, stimulus data-generation methods, estimating the time and costs) in order to allow necessary adjustment before the final commitment to the design (Robson, 2011).

I piloted the interview questions in December 2014 with five acquaintances who were Libyan students who took MA TESOL at the School of Education and Professional Development at the University of Huddersfield. The pilot study was conducted to check the clarity of the questions for effective practice in the actual interviews. The acquaintances were interviewed through a focus group to pilot the approach to the students. The teachers' interview questions were also piloted by interviewing two of the acquaintances who have experience in teaching English at state schools. The pilot

focus group and interviews were transcribed and initially analysed using thematic qualitative data analysis. The pilot study provided the researcher with an opportunity to check the intelligibility and unambiguousness of questions in order to revise, remove and add to the questions and sub-questions. The pilot study confirmed that the questions were clear, acceptable and answerable. From the pilot study, I perceived that 45 minutes was an optimal time for conducting interviews to avoid making participants impatient since after that no new information seemed to be obtained from them.

The piloting stage was beneficial for assessing the interview protocol and allowed me to familiarise myself with the process involved in carrying out a focus group. I realised the difficulties associated with the translation and transcription of the interview data. Transcribing the interviews also enabled me to thoroughly understand the data and the distinctiveness of each interviewee. Likewise, it enabled me to get the experience of dealing with the data such as how to identify and retrieve the perceived significant data from the transcript.

5.7   Issues of ethical consideration

Research ethics concerns protecting participants involved in a research from emotional, psychological and physical harm (Cohen et al., 2011; Silverman, 2010). As educational research involves people and their institutions, it was important for this research to protect and respect individuals' identities and practice (Cohen et al., 2011) as well as sites. As dealing with these issues was a priority in this study, the Ethical Guidelines for Educational Research (2011) from the British Education Research Association (BERA), as well as the Code of Ethics for the International Language Testing Association, were used as the benchmark for the ethical processes in this research study. Language testers developed the Code of Ethics for the International Language Testing Association to offer satisfactory ethical behaviour when researching language testing, especially when it includes test-takers – participant students in my case. The principles of the code highlight the importance of considering all relevant ethical principles embodied in public guidelines when undertaking any research. For that, the two codes guided the research study. The following section discusses the ethical issues, taking into consideration acceptance and consent, access to the research site, confidentiality, anonymity and safety.

5.7.1 Acceptance and informed consent

From the beginning, the participants in this research were clearly informed of the purpose and the significance of the research study to the context. Cohen et al. (2011) and Creswell (2007) reported that such action encourages participants to give comprehensive responses and consequently increases the reliability and validity of the research findings. The participants were given an abstract of the objective and the focus of the research. One of the teachers asked about the nature of the questions which would be asked in the interview before they gave their consent. Therefore, I gave them a copy of the themes of the interview guide but not the interview questions. That decision was made to be able to get the data in the most honest and complete format. The participants were also informed about my plan to record the interviews for personal reference. Consequently, all participants signed a consent form before giving any information or documents related to the research as recommended by Bryman (2008). I adopted the University of Huddersfield consent form for my study (see Appendix 6). They were also verbally informed of their right to withdraw from the research at any time and that I would understand and respect their decision without the necessity to give reasons behind that. In that case, all their given information would be erased. However, none of the participants highlighted an intention to withdraw from the study.

5.7.2 Access to the research site

It is important for researchers to get the required consent and permission to access the research site to legitimate the data generation process (Aldridge & Levine, 2001; Bell, 2014; Gay & Airasian, 2003). Thus, I obtained a consent letter from the education authority in the region to reach teachers in their schools and inspectors in their offices. First, I gained an official letter from the University of Huddersfield to state the process of my research in which I was permitted to collect data procedures. That letter was directed to the Cultural Affairs Bureau within the Libyan Embassy in the UK, to issue a further authorisation letter to be used for the Secretariat of Education in the region where the research was conducted. Then, the regional educational authority issued a letter to the schools where the study was conducted (see Appendix 7: a, b & c). The same letter was used with other participants such as the inspection office participants and assessment office.

5.7.3 Confidentiality

Confidentiality deals with the protection of the participants' rights to privacy at the beginning of the research to gain the participants' trust. Confidentiality encourages participants to speak openly and in good conscience (Simons, 2009). Thus, I ensured and guaranteed all participants that the data collected would be confidential and only used for research purposes. Since I conducted interviews with teenage students, confidentiality was an important aspect for them, so that they were willing to talk about their actual perceptions of the classroom testing and teaching practices. Therefore, I ensured that their information would not be accessed by their teachers and other participants in the context, although it would be anonymised and used for research purposes. Practically, as the interview data was electronically recorded and transcribed, it was securely stored in a personal protected computer with password access. The field notes, observation sheets and other documents were kept filed in a secure place with anonymous references.

5.7.4 Anonymity

In terms of anonymity, the researcher ensured that participants' real names were not published in the research or transcripts. The participants were given numbers to be used in the research, such as P1 and P2 etc., whereas the teachers were numbered using T1 and T2. The students were given a different number strategy, using S1 and S2, etc. I also ensured that no information would be used which could lead to the identification of the research site. I also guaranteed the protection and privacy of the information access.

5.7.5 Safety

Safety concerns the avoidance of causing physical or emotional harm to participants (Fontana & Frey, 2005). In this research, safety was carefully considered on different levels: the physical and emotional protection from harm as well as the safety of the participant position and status. First, as the region was undergoing an armed conflict, making appointments for interviews was based on safe access, time and the place for both me and participants. That was to avoid the violent areas and ensure the physical safety of myself and the participants. Working in an ongoing conflict environment, such as Libya, brings further issues to be taken into consideration to ensure safety. Due to the demographic social construction of the study context, the participants' social

identities cannot be ignored, especially their socio-political perspectives. As the Libyan context was still in a state of flux following the conflict of 2011, careful consideration was given to the dynamic of the students' group interviews. The sample selection aimed to ensure the physical and emotional safety of participants, as mentioned in section 5.5.4. The decision of the choice of the students within the group interviews was also made for safety purposes. As explained in the sampling section (5.4), as the headmaster knew the students personally and their characteristics and socio-political perspectives, he made sure to avoid engaging those who could lead the discussion to a volatile situation. As a researcher, I treated all participants with dignity and respect without any regard to their ideas and political beliefs.

Regarding the safety of position, I assured that the generated data would not be handed to or used by any official or unofficial authority which could negatively affect their status or practices of the participants at any level and under any conditions. In addition, I ensured that the participants' attitudes, perceptions and problems would not be mentioned to the other participants or discussed with anyone else.

5.8   Data analysis

Qualitative data analysis and interpretation aim to make sense of ample amounts of data in a framework in order to explain and present the findings that the data reveal. It is a systematic process whereby we search and arrange data in order to increase our understanding of the data, and to enable us to present what has been learned to others (Baxter & Jack, 2008; Maxwell, 2005). Qualitative data analysis and interpretation is a challenging task in the research process. The process of all qualitative data analysis involves organising and reducing the data, synthesising information, identifying significant and/or important patterns and relationships, theorising about how and why the relationships appear as they do, and reconnecting the new knowledge with what is already known (Given, 2008). These processes can be used to develop theories, create explanations or pose new questions or direction in the field. As the qualitative data is usually generated over a period of time, that requires starting the process of data analysis simultaneously with the data generation. This helps researchers to reduce the overload of data and offers the opportunity to become more familiar with the data. It also enables classification of significant themes that can be explored further in the remaining time of data generation.

Since the interview questions and observation schedule were built on a pre-theme, the generated data is thematic in nature. Therefore, I used thematic analysis as the analytic procedure to interpret the data of this study. Silverman (2010) defined the theme as a pattern found in the data information that describes and organises the possible observations or interprets features of the phenomenon. Thematic analysis procedures have been defined through different steps or processes by different researchers. For instance, Braun & Clarke (2006) identified six phases of thematic analysis: the familiarisation with your data; generating initial codes; searching for themes; reviewing themes; defining and naming themes; and producing the report. Ary et al. (2010) described thematic analysis in three main steps: organising and familiarising; coding and reducing; and interpreting and representing. I adopted the process of Ary et al. (2010) as it is an inclusive classification. It helped me to work flexibly to generate initial themes and then to reduce and categorise to identify the codes, and to develop and revise my work with reference to the other described process of data analysis (e.g. Creswell, 2007; Flick, 2014; Marshall & Rossman, 2006; Maxwell, 2005). These processes are discussed in detail in the following section.

5.8.1 Familiarising and organising

The first stage of thematic analysis began with familiarisation with the data through listening repeatedly to ensure data can easily be retrieved and identify other data of great importance and value for the data analysis process (Ritchie & Lewis, 2003). Preferably, transcriptions should be made of all data, including tape-recorded interviews (Guest et al., 2012). Therefore, I transcribed all interviews with teachers, teachers' inspectors, head teachers, head of English language subject units of the inspection office and the representative of the Examination Office. Transcribing the data by myself has helped me to recall the memory of the interviews' context and generates initial codes and patterns within the interview data and other data-generation methods such as field notes.

Transcription was conducted in two stages: transcribing the audio record in Arabic and translating the transcript into English. The first stage was transcribing the audio records in the original language used in the interview (Arabic). I decided to transcribe the whole of the recorded interviews and not to summarise them by considering only the words, sentences or other utterances related to the focus of the questions. This process started during the data-generation time. Before starting the second stage of

translation, the interview transcripts were verified with the participants to ensure its reflection of the participants' responses during the interviews since the plausibility, sturdiness and confirmability of meanings have to be ensured in such a process. For those whose interviews had not been transcribed during the time of the data generation, the transcription was sent to them via email to check it. All participants confirmed the consistency between their original ideas and the transcripts.

The second stage was the translation. In spite of my experience in translation and interpretation, the translation was not an easy task where pragmatic knowledge of the social use of words was very complicated; it was important to identify the original meaning of the words and its meaning in the translation. For instance, the participants used the word "مساعدة" [help] with different meanings in different situations in this study. In teaching and learning, it meant [to facilitate/to ease] whereas in the assessments it was referring to [cheating]. Therefore, I adopt the literal translation of the meaning. I translated the meaning from Arabic into the appropriate meaning in English in order to achieve equivalence in meanings as much as possible in order to convey the message clearly. Translating accurately is an ethical responsibility, as well as important for the validity of the data. Occasionally, I complemented or replaced concepts, which could not be translated exactly from Arabic to English as the equivalence or precise meaning does not exist. This process is known as "transliteration" (Regmi et al., 2010). I reviewed the translation repeatedly to be sure of achieving the most accurate meaning and highlighted some chunks as important. For the credibility of the translation, I used the member-check techniques to validate translation. For confidentiality of the data, I gave separate parts of different interviews, but not the complete interviews, to a friend who is trustworthy and a specialist in English – Arabic translation to check these translated parts, especially those highlighted as important. This process was done conversely (English to Arabic and Arabic to English). A high level of compatibility was found between the interpretations. I read each written transcript several times in order to immerse myself in the data and to seek a better overall understanding of the whole contexts. The member check gave some advice to take into consideration in different aspects of the translation process. Although those processes were time-consuming, I did that for three reasons. The first reason was to ensure that the whole picture of the context of each episode of the interview was kept original and clear. The second reason was to use some parts of the two versions to check the meaning by member

check, especially the used episodes. Finally, the process enabled me to live with the data set, familiarising myself by grasping data and organising it.

Dörnyei (2007) used the term of the "pre-coding" stage of qualitative analysis and described it as a simultaneous step in the transcription process. Therefore, I started writing my general comments and notes about initial thoughts and interesting issues that were emerging from the data in the margins. All these strategies were the beginning of creating codes, sub-codes and themes.

5.8.2 Coding and reducing

The beginning of creating initial codes occurred through different processes of the research. Basically, this initial analysis occurred between interviews as part of the basic preparation procedure before meeting other participants. Maxwell (2005) argues that researchers should begin data analysis immediately after finishing the first interview or observation and continue to analyse the data as long as they are working on the research. This enabled me to become familiar with the data so as to classify significant themes that can be explored further in the remaining time of data generation. I wrote down the important points and ideas that had been discussed in each interview, and how these could be developed in the next interviews. At the same time, I kept track of categories or patterns emerging from the data with brief descriptions and identification of possible relationships, trying to create initial codes in Word files (see Appendix 8).

Creswell (2007) identifies five phases for coding process; reducing the data into meaningful fragments, naming these fragments into codes, categorising the codes into themes, and making comparisons between themes, and displaying the final themes. I followed a similar pattern in coding my research data. As explained earlier, the beginning of the coding process was incorporated into the transcription and translation process. I highlighted the important and meaningful segments in relation to the phenomenon under study in the transcript and noted down my own comments. These codes were generated about a word, a phrase, sentence or sentences, or a paragraph. Finally, the codes were categorised and finally defined, described and re-written until I was assured of their appropriateness and interrelatedness.

Identifying themes involves organising different codes into potential themes and grouping all the relevant coded data and extracts within the identified themes.

Essentially, this process was the start of grouping and combining different codes to form an overarching theme (Braun & Clarke, 2006). I combined some codes that carried the same idea and compared these codes that had different ideas to identify themes as showed in Table 6. Some of these codes were grouped according to the research pre-themes identified based on the review of research on washback research and the research's theoretical framework. Other emergent themes were categorised based on the interpretation of data segments. In this process, it is important to ensure that thematic data fit together meaningfully and distinctively across the identified themes.

Table 6: The research themes, codes and sub-codes

Curriculum, assessment and pedagogy

- The curriculum and assessment alignment
- The power of public examination policy
- The curriculum and timetable
- Motivation and assessment policy
  - The SECE 's score use

The SECE assessment in practice

- Perceptions and attitudes towards the SECE approach

  - Public examination administration
  - Grading in public examination
  - Policy implementation
- Evaluation of the SECE approach

Classroom experiences

- Teaching strategies

  - Teacher-centred approach
  - Grammar Translation Method
- Classroom activities
  - Teaching to the test
- Assessment strategies

  - Enhancing students' performance
  - Eliminating cheating
  - Teachers lack of language assessment literacy
  - Self-protection
- Learning strategies

  - Translating the textbook
  - Engaging in test preparation activities
  - Rote-learning
  - Developing coping strategies

English learning outcomes
- The candidates SECE score
- The Level of English language knowledge and ability

5.8.3 Interpreting and representing

In the final process of the thematic analysis, I reviewed and refined themes until they captured the coded data and the extracts. At the end of this stage, I had a sense of how to fit themes together in order to tell and discuss their story in relation to the research questions. This is evident in the following chapters (Chapters six, seven and eight). My research journal was used to record my reflection on the research process and the collected data was an important analysis tool in this study. I was constantly involved in an analytical process of both direct interpretation and categorical aggregation as Stake (1995) recommended.

5.1 Trustworthiness

In response to the criticism made by positivist researchers, the terminology "trustworthiness" is used in interpretivist research as an equivalent measurement of the concepts of validity and reliability in positivism (Guba, 1981; Shenton, 2004; Silverman, 2001). The term was developed to show the differences between positivism and naturalistic research. Guba (1981) constructs four criteria that should be considered to ensure trustworthiness in qualitative research compared with the concepts in positivist research:

1- Credibility (internal validity)
2- Transferability (external validity/generalisability)
3- Dependability (reliability)
4- Confirmability (objectivity)

Trustworthiness criteria can be functionalised through different strategies and considerations within the research process. To decrease threats to trustworthiness, I employed different strategies that assisted me to authentically present research findings in the meaningful way described by the participants in this study. Guba's four criteria are discussed in the following section to explain the different strategies used throughout the research process to ensure trustworthiness.

5.1.1 Credibility

Credibility refers to the processes that attempt to determine that a true picture of participants' views of the phenomenon under scrutiny is being presented (Erlandson et al., 1993). Different techniques can be utilised to establish confidence that those

views of the participant have accurately recorded the phenomenon under scrutiny. This study used techniques such as the adoption of well-established research methods, prolonged engagement, triangulation, member checks and a thick description of the phenomenon under scrutiny. This section discusses some of these techniques that are relevant to this study.

5.1.1.1 The adoption of well-established research methods

Yin (2013) recognises the importance of incorporating specific procedures and the methods of data collection from those that have been successfully utilised in previous comparable projects. Therefore, I considered some pre-themes from different research with a similar focus and characteristics to the study context to guide the overall research techniques for the data-generation process. Some of the themes of the interview guides were adopted from previous research (e.g. Burrows, 2004; Wu, 2014; Yang, 2013) and developed to suit my research aims and context. In addition, the observation schedule was adopted from Shill's (2009) research

5.1.1.2 Prolonged engagement

Prolonged engagement refers to developing an early familiarity with the culture of participating organisations to establish a relationship of trust between the researcher and participants before the first data collection dialogues take place. It can enable researchers to be aware of the cultural/social setting and thus help in understanding and interpreting misinformation introduced by distortion (Lincoln & Guba, 1985). As a result of my work experience in the region of the research context, I had quite good experience with the school education system in general and within the region. I took into consideration Silverman's (2000) argument of the need to be aware of the influence of such a relationship on my own interpretation of washback in the context. I also developed a good trust relationship with participating practitioners, and I was in contact with them before and during the whole period of data generation. This engagement entailed contact with those participants before and after conducting the interviews and even after I travelled back to the UK. Most of them offered for me to contact them if I needed any further help or explanation regarding my data.

5.1.1.3 Triangulation

Triangulation is considered as one of the great strengths of qualitative research in which evidence can be collected by using multiple or different data generation sources

and instruments (Rowley, 2002; Stake, 2005). Concerning this case study research, triangulation was attained by combining three different data collection instruments (interviews, observation and document research) and different data sources of information (e.g. students, teachers, inspectors and documents). These methods and sources provided rich data that allowed sufficiently deep understanding and insightful interpretations on the washback effect in the context. Credibility was also enhanced through minimising the potential distortion that could be elicited from the use of a single data source by comparing data obtained from different sources and instruments. This provided different perceptions and practices which offer opportunities to clarify meanings and enhance getting a deeper insight into the washback phenomenon.

5.1.1.4 Member-checking

A copy of the transcripts was taken back to the participants to check if they represented the ideas they mentioned in their interviews. They were satisfied with the accuracy of the transcribed interviews. Besides, different parts from different translated transcriptions were checked by a specialised friend for its accuracy as discussed earlier (section 5.8.1). This process has been done conversely (English to Arabic & Arabic to English). A high level of concordance was found between the translations. The constructive feedback that were offered by participants, my research supervisors and colleagues were welcomed and taken into consideration during the whole of the research process to refine research methods, enhance the rigour of the research design and to strengthen the research arguments. Other feedback and discussion from my contribution in conferences were also valuable in strengthening the research arguments (see Appendix 11).

5.1.2 Transferability

Generalisability is still a matter of argument in qualitative research. Hammersley et al. (2000) demonstrated an argument emphasising the importance of identifying the contextual factors that impinge on the case in naturalistic research which could make generalisability impossible. However, Stake (2005) suggests that while a case may be unique, it is also an example of a broader group context. Therefore, the prospect of transferability should not be immediately rejected. Others (e.g. Shenton, 2004) seem to come in between where they mentioned that applying the transformability approach is possible with high attention and consideration of the specific contextual factors.

Transferability can be achieved through different techniques, such as thick description, purposive sampling and maintaining a reflexive journal to support transferability. This study provided sufficient details of the research context to enable readers making a decision whether the prevailing context is similar to another familiar situation and are able to justify whether findings can be applied to the other setting (Denscombe, 2010; Erlandson et al., 1993; Hammersley, 2008). Therefore, thick descriptions that contain details of the interrelationships and intricacies of the study context, where the participants lived and constructed their experience, as well as sufficient detailed descriptions of the techniques of generating data were presented to facilitate judgements to be made about the research findings' transferability. In addition, a detailed description of each step of the research methodological perspective of the research process was described. The research diary of the data collection part can be considered as a reflexive journal that assists in the achievement of transferability. It contained data about the daily agenda of the study, a personal diary of what happened, the methodological decisions made and the reasons for making them, and any speculation about insights researchers collect and a record of any methodological decisions and their rationales.

5.1.3 Dependability

Dependability is concerned with researchers' responsibilities for making certain that the findings of the study can be repeated with the same or similar participants. It can be achieved through keeping a detailed record of the process and decisions about the research (Creswell, 2012; Denscombe, 2010; Yin, 2013). With respect to the subjectivity of interpretivist research, it has been argued that because of the close interrelationship between credibility and dependability, ensuring the former could be expanded in confirming the latter in practice (Hammersley, 2005; Lincoln & Guba, 1985). However, dependability may be achieved using overlapping of data collection methods and sources, as discussed earlier. It could also be addressed through adopting the prototype model of research design in which the processes of the study should be reported in detail. Thereby, that would enable future researchers to repeat the work, if not necessarily to gain the same results. Those techniques go in line with my practice of the use of the thick description, data triangulation and well-established research techniques addressed in earlier parts.

5.1.4 Confirmability

The concept of confirmability refers to the process that should be undertaken to help ensure that the presented findings reflect the ideas and experiences of the participant, rather than the preferences and characteristics of the researcher (Patton, 2005). Confirmability enhances transferability since there is no transferability if there is no credibility available. Therefore, the role of triangulation in promoting such confirmability must be emphasised to reduce the effect of investigator bias in the study. The pilot study was conducted to ensure the participants in the sample would be capable of answering the interview questions without being misunderstood as well as to ensure that the generated data is suitable for answering the research questions.

To this end, researchers should acknowledge and report all their beliefs and perception that influence their decisions about their research strategies than other approaches in their contexts, and explain and admit the weaknesses in these strategies. I used a reflective diary, as explained earlier in detail (section 5.5.4).

5.2 My role as a researcher

The role of the researcher involves integrity, sensitivity, honesty and fairness in terms of data generation and analysis (Silverman, 2010). Miles & Huberman (1994) highlighted that one of the principles for confirmability is the extent to which researchers admit their own predispositions. Adding to my position as a researcher, I am known by all participants, except students, as a former secondary school teacher of English language and a teacher educator and trainer in the study context. Therefore, I have a number of shared experiences with the participants, which define me as an insider researcher. Being an insider researcher can be an advantage due to my knowledge of and experience of the research context. Ryan (2011) stated that insider knowledge provided researchers with a "surplus of seeing" (p. 642) and make them highly informative than of being an outsider. As a researcher, I was fully aware that such knowledge, personal experience and preconceived ideas also had the potential to affect the interpretation of data and introduce bias. Therefore, I was aware of the relevance of such a role to the research since the beginning of this research. Accordingly, different critical scrutiny was employed to minimise this effect, such as undermining power. Mainly, I tried to be flexible, objective, empathetic and persuasive.

For instance, I chose to be open to equalising the relationship and being a good listener. During the interview, I ensured that the participants took enough time to speak openly without being forced to talk about certain points. I showed interest in listening to participants and explained to them how valuable their experience was to me in my research and in my work. Within this role, I adopted an outsider perspective in order to look at things with critical fresh eyes, especially since I have been away from the profession of school education for a while and significant changes have happened in the educational system after the civil war in 2011. Rabe (2003) declared that an outsider researcher would have the ability to investigate the phenomenon and notice things with 'new' eyes than an insider researcher who could consider things as granted. An outsider perspective gave me new insight and understanding, and a fresh perspective of the practitioners' experiences of the phenomenon in the context. However, it was also important to share some of my experience with the participants to show understanding and to build up trust with participants so they were willing to share their experiences as Seidman (2006) advised. Therefore, I had to share some of my teaching experience if they asked me about a similar situation to what the participant had already described in detail.

Moreover, I also ensured to clarify everything clearly in the data even though that is pragmatically and carefully understood. For instance, almost all participants started or developed their conversation and responses using the expression "*as you know*". They tended to rely on my knowledge of the education system and practice by using that expression in their speech. However, I was always asking them to clarify the meaning of that by asking them "what do you mean by that?" or by saying clearly what they meant by that. I was always trying to describe my understanding of the situation clearly to elicit agreement on the mutual knowledge between myself and participants.

In this thesis, therefore, when I am being reflexive, I use the first person "I", but, when I am aiming to assume a more objective position, I use the third person and identify myself as "the researcher".

5.3 Chapter summary

This chapter has discussed the philosophical perspectives underpinning the research paradigm and approach. It has also outlined the research strategy, the study context and design including the sampling techniques, data-generation instruments and data

analysis techniques. The chapter has also described the issues of ethical considerations, the role of the researcher and the measurement of the study's trustworthiness.

This study is an interpretive qualitative case study research. The research forms a case study since it focuses on investigating the washback of the Libyan SECE on perceptions and language teaching and learning strategies and outcomes in a particular context. This case study employed interviews as the main data generation method. One-to-one semi-structured interviews with the teachers, inspectors, the school head teacher and the representative of the Examination Office were conducted. A group interview was used to generate data from 15 participant students. Those students were interviewed in three groups before they undertook the SECE. Only a few of those students were interviewed after they undertook the SECE to reflect on their test performance and their perceptions of their language learning outcomes. All the interview data was generated in the participants' first language (Arabic) to ensure that the interviewees expressed their feelings and perceptions freely without pressure, and to ensure that no misunderstanding happened between myself and the interviewees. The data was transcribed in the same interview language (Arabic) and subsequently translated into English by the researcher.

Classroom observation was also used to collect data about the teachers' and students' classroom practices. The teachers' and inspectors' practices were also observed to examine the nature and the effect of these practices on the classroom teaching and testing practices. Analysing documents related to the research subject was also used to generate data to answer the research questions. A sample of SECE and the teachers' classroom-based tests, supplementary materials in classrooms, teachers' training materials, criteria for assessing students and teachers' evaluation criteria were collected. These documents were used to gain in-depth understanding during the interviews and to analyse the generated data.

Thematic data analysis procedures were employed to analyse the research data through the identification of codes and themes. The processes of organising and familiarising, coding and reducing, and interpreting and representing were discussed in this chapter. The ethical issues were also discussed in the perspectives of acceptance and informed consent, access to the research site, confidentiality, anonymity and safety. Being a language teacher and teacher educator gives me the

ability to take on a dual role as a researcher. The experience gives me insider knowledge; I have outsider knowledge because I am also a researcher. The study's trustworthiness was also discussed according to the criteria of credibility, transferability, dependability and confirmability.

The following chapters next are going to present the findings in relation to the washback effect on teachers and their teaching and assessment practices in the classroom. The chapters will provide interpretation and discussion of the data in relation to the research questions and the theoretical considerations.

6 Chapter Six: participants' perceptions of washback and the SECE

This chapter discusses the key issues related to the participants' perceptions and attitudes towards the relationship between the English language curriculum, assessment and their classroom practice, and how their experience of this relationship yielded the nature and the intensity of washback in the Libyan context. First, it critically examines the participants' perceptions of the English language curriculum, assessment and pedagogy practices. It also presents the effect of the assessment policy on the secondary school students' language learning motivation. It also presents a discussion of ELT teachers', inspectors' and secondary school students' perceptions and attitudes towards the reformed public examination system, and its use for assessing the English language learning outcomes in the Libyan education system. It should be noted that the ELT inspectors, Examination Office representative and the school head teachers are described as "practitioners" in this chapter and the following chapters. In the data narrative, they are anonymised by using P1, P2, P3, P4 and P5 for ethical consideration while teachers are referenced as T1 and T2, and students as S1 and S2, etc.

6.1 Curriculum, assessment and pedagogy policy

Test design and function have a vital influence on the quality and strategies of teaching and learning. Earl (2003) stated that assessment must be aligned with the curriculum objectives and teaching strategies as all these processes aim to evaluate the students' achievements of the objectives. Accordingly, assessment can undermine both student learning motivation and learning outcomes if it is misaligned with learning objectives or instructional strategies. In this study, the findings identified a lack of alignment in the planning of English language education related to the connection between the curriculum elements (including the objective and the time allocated for teaching the curriculum) and public examination policy and focus in Libyan secondary school education. This gap was found to affect the participant Libyan ELT teachers', inspectors' and students' perceptions about different aspects of language education (including curriculum, assessment and practices) as well as their role and educational practices in the study context. These perceptions gave insights into the contemporary practices of how the participant ELT teachers, practitioners and students experienced washback in the Libyan secondary school education.

6.1.1 The curriculum and assessment alignment

The documentary analysis of the SECE test shows that there is a significant gap between the objectives of the English language curriculum and what is measured in the English public examination. Whereas the ELT curriculum focuses on the teaching of English language communicatively to develop students' language skills, the public examination focuses exclusively on assessing grammar and reading comprehension (see Appendix 1 and 2). The participant ELT inspectors showed an awareness of this lack of alignment. For example, in an interview, one of the inspectors criticised this practice as follows:

> *In my point of view, the alignment between the input [the curriculum] and examination method of the outcomes leads to confusion. Here, I think that each of these things goes in its own direction. That means these processes are planned and implemented individually, and the relation between curriculum and examinations is negative* (Interview P4).

The negative relationship described by P4 in the above comment affected his perception and that of the other ELT teachers towards the aim of the English language teaching in Libyan school education. The participant teachers and inspectors constructed the aim of teaching English at secondary school education level according to the public examination focus and away from the pedagogical objectives outlined in the curriculum. Since assessing language skills holistically has never been considered in the Libyan public examination of secondary school education, the participant teachers and inspectors believed that developing English language communicative competence was not the aim of teaching English in Libyan secondary schools. They claimed that the focus of the English public examination requires secondary school students to promote and upgrade their learning of grammar and vocabulary from the prescribed textbooks due to its focus on assessing grammar and reading comprehension.

Consequently, the teachers considered the aim of teaching the English language textbook was to introduce subject-specific information but not to develop English language skills. One of the teachers explained:

> *I think this curriculum has been done to prepare students for the next stage after finishing secondary school by providing scientific information for students*

in English language to know the specialised terms in English but not how to use them because these terms are not for daily use but for studying at the university (Interview T1).

T1's perception about the goal of language education was attributed to the focus of the public examination as well as the content material of the English language curriculum. These perceptions of the participant ELT teachers and practitioners can be understood as an awareness of the traditional focus of assessing English language learning outcomes through assessing grammar and vocabulary despite the reforms in the language curriculum and public examination system that have occurred in the last two decades in the Libyan context. The lack of alignment was also found to affect the participant Libyan students' perceptions about the aim of English language education in the school context. The research data showed that 10 out of 15 participant students (e.g. S1, S3, S9, S12; S2 and S7) had negative perceptions about English language education in the school context. They perceived English language use in a social context to be completely different from learning English as a school subject since the latter did not relate in any meaningful way to the former. For instance, S7 stated:

There are more challenges and difficulties in learning English subject now rather than just learning for developing our English communicative abilities, such as studying certain topics in a certain time and with a very specific focus. At school, they only focus on teaching the curriculum and doing tests and examinations on grammar and reading information [comprehension]. However, if we study English as a language, I would have the opportunity to use it whether I speak correctly or not that doesn't matter. The important thing is that the person I am talking to is able to understand me (Pre-examination group interview 2).

Another student, S12, said:

At school, my focus will depend on more than just learning English. I will think about how many classes we will take and how we will finish the curriculum; and how to get high marks to succeed in the examination. Here, I have to say and do everything correctly to succeed; otherwise, there is no benefit from studying English at school (Pre-examination group interview 2).

In both comments, these students emphasised the significance of studying the curriculum for examination rather than as an opportunity to practise the language as a means of communication within the school context. As Manjarrés (2005) argued, the participants' perceptions and attitudes towards the value of language education in school were resulted from the under-representation of communicative competence, especially when only linguistic competence is included in the public examination. The participant Libyan secondary school students did not value the development of holistic English language skills since the SECE only focused on grammar and vocabulary. They perceived the aim of language education in school to be studying the tested content items of the curriculum. This perception was based on the focus of the examination rather than on classroom practice. That perception de-valued the significance of language education in schools for the participant Libyan students. Alderson & Hamp-Lyons (1996) and Qi (2005) emphasised that the perceived function and interpretation of the test construct compels teachers to work for the immediate goal of raising scores through the focus on linguistic knowledge instead of focusing on language use. In this study, the illustrated lack of alignment between the objectives of the English language curriculum and the focus on the public examination was a major factor influencing the participant Libyan ELT teachers and inspectors, as well as school students' perceptions towards the objectives of English education in general secondary schools.

6.1.2 The power of the public examination policy

Another finding is the impact of the public examination policy on ELT teachers and inspectors. In particular, the study found that the lack of alignment in the planning of English language education influenced the participant ELT inspectors' and teachers' perceptions about their conceptual role within the education system. Whereas the ELT teachers and inspectors were required to teach the curriculum communicatively to the secondary school students to develop their language knowledge and competence (see Appendix 9), the public examination policy only required the teachers to finish teaching textbook content according to the curriculum time plan. In other words, the public examination policy only demanded to teach students the content of the public examination from the curriculum. According to the representative of the Examination Office (P2) and English language inspectors (P1, P3, P4), the main goal of their supervision in the final year of secondary school education was to ensure that the

teachers finished teaching the curriculum before the end of the academic year. For example, the representative of the Examination Office (P2) explained:

The only specified goal in the education policy is to finish the curriculum. For final year teachers, it is not their job to teach what students have missed in the previous academic years such as developing students' spelling and writing skills either in Arabic or the subject of English. By doing that, the teachers would not be able to finish the textbook because of the time constraint. Then, they would be blamed for not doing their job properly as the public examination questions related to the whole textbooks (Interview P2).

P2's narrative can be understood as reflecting the perception that completion of teaching the prescribed textbooks would fulfil the objectives of English teaching in secondary education. P2, as an assessment practitioner, prioritised the importance of finishing the textbook for final year students over the educational goal of building coherent language knowledge for students. This prioritisation demonstrates the power of the public examination policy over the pedagogy outlined in the curricula of Libyan secondary schools. This power is yielded by the significance of students' success in the public examination when evaluating English language teachers' and inspectors' proficiency for their career progression within the Libyan context. The participant ELT teachers and inspectors considered finishing the English curriculum, even superficially for the final year secondary school students, as their main goal in their educational practices rather than the development of students' knowledge and skills. This led the participant teachers to experience a dilemma between pedagogical and ethical decisions as Spratt (2005) described. They either teach according to the curriculum guideline and their own philosophy of language education or teach to the test to enable their students to pass exams, especially since the public examination does not measure the curriculum comprehensively. Hence, following the intended pedagogical practices of a CLT approach integrated into the curriculum it became less important for the participant Libyan English school teachers and inspectors than finishing the curriculum and preparing for the public examination. Even though the Ministry of Education encouraged the Libyan English language school teachers to implement the CLT approach, the participant teachers marginalised the use of CLT in their classrooms. For instance, the Ministry of Education – in association with the British Council – produced a booklet, which was given to English language teachers and

inspectors, explaining the difficulties facing teachers in implementing CLT in a worldwide context, and how to overcome these difficulties (see Appendix 10). The Libyan school ELT teachers and inspectors were supposed to apply some of these techniques to integrate CLT into their classroom practice. However, the participant teachers considered the CLT approach as incompatible with their context. One of the teachers reflecting on the booklet declared:

> *I had a look at the booklet, but I feel like it doesn't work with us. It contains lots of philosophical perspectives, which are a waste of time. In the end, the most important thing for them (Ministry of Education, inspection office, and the school) is to finish the curriculum. If I followed the practices mentioned in that handout, I wouldn't be able to teach a quarter of the curriculum. In that situation, I would be the loser because my main objective is to prepare students to focus on the things that will be included in the final examination* (Interview T2)*.*

T2's statement represents a negative view of CLT in comparison with her main duty as a teacher of final year secondary school according to the public examination policy. This finding implied a conflict in the participant ELT teachers' and inspectors' perceptions of their conceptual role as a language pedagogical tutor and as an assessment policy implementer in the final year of secondary school education in the Libyan context. T2 had chosen to prioritise her role towards the implementation of the public examination policy since it was the main concern in evaluating her proficiency in relation to academic promotions in the context. Although finishing the English curriculum for final year students as an assessment policy was not new in the education system, the intensity of the focus on that became crucial for the practitioners since the implementation of the current public examination system. The head of the English inspection unit explained:

> *There is no change in the teaching focus, but the only thing that teachers have to do now, and should focus on, is finishing the textbook as the questions will be inclusive. That means from each section of the book they should take one or two questions* (Interview P1).

In his comment, P1 emphasised the intensity of the focus on finishing the textbook in relation to the new public examination system compared with the previous system, since it covers a wider range of content in the curriculum.

The Libyan government's assessment policy affected not only the ELT teachers' perceptions of their role but also determined their classroom testing approach. According to the assessment policy, it is important that SECE candidates understand the nature of the examination system and its type of questions in each school subject. It compelled teachers to train final year secondary school students in the public examination. The Libyan school system explicitly tend to illustrate the importance of high-stakes examinations for students and the necessity for them to be trained in answering the examination questions through policy directed towards that as Smith (1996) argued. In this study, the representative of the regional examination department, P2, stated that one of their responsibilities was to ensure that final year students are coached on the final examination approach through the inspection office and the school administration.

> *We emphasise in our supervision of the school teachers through our work with the inspection office on completing the textbooks for students in order to prepare them for the final examination* (Interview P2).

Consequently, the participant teachers ignored the classroom activities and assessment criteria associated with the curriculum and trained their students in a select-response format in the classroom-based tests. In the observed classrooms, the inspectors (P3 and P4) requested a copy of the teachers' classroom and periodical tests to ensure their compatibility with the public examination approach (T1's observation 3, T2's observation 2). The school head teacher (P5) was also keen to ensure the inclusion of questions from previous public examinations in the periodical and classroom tests of the final year students. One of the inspectors, P3, attributed the necessity of training the students in the public examination system and questions because they differ from the type of questions and activities in the textbooks. Thus, it can be argued that the Libyan English language teachers were compelled to teach to the test rather than implement the CLT principles explicit in the curriculum. The public examination policy has significantly affected their practices as their professional promotion was linked to the teaching and assessment practices attached to the public examination policy.

6.1.3 Curriculum and timetable

This study found that the misalignment between the curriculum content and the time plan of English language teaching affected the participant Libyan teachers' and students' perceptions about English teaching and learning in the secondary school context. The participant ELT teachers and inspectors perceived the allocated time for teaching English language within the education policy to be insufficient to complete the curriculum in the designated time. The participant teachers reported that the length of each school period (40 minutes) was not enough to finish a lesson as prescribed in the teaching plan, which then has an effect on the long-term teaching plan where four periods per week became not enough to finish mastering the English language curriculum. The reason for that was identified as the difficulty of the English language textbook content. The participant teachers and inspectors considered the amount of subject-specific information included in the textbook as being beyond the Libyan secondary school students' English language competence. Therefore, this study reinforces the significant effect of the shortage of time allocated to teaching English language textbooks identified by several research studies in the Libyan context (e.g. Asker, 2011; Omar, 2013) on language teaching and learning in schools. This misalignment became problematic in English language education in the Libyan education system, especially in the final year of educational certificates levels. For instance, the participant teachers considered finishing the final year English language curriculum within the time plan for the sake of the SECE as a real challenge. Hence, they prioritised mastering the curriculum through using traditional teaching approaches rather than implementing CLT to rapidly finish teaching before the end of the academic year. S13 described T1 as follows:

> *The teacher is highly regulated by the time this year. Because in the second year, the exam will depend on what the teacher taught. But this year, either we understand or not, the teacher will continue giving lessons to finish the textbook, even if we told him that we didn't understand. Because he knew that he can't cope with the prescribed plan. It is possible that we arrive at the time of the final examinations and we still haven't finished the textbook* (Pre-examination group interview 3).

The above comment indicates the students' awareness of the problem. They seemed to be aware that the teachers' practices were governed by the time constraints rather than their needs. They recognised the teachers' prioritisation and its negative effect on their language learning practices. The students' experiences of this produced a negative response towards language education in school when it is connected to the time pressure of public examinations.

English language teaching is also problematic at a lower level of school education when teachers focus on developing the students' language knowledge and abilities without the pressure of public examinations. However, teachers were also not able to finish the curriculum for students in that situation. The students (e.g. S11) reflected on their previous experiences as follows:

> *Last year [in different school], we were taught by a teacher who always highlighted the important words in real life, and encouraged the whole class to memorise those words. She never passed a lesson unless we all fully understood it; we might spend two to three weeks on one lesson. For that, we studied only six to seven lessons from the textbook, but we all developed in English by the end of the year* (Pre-examination group interview 3)*.

This comment reflects an experience of effective pedagogical practices in the Libyan school context. The teacher provided enough time for the students to be able to understand the curriculum and was able to afford this focus since class teachers controlled the final test questions at lower school levels when students do not undertake public examinations. Although the students were satisfied with their language development, the time was still insufficient to complete the English language textbooks. S14 explained the impact of this as follows:

> *It is also problematic because there is no consistency in learning English at school; we study some English at school each year and, after school finishes at the end of the year, that is it. We don't revise what we studied or missed from the curriculum in the next level* (Pre-examination group interview 3).

From the students' perspectives, there is no development of a broader linguistic knowledge through the English language curriculum. That negatively affected school students' beliefs about the objectives and the value of language education within the

school education system. They treated English as a subject rather than being concerned with acquiring the skills to become proficient in the language. The participant students accepted concentration simply on the English language textbooks and their examinations without any expectation of the sort of language knowledge and skills, which would enable them to use English in a social context. Therefore, it is important to revise the alignment between the intensity of the English curriculum and allocated time for teaching it. This alignment could mediate positive washback on teachers' and students' perceptions and practices towards the focus on building comprehensive language knowledge and developing language skills in the Libyan school education.

The above data suggest that the lack of alignment between different aspects of language education planning had a significant effect on washback and on the perceptions of the ELT teachers, practitioners and students about language education in Libyan secondary schools. The mismatch between the curriculum objectives and public examination focus were the most significant factors eliciting negative perceptions and attitudes towards the school language education.

6.1.4 Policy and students' learning motivations

The literature chapter has discussed several research studies investigating the effect of testing on students' motivation. Most of these studies showed that students' motivation was mostly affected by the function of the examination and its score use rather than its approach or design (e.g. Guilloteaux & Dörnyei, 2008; Hartaty, 2017; Madaus & Clarke, 2001; Zhao, 2016). This research finding reinforced other studies where the standardised format has not been identified of a significant effect on the participant Libyan students' learning motivation. The results could be related to the fact that the public examination focus is still the same in educational perspectives, although the examination approach has been reformed. The participant students' motivation to learn English was connected with the significance of the SECE score in the Libyan education context. Therefore, the students' learning motivation is discussed as a policy related factor in this thesis.

In this study, the participant Libyan school students showed an interest in learning English for the test rather than developing their English language knowledge and ability in their classrooms. Ainley (2006) acknowledged that motivation for learning is

a complex overarching concept, which constitutes a range of psychosocial factors related to learners' personalities, social environments and test-use factors. These factors affect the students' language-learning motivation and directing the nature of test washback on motivation. In this thesis, these factors are categorised into two sections: intrinsic and extrinsic factors.

6.1.4.1 Intrinsic factors

Dörnyei & Ushioda (2013) highlighted that the general perceptions and attitudes towards language learning are tightly linked to motivation. The research data revealed that the participant secondary school students were not intrinsically motivated to develop their English language competence in the school context. This study found that the participant Libyan students were motivated to learn English only for testing because of their negative perceptions and attitudes towards English language learning during their school education (discussed earlier in 6.1.1). Students highlighted this as follows:

> *S5: The English language curriculum is not about the real life. It talks about dinosaurs, ancient human beings and types of fish. I mean these are things that we don't have in our real-life context at all.*
> *S4: If they brought something from our culture and customs, we would be better able to understand it than we do now. But these customs in the textbook aren't ours, and we have difficulties in understanding them* (Pre-examination group interview 1).

S4 and S5 above reflect the students' difficulties in understanding the content of English language school textbooks, which rely on the target language culture and scientific subject-specific information. The students realised that in the subject-specific information, the source of difficulty was not important for future benefit because it is not relevant to the higher level of study. That is because Libyan university education is mostly instructed in Arabic. However, Tsagari (2007) stated that motivational behaviour is shaped over time and is influenced by the wider socio-cultural and contextual environment impinging on language learners. In this study, the use of more western culture in the curriculum emerged as the most significant factor affecting the students' motivation. The participant students classified the curriculum content as unimportant to be learnt for their contextual environment or to develop English

language competence. The participant ELT inspectors also held the same negative perceptions and attitudes towards the effectiveness of secondary school English language textbooks. The head of the English language inspection office (P1) explained:

> *I wish that a new English language textbook could be designed by local ELT practitioners from higher education in association with experienced teachers and inspectors to be compatible with the needs of Libyan students in language and its use in context. Libya is a big country, which has a range of cultural differences. Students could be motivated to learn about others in the country and in doing so they learn how to use English* (Interview P1)*.*

The above comment also reflects the difficulties encountered while using the school textbook content in the social context. The participant ELT inspectors' and school students' desires to change the curriculum reflect the significance of the gap between cultural discourse on the motivation of English language learning. That produced negative washback on motivation in which it encouraged cramming in teaching and learning English at the study context. This also reinforces Bailey & Masuhara's (2012) argument that the intensity of washback through coursebooks is linked to the importance placed on both examinations and coursebooks by administrators' attitudes and needs, as well as the desires of teachers and learners. Thus, the participants' attitudes towards the prescribed textbooks have had a significant role in the washback on the participant secondary school students' motivation towards learning English.

6.1.4.2 Extrinsic factors

This study also identified two extrinsic factors influencing the participant students' language learning motivation: the assessment policy and the score use. Recently, the Ministry of Education extended the opportunity for failed students from the specialised secondary school division, giving them extra chances to undertake the SECE without repeating the school academic year, as they had been required to do previously. This policy adversely influenced the participant Libyan secondary school student motivation towards developing their English language competence in the school context.   An example of this was highlighted by S11:

> *About 50% of students don't care about school subjects because they know that they will succeed in the end even though they didn't study. In the end, they*

will succeed because they will be given chances to undertake a third and fourth examination sitting until they pass the examination (Pre-exam interview 3).

The policy was also highlighted by P2 and other students as negatively influencing students' learning motivations and their perceptions of the level of difficulty of the SECE. The possibility of repeating the public examination more than twice decreased the participant secondary school students' motivation to work hard in their studies, especially those with low levels of cognitive ability. Watanabe (1996) argued that the learners' perceptions of the difficulty of the test cause washback. Therefore, the decision to allow repeat examinations extended negative washback with respect to the participant Libyan students' motivation. It is also possible that such a decision had implications on the standards and quality of the Libyan students' achievements. The participant students held a view, similar to that reported by Hartaty (2017), in which they believed that every student could pass the national examination. Such perceptions negatively affected their motivation to develop their language knowledge and competence. Thus, the participant students did not consider the SECE as an important standard of measurement for the quality of language learning.

In addition, the use of the SECE scores, which are identified as test-related factors by Shih (2007), was another extrinsic factor influencing the participant students' language learning motivation in the Libyan context. It diminished students' interests in learning to develop their language knowledge and ability. Research studies in different contexts identified the use of high-stakes scores within the education system of a significant negative influence on students' motivation (e.g. Al-Yaqoobi et al., 2010; Pan, 2014; Stan, 2012; Tayeb et al., 2014). In this study, the importance of SECE scores motivated the participant secondary school students to focus on test-oriented learning behaviour in order to pass the SECE and achieve a high score. Ten of the fifteen participant students (e.g. S4, S7, S12 and S13) had no interest in developing English language skills in the school context. That is because language skills were considered as insignificant for the SECE score or in the Libyan context in general. They were only motivated to study English in order to succeed in the examination and get a high score (classified as 75% and above). Scott (2007) found that the higher the grades, the more the intensive testing affects students. In particular, two factors related to the educational policy were identified as motivating participant students to value overall high scores in the SECE and the English language subject in particular. These factors

are getting a scholarship and the entry requirements for academic courses in the Libyan university education. First, according to the Libyan education policy, the students with the highest scores in the SECE are awarded a scholarship to study at a university abroad. For example, S7 mentioned that he intended to be diligent in his studies in order to perform well in the examination and achieve the highest score possible in his SECE:

> *Mainly, I focus on passing all examinations with the best score possible. So I am preparing for that from now, and I will be ready for it. It is possible, if I get a good score, that I will get a scholarship to study abroad and be able to get my degree from a developed country* (Pre-examination interview 2).

The fact that internationally qualified graduates tend to have better opportunities for good jobs, especially in academic professions compared to local graduate students, strengthens the value of getting a high overall score in public examinations in the Libyan context. Thus, this influenced the participant students' motivation to learn to the test, especially since this reward is based only on the norm-referenced examination of the SECE in the first instance. Zhao et al. (2015) considered test anxiety and academic stress connected with higher test scores as negative outcomes of high-stakes assessment. Since the scholarships are only given to those achieving the highest scores in the public examination, this washback effect is classified as intensively high in the Libyan context. However, for Pan & Newfields (2012), the instrumental motivation connected with extrinsic factors to earn certificates and to move on to higher or further education cannot last in the long term after the rewards or the target are achieved. Given that, it is possible for such washback to be detrimental to learning motivation and lifelong learning, especially since students are likely to forget the information they have memorised after they pass the test. Therefore, the excessive focus on increasing test scores in the Libyan secondary school education system seemed to hinder the development of multiple language competencies.

Second, the participant students were more motivated to get high scores on their certificates rather than to develop their language ability because it would enable them to fulfil the entry requirements of different academic courses in university education. The SECE score is the only factor that determines the Libyan school-leavers' future in tertiary education, according to the Ministry of Higher Education (MHE, 2010). In the

pre-examination interviews, almost all students expressed an interest in achieving high overall scores of 85% or over. The students' aspiration for high scores was affected by the requirements of their proposed academic courses, as can be seen in the participant students' profile table in the methodology chapter (section 5.4.5). In the interviews, for instance, S4 stated:

> *Yes, it is important, because most of the academic courses at university level require school-leaving students to have a high overall score in the SECE to go to university, otherwise, we can only attend higher vocational institutions. They also require a particularly high score in English language to be accepted on certain courses such as medicine and aviation engineering. So, the entry criteria for various subjects depend on the overall score and English mark on the certificate* (Pre-examination group interview 1).

Another possible explanation for the required high SECE score is that it enables students to join different academic courses in their specialised provision if they decide to change their intentions. The above comment also indicates the social value of academic education over vocational education in the Libyan context. The distinction between academia and higher vocational education also affects students' future employment opportunities in Libyan society. Cheng et al. (2014) highlighted that the nature and intensity of washback on motivation can be affected by different socio-contextual factors related to testing in educational and social contexts. Therefore, students were prompted to study in order to achieve in a way that will satisfy their parents and allow them to attend their university of choice. Zhao (2016) also emphasised the significant role of the intergenerational motivation of students' parents and other socio-contextual factors on students in Asian contexts where a test-driven education system is paramount. This study suggests that the participant Libyan students conform to this stereotype. It was also found that because there are a few university courses that are taught in English, some students were motivated to develop their knowledge of English in an ESP context. A notable example of this was students who were planning to study medicine or engineering (e.g. S1, S4, S6 and S12). They highlighted their enthusiasm to achieve high scores in the examination because an appropriate level of English is one of the entry requirements to study these specialisations.

This study's findings suggest that the participant secondary school students' motivation to learn English to the test has been shaped by their perceptions and attitudes towards the English language curriculum and the nature and score use of the SECE in the Libyan education system. Since the test-takers' motivation may affect test preparation and performance (Weir, 2005), the participant school students adopted "cramming" in the English language classrooms in order to ensure success in public examinations rather than developing their English language competence (see more discussion in section 7.2).

6.2 The SECE assessment in practice

Despite the participants' negative perceptions of the public examination focus and its effect on developing English language skills, they had a positive attitude towards the use of a select-response format in the SECE. This section addresses the participants' perceptions and attitudes towards the standardised format of the public examination. It also addresses the participant ELT teachers' and inspectors' perceptions of the effectiveness of using this standardised format in public examinations within the Libyan secondary school context.

6.2.1 Perceptions and attitudes towards the SECE approach

All participants of this study had a positive attitude towards the use of a select-response examination and the automated marking system in the SECE. Wall (2000) declared that perceptions about adopting an innovation are determined by many factors, including the characteristics of the innovation (such as the assessment approach in the study context), the context (classroom, school, the educational and cultural system), models for introducing change (top-down or bottom-up policy implementation) and the potential users. In this study, the positive attitudes towards the reformed public examination system were a result of the perceived improvement in public examination administration, and the candidates' scores, as well as the effect of the top-down policy implementation approach used in the Libyan context.

6.2.1.1 Public examination administration

The participant ELT teachers and other practitioners revealed the reformed examination system had improved the administration of the public examination in the Libyan education system. The teachers revealed the current examination system

increased the efficiency of the marking process and publication of candidates' results. T2 explained:

> *Integrating technology into marking public examinations makes it more accurate and objective. That helped the Ministry of Education to avoid the imperfections of the previous method, such as shortage of marking time, speeding publication of the students' results this means that their certificates will be issued early so that they can register for the universities including those students who undertook the examination re-sits* (Interview T2).

Another teacher (T1) said:

> *In fact, the reformed public assessment system was implemented for different technical reasons, such as the speed and accuracy of marking the examination for those who engaged in this process before* (Interview T1).

The teachers' emphasis on marking, in the above interview extracts, reflects their awareness of the issues attached to grading the public examination. The teachers reacted positively to the electronic marking process because they perceived it as objective, and it diminished their responsibility for the marking criteria of the public examination. In the previous examination system, the students' papers were marked manually by different examiners, and although standardised, this required academic judgement when applying the marking criteria as each individual considered the students' scripts. Thus, the perceived fairness and efficiency of the new system influenced teachers' perceptions of the public examination. The automated scoring was found to have a positive impact on the quality of marking (Bejar, 2011). In different washback research, factors pertinent to the exam itself were found, determining the existence, quality and degree of washback, such as the level of stakes and the status of the language test, and the purposes and uses of the exam scores (Shohamy et al., 1996; Spratt, 2005; Wall, 2012). However, the issues of objectivity and the associated work and responsibility have not been discussed as a significant factor in washback research. Only Taqizadeh & Birjandi (2015) highlighted a similar finding to the current study where some of their participant teachers referred to the fairness of the Iranian University Entrance Examination (due to the scoring technique), clarity of accurate answers and easy scoring as factors eliciting positive attitudes towards the test. This

implied that the issue of objectivity could be a matter of concern in the MENA context or other developing countries.

Additionally, the participant teachers and other practitioners (e.g. T1, P1, P2 and P3) also declared that the new public examination system eliminated cheating during the examination as the order of examination questions for each candidate is different from others in the same sitting. One of the teachers (T1) emphasised:

> *I think this public examination system is better than the previous approach as it helped in eliminating cheating. The new examination technique grants candidates the chance to be independent as examination papers are different from one candidate to another* (Interview T1).

In the previous examination approach, candidate-to-candidate cheating could easily occur because they were tested with similar examination papers and sat the examination close to each other in the same sites. In the new examination system, candidates are less likely to cheat since they cannot look over at each other's responses as the examination papers are different from one to another. Fulcher (2011) emphasised that the test scores integrity could be ensured by diminishing candidates chances for gain marks by fraudulent means, which threatens test validity. The participant teachers and practitioners had been concerned about the reliability of the examination, and eliminating cheating in that way reassured them that the public examination scores would likely more accurately be representative of individual candidate's ability. This technique used to eliminate cheating also decreased the pressure on teachers in securing the public examination when supervising examination sittings. However, in the Libyan context, the phenomenon of cheating within public examinations is complex and multi-faceted. These complex issues will be discussed in more detail in section 8.1.2.3.

6.2.1.2 Grading in public examinations

The use of a select-response format was perceived as having a positive impact on candidates' scores improvement in the Libyan public examination. This improvement yielded positive attitudes towards the new public examination system among all participants in this study. The select-response approach was perceived as diminishing the possibility of losing marks because of spelling mistakes or lack of clarity when writing as in the previous examination system, which was based on a constructed-

response approach with open-ended, wh-questions. For example, one of the teachers stated:

> *This examination approach is better for students because it helps them to get high marks and not lose marks because of bad handwriting. The students get the benefit of being stress-free and avoid the other psychological pressure that students used to experience because the student's goal is to succeed, not to develop their English language skills* (Interview T2).

In another interview, T1 said that the new public examination system offered students options, which simplified the required cognitive process needed by students when undertaking the public examination:

> *The new public examination approach has shortened what students need to remember by giving them choices and not leaving them wondering how to answer the public examinations' questions* (Interview T1).

The above comment reveals the participant teachers' satisfaction with the select-response approach in enhancing the performance of students with lower language ability, especially since the primary goal of the teachers and students of the final year was only to pass the public examination, not the development of language knowledge and skills. This reinforces the described finding that standardised examination has a significant effect on improving test scores in different contexts (Hsieh, 2017: Taiwan; Toksöz & Ertunç, 2017: Turkey; Yu & Jie, 2013: China). The ease of the cognitive process emphasised in this research refers to the learners' abilities to apply test-taking strategies described by those research studies, which develop their performance in the test. This finding showed different results from Onaiba (2014) regarding attitudes and perceptions about the public examination approach. In this study, the participant teachers' positive attitudes towards the public examination system were linked to the improvement of candidates' scores, and the accuracy and perceived objectivity of the public examination. Onaiba's study was conducted close to the time of examination when his participants did not recognise these improvements. This difference reflects the changing nature of washback over time.

The participant students also perceived the select-response questions facilitated higher scores in public examinations. This positive perception was repeatedly highlighted in the students' pre-examination group interviews and the post-

examination group interview. In the pre-examination group interviews, students (S2, S4, S8, S9, S10, S11 and S14) declared that the focus on written tests was necessary to increase their coursework marks towards the final score of the SECE. This was explained by S8, S9 and S10 as follows:

> *S9: The focus on multiple-choice questions in classroom tests helps us to increase our marks better than other types of examination or the evaluation criteria of the coursework, especially since we do not do a lot in the classroom. We only follow the teacher while she is reading and explaining texts and grammar from the coursebook.*
>
> *S8: Yes, because in this type of examination we can study for the test and perform well to increase our marks. This is better than an examination where we need to write, or when evaluating our reading or speaking in the classroom participation because we are not good in these skills.*
>
> *S10: Multiple-choice questions are valuable for us, as it would enable us high grades or distinction* (Pre-examination group interview 2).

In the post-examination group interview, S15 explained:

> *The type of public examination question was easy and good. It helped me to succeed in examinations better than the type of question in the textbooks. The questions summarised all the information we have studied during the year. It helped me to remember the information and be able to choose the right answer to get a good score* (Post-examination group interview).

The above comments suggest that the participant final year secondary school teachers and students preferred the use of the public examination approach mainly to increase their marks towards the SECE. That was because of the value of the test score in the Libyan educational and social context over the students' language development and the ability to use English. However, such focus on the multiple-choice approach had an adverse impact, not only on the value of the coursework activities but also on the participant teachers' and students' perceptions towards developing English language skills in the Libyan school context. Therefore, this examination approach yielded undesirable and ineffective assessment practice in language teaching and learning in the study context.

6.2.1.3 Policy implementation in practice

The participant ELT teachers' and other practitioners' positive attitudes towards the reformed SECE system were also affected by the traditional top-down policy implementation approach used in the Libyan education system. Gunn et al. (2016) highlighted that education policy is often used or instructed teachers to administer a certain type of tests with little or no chance to give feedback regarding whether or not they feel that these tests are useful, valid or should even be given. During the interviews, the practitioners (P1 P2, P3 and P4) claimed that the reformed public examination was intended to integrate modern educational innovation, such as automated scoring in the examination, in the Libyan education system. The participant teachers and practitioners (e.g. T1, T2, P1, P2 and P3) recited the exact objectives published by the Ministry of Education in the policy documents as their own perception of the reasons behind the incorporation of a select-response approach in public examinations. The reiteration of the Ministry of Education claims regarding the reformed public examination system reflects the power of the Ministry on policy implementation, not only on practitioners' practices but also on their perceptions about educational practices. For instance, the representative of the Examination Office stated:

> *This approach was used as a way of simplifying the questions of the public exams to students, but we don't know if it actually does that. Nobody, however, can say that the specialised people who advised this method are wrong or that wasn't a good choice as this is their point of view in doing this* (Interview P2).

The above statement by P2 reflects the effect of top-down policy implementation, since those practitioners have not been given a chance to express their own opinions or give feedback regarding government innovation in the educational context. What is also clear is P2's hesitation to challenge the argument presented by those perceived to be at a higher level in the hierarchy of the education system. Moreover, the recognition of the enhanced process and improved score also diminished the participant English language teachers' and inspectors' willingness to express criticism about these changes. The satisfaction of the Libyan Ministry of Education with the increased number of school-leavers passing the SECE from the first sitting was considered as an indication of the positive effect of the use of the new system. Gebril & Brown (2014) argued that the public examination scores tended to be linked to the quality of schools

and educational improvement within an examination-driven society such as MENA context and many Asian countries, where teachers considers a high quality school teaching is judged by student exam results. However, these finding is not in line with other research studies in other context (e.g. Menken, 2008; Turner, 2006) which argued that teachers have a positive attitude and motivation towards integrating changes into their teaching or assessment practice according to their own beliefs and professional stances on high-stakes examination. The study indicated that the power of the assessment policy and the policy implementation approach had a significant effect on the participants' perceptions and attitudes towards the high-stakes examination system.

6.2.2 Evaluation of the SECE approach

The research data revealed a variance and contradiction in the perceptions of the participant ELT teachers and inspectors in the efficiency of using the select-response format in assessing English teaching and learning outcomes in the Libyan school context. This discrepancy and inconsistency were caused by the lack of alignment between the focus of the curriculum and public examination, and the participant ELT teachers' and practitioners' beliefs regarding language learning and acquisition.

The participant ELT teachers believed that the select-response approach used in public examinations was suitable and adequate to assess students' knowledge of the ELT curriculum. Spratt (2005) argued that the teachers' goals and positive attitudes towards the approach and design of an examination determine the extent of washback in the micro context. Accordingly, the participant teachers' perceptions were linked to their beliefs that the development of students' language skills was not the aim of teaching English in secondary school education. For example, one of the teachers stated that:

> *The public examination approach is appropriate to assess students' knowledge of linguistic aspects and the information included in the ELT curriculum* (Interview T1)*.*

The teachers' perceptions relied on the traditional focus of the public examination of the English language subject, assessing grammar and comprehension, as language skills development was not valued or rewarded (see 6.1.1). Based on their experience of preparing students for public examinations, they believed that the current public

examination system presented the traditional focus of English language teaching and assessment within the Libyan school context. Accordingly, the participant teacher experience plays a significant role in their beliefs, perspectives and decisions about their classroom teaching and assessment practices. The participant teachers' translations of the goals of the ELT curriculum into test specifications yielded undesirable perceptions of the link between the curriculum and testing. Thus, they considered the current examination system as more effective practice than ever before in the Libyan education system. Similarly, the students considered the select-response approach as compatible with the English teaching and learning focus in the school context. S1 expressed his view about that as follows:

> *The public examination is appropriate to assess what we learn in school. All of what we study in the classroom comes in the public examination. That is what the education policy requires. We should study the whole curriculum and undertake the examinations to make sure that we finished the curriculum* (Pre-examination group interview 1).

This comment is illustrative of the fact that the students' perceptions of assessment are closely related to the classroom teaching focus rather than to the objectives of the English curriculum. This account also suggests that the public examination policy is considered as the cornerstone of the Libyan educational system. The focus and practice of the examination were perceived as the main objective of teaching and learning English in schools. These findings support the argument of Barnes (2017) that the perceived aim of language teaching context (e.g., test preparation versus general English in this context) affects the teachers' beliefs about the effective teaching practices in their classes.

However, the participant ELT inspectors considered the public examination approach as unsuitable for assessing the students' development in the prescribed language knowledge and skills although they shared the teachers' beliefs that developing language skills were not important for secondary school students. The head of the English inspection unit stated that:

> *I think the examination should assess the student ability in using English in communication and writing at least at a minimum level to encourage students to learn how to use the language in order to achieve the curriculum objectives.*

> *I mean we should make sure that school students pass at least the beginner level of English communicative competence* (Interview P1).

P1's statement reflects his preference to align the public examination with the focus of the curriculum in order to promote beneficial washback on language teaching and learning practices. This perspective can be attributed to the ELT inspectors' understanding of their role in supervising ELT teaching in order to implement the curriculum to achieve its intended outcomes. In this regard, the representative of the Examination Office (P2) stated that the inspectors are responsible for the quality of language education, especially the non-tested aspects. Hence, P1 considered the inclusion of language skills in the SECE examination would enhance the teachers' and students' focus in their classroom on the curriculum objectives. One of the teachers (T1) also responded to a question about whether the inclusion of assessing language skills would affect the development of students' language skills in the school context as follows:

> *Sure it would! Because students are studying for examinations, and its inclusion in the examination, sure would lead teachers and students to focus on these skills so that students succeed* (Interview T1).

The above statement emphasises the impact of the public examination on the English teaching and learning focus and motivation in the study context. This desire for change can be interpreted as an indication of the adverse washback of the current SECE approach on the goals of teaching and learning English in Libyan secondary school education. In addition, the participant ELT inspectors argued that the select-response approach was inadequate for assessing different cognitive abilities necessary for successful language learning and acquisition. For instance, one of the inspectors (P3) indicated the following at the end of the interview:

> *I just want to insist on the fact that, even though the previous method of examination was difficult, it was more effective for assessing and evaluating students' actual learning outcomes. The new examination system has the advantages of accuracy in marking, specificity and speed in getting the results, however, it doesn't assess the students' language ability and learning outcomes* (Interview P3).

In the above statement, P3 seems to refer to the students' writing skills as an important aspect of language learning that is missing from the public examination. The participant teachers also thought that the public examination did not promote effective language learning practices through not assessing writing. One of the teachers (T1) recommended the following:

> *This examination should focus more on reading comprehension through presenting a passage in the examination and giving the choices accordingly, it should not rely only on the memorisation of the information from the textbook. Adding a question about writing would encourage students to remember a reasonable number of vocabulary only for writing* (Interview T1).

This implies some contradiction in the participant ELT teachers' attitudes towards the SECE. The teachers seemed to consider the public examination approach as effective according to the language education focus in the school context. However, they believed that the SECE did not encourage effective language learning practice. This belief was based on their understanding of learning a second/foreign language. Other participant practitioners in secondary education, such as the school head teachers (P5) and the representative of the Examination Office (P2), also held the same view regarding assessing general education outcomes. For example, the representative of the Examination Office also stated:

> *In this examination system, the students write nothing at all, even their names are already written on the examination paper, so with that, we can see that students neglect the skills of handwriting in the first instance and the development of writing strategies in general. For that, essay writing questions are important to see how students write and express their opinion* (Interview P2).

The above comment reflects the traditional view regarding the value of written composition in the Libyan education system. This perception of the value of assessing learning outcomes through constructed response examination was one of the most significant factors influencing the participant teachers' and practitioners' views of the SECE approach. The comments by P3, T1 and P2 above indicate the failure to assess students' writing ability and skills was seen as ignoring an important cognitive ability, which promotes development in language learning and acquisition. The participant

ELT teachers and other practitioners held the perceptions that the SECE was not suitable for assessing students' language learning, as it does not provide the opportunity for students to express their language learning outcomes through written composition.

In summary, the lack of alignment between the examination focus and the objectives of the curriculum has had a negative impact on the participant ELT teachers' and inspectors' perceptions about English language education. They valued the implementation of the public examination policy and ignored their role to use the pedagogical strategy of English education in secondary schools. Nevertheless, Shohamy (2013) admitted that the quality of tests should be judged not only by how well tests measure language knowledge from a measurement perspective, but should also take into consideration the reasons for introducing these tests, and the consequences of tests on individuals, or the whole education system and social groups. Hence, the importance of perceived positive washback of the reformed SECE system on scoring the public examination and the improvement of candidates' scores within the Libyan education system should not be under-valued, especially since these positive effects were reported as the main motive for implementing the new public examination system in the Libyan education context. Therefore, the findings suggest that the SECE has multi-directional washback on different individuals in the study context; however, it has achieved the intended washback as determined by stakeholders.

6.3   Chapter summary

This chapter has discussed the participant ELT teachers' and inspectors' perceptions of washback and the effect of the SECE on perceptions about language education in the school system. It has focused on the nature of the relationship between the educational policy regarding the objectives of the English language curriculum and the associated time plan in connection with the SECE policy. Throughout this chapter, I have argued that the lack of alignment between the focus of the English SECE and the objectives of the curriculum had a significant effect on the participant ELT teachers', inspectors' and students' perceptions of the goal of language education in the Libyan school context and their view of the recent public examination approach. In particular, the participant English language teachers and inspectors, as well as

students, held the perception that developing language skills is not the aim of teaching English in the secondary school context since these skills have never been assessed in the Libyan SECE despite the integration of a communicative approach. Therefore, they identified the goal of language education at secondary school based on the traditional focus of the public examination of English language and claimed that it should promote the school students' knowledge of vocabulary and grammar. Accordingly, the participant ELT teachers and school students held positive attitudes towards the use of the select-response approach in the Libyan public examination since it is compatible with the traditional focus of language teaching and assessment in the school education system. However, the ELT inspectors believed that such an assessment approach neglected to assess writing composition, which they considered as essential for students' cognitive development and for successful language learning and acquisition. From this standpoint, these inspectors held a negative attitude towards the reformed public examination approach.

I have also argued that the lack of alignment in the planning of English language education influenced the participant ELT inspectors' and teachers' perceptions of their role within the policy education system. The public examination policy had a powerful impact on the evaluation of the participant teachers' and inspectors' professional careers, so their language teaching and assessment practices were conducted to fulfil this rather than to meet the pedagogical objectives of the English school curriculum. For instance, the teachers and students prioritised the aim of finishing the curriculum even through using traditional teaching approaches rather than focusing on building comprehensive English language knowledge or implementing the CLT proposed by the curriculum. This perception was also influenced by the lack of alignment between the scope of English language teaching content and the time plan allocated for teaching and learning. This mismatch had a significant effect on the participant ELT practitioners' and school students' perceptions about the aim and value of English teaching and learning in school education.

The participant students' experiences of this lack of alignment and the under-representation of English language use and language skills also produced negative perceptions among them about the value of English language education in a school context, which subsequently affected their language-learning motivation. The participant students were motivated to learn only for the test. The educational rewards

and the use of the SECE in the higher education system endorsed the participant secondary school students' emphasis on measurement-driven learning practices since both focus only on examination results, not on language competence. It is also important to highlight the emerging theme related to the perceived gap between the western socio-cultural context presented in English language textbooks and the local characteristic of the Libyan society was an important factor driving students' motivations towards the measurement-driven learning. This finding reinforces the importance of considering washback by textbook highlighted by Bailey & Masuhara (2012).

The next chapter will discuss how the participant ELT practitioners and students experience this misalignment and how it has affected the nature and intensity of washback in the study context.

7 Chapter Seven: washback on classroom experiences

This chapter critically examines how the participant Libyan ELT teachers and their students experienced the implementation of pedagogical and assessment policy in classroom practices. In particular, it examines how the participants' perceptions discussed in the previous chapter about the SECE and language education affected and shaped washback on their classroom practices. The chapter is divided into two main sections: teaching and assessment strategies, and learning strategies and outcomes.

7.1 Teaching and assessment strategies

The key finding of this chapter reveals that the policy and practice of the SECE crucially affected the participant teachers' classroom practices. The following section discusses SECE washback on the English language secondary school classroom teaching, activities and assessment in detail.

7.1.1 Classroom teaching approaches

Findings from the classroom observation show that participant teachers used different teaching practices from those prescribed in the ELT textbooks and outlined in the teachers' book (T1 and T2's classroom observations 1, 2 and 3). Although the prescribed ELT curriculum imposes the use of the CLT approach (see Appendix 1), teachers justified their neglect of CLT as being ineffective in fulfilling their duty as teachers of final year secondary school students according to the high-stakes examination policy; they used a teacher-centred instructional approach and grammar translation method (GTM). These approaches have been widely considered as the traditional English language teaching approach in the Libyan language teaching classrooms (Abusrewel, 2014; Orafi, 2013). However, this study showed that the participant teachers' perceptions and practice of these teaching approaches represent the washback of the public examination rather than being related to institutional and personal factors as highlighted in the majority of research studies in the Libyan context. Findings from this study also show that the lack of alignment in the planning of English language education (discussed in section 6.1) affected the participant teachers' choice of teaching approach in Libyan secondary schools as follows.

7.1.1.1 Teacher-centred instructional approach

The participant teachers relied on a teacher-centred instructional approach for teaching English language curriculum content in their language classrooms. Due to the importance of finishing the curriculum and the shortage of allocated teaching time (discussed in 6.1.2 and 6.1.3), the teachers felt compelled to use a teacher-centred instructional approach in order to be able to finish the curriculum. It is clear from the classroom observation that the teachers dominated the classroom talk to a large extent. The following interview extract from S11 and S13's conversation illustrates the significant influence of time pressure on T2 and students and its effect on classroom management.

> *S13: The teacher is highly regulated by the time, because this year [the final year] either we understand or not the teacher will continue giving lessons to finish the textbook, even if we told her that we didn't understand. Because she knows that she can't cope with the prescribed plan.*
> *S11: Yes, that is why she gives several lessons and sometimes she finishes one or two units in one period.*
> *S13: We don't understand because of the number of periods we take every day. So we only listen, we don't understand much (*Pre-examination group interview 3).

S9 also added about T2's practice:

> *All the time is fully occupied by minutes and seconds for her, she just follows her plan and continues reading from the textbook. She doesn't waste a single minute of the period. Because she only wants to finish the lesson, write the grammatical rule and give the homework before the period finishes* (Pre-examination group interview 2).

S14 described T1's classroom as follows:

> *There is no time for us to speak; it is the teacher who keeps the class busy with explaining grammar and reading from the book all the time* (Pre-examination group interview 3).

These examples demonstrate the participant teachers' intensive use of the teacher-centred instructional approach in their classrooms and their need to devote time for extra classes to finish the curriculum. The students in the above narrative indicated

that the described practices were different from teachers' classroom practices the previous year. Even though the teachers seemed to use the same teaching approach, they focused more on teaching the curriculum content right until the end of the period in order to be able to cover all the content of the public examination in the final year. This indication represents the existence of washback in the research context. Thus, these findings suggest that the policy requirement for the teachers to finish the curriculum within a restricted allocated time is a significant factor that has an impact on the processing of task material, and influencing the participant teachers' considerable focus on using the teacher-centred approach in their classrooms. This finding reinforces the findings of Pathan et al. (2016) who highlighted in a different study that the Libyan inspectors demand that teachers complete the curriculum in time without being concerned with authentic language learning. Therefore, the tight time allocation adversely affected pedagogical teaching practices in a way that can be considered as a type of negative washback similar to that observed in other Asian contexts by Qi (2005). In this study, the use of the teacher-centred approach undermined opportunities for extended classroom discussion and activities. Hence, the English language classrooms were transformed into a lecture environment. The identification of reading and explaining the textbook content by teachers as the only way to finish the curriculum also implies the reinforcement of the traditional role of the teachers as a transmitter of the knowledge from the curriculum by the public examination policy. Hence, applying CLT that promotes a student-centred learning approach became meaningless in the study context.

7.1.1.2 Grammar translation method (GTM)

This study also found that the participant teachers used the grammar-translation method in their classroom instructions. They knew that the public examination focus was still on assessing grammar and vocabulary even though the curriculum and examination system had been reformed. Therefore, they continued using their traditional teaching approach of GTM to teach the grammar and vocabulary required for the SECE. In the interviews with teachers, one of the teachers (T2) replied to a question about whether the reformed public examination approach had changed her teaching focus in teaching as follows:

> *It is grammar and grammar, then grammar. I also focus on teaching the information about famous characters from history and science; geographical and statistics information from the textbooks. That is because there are always questions about facts such as dates of birth and death information; and scientific achievements in the public examinations* (Interview T2).

In the above statement, T2's repetition of the word "grammar" can be interpreted as a representation of the intensity of her usual focus on teaching grammar in classrooms for the public examination. Chen & Tsai (2012) and Pizarro (2010) highlighted that language examinations tend to reduce the teachers' freedom to choose a teaching methodology, especially one that is incompatible with the format of the test. In the study context, the teachers' teaching methodology was based on the traditional focus on grammar and vocabulary even after reforming the public examination system. T2 continued:

> *As a final year teacher, my experience of teaching the final year students and preparing them for examination, I have full knowledge of what is important for the final year examination and how to teach it for students. So, we should focus on that during students' preparation for the public examination* (Interview T2).

In this narrative, T2 emphasised the role of her experience in preparing students for the public examination as a justification for her choice of the classroom teaching approach. That experience provided T2 with knowledge of the tested aspects of the curriculum in the public examination and that influenced her perception of the effectiveness of the GTM to prepare students for the SECE. The success of former students in the SECE was considered by the teachers as evidence of the effectiveness of using GTM as opposed to a more integrative communicative approach in their classroom practice. Nevertheless, the new public examination format negatively affected the development of the students' vocabulary for writing. In spite of the teachers' beliefs that assessing writing is important for promoting better language learning outcomes, they ignored it in their teaching. Three students (S3, S8 and S11) highlighted that they had not been engaged in any writing activities in their classrooms. S3 stated:

> *There is no revision of what we studied so far in this year. We don't do reading practice, writing or spelling. She said that she would produce a test to give us*

the marks given for these activities in the assessment criteria, as there are only three marks on reading practice. However, she hasn't done it (Pre-examination group interview 1).

S3's comment emphasised the complete alignment of the teachers' practices to the SECE focus in the final year of secondary school. Writing ability is ignored because it is not part of the public examination. This finding seems incompatible with Cheng & Watanabe (2004) and Watanabe (1996) who argued that the effect of the test content and use is insignificant in determining the teaching methodology a teacher employs; rather, teacher factors, personal beliefs, past education and academic background are more powerful. This study shows that the use of GTM was affected by the SECE content in the study school context. The participant language teachers had neglected the development of their students' spelling and writing skills since the implementation of the select-response items. The change was not only avoiding teaching broader language skills but it also included all the classroom activities as explained below.

7.1.2 Classroom activities

The findings of the study show that the classroom activities were also geared in the direction of the public examination. In other words, *teaching to the test* was the focus of the participant teachers' overall classroom teaching practices. The participant teachers' classroom activities were broadly identified as test-preparation activities. The students revealed that the classroom activities in the final year were different from the earlier years of schooling. For instance, the following extract from the group interviews represents a description of the students' experiences with T1 classroom activities:

> *S5: Yes, it is only this year. In our first year of the secondary school, the teacher used to ask us to read in the class [reading aloud practice] and we used to study new words to remember them, but now we do not read and study new vocabulary.*
> *S2: Even phonetics we used to study them but now we don't study that and don't do the reading practice. We just study grammar and passages* (Pre-examination group interview 1).

The classroom activities mentioned above mirror the participant teachers' perceptions (as discussed in section 6.2.2) of language education. These practices were reported

as the normal teaching practices by the students at different educational levels including when they did not undertake a public examination. However, as these language activities were not tested in the public examination, the participant teachers did not teach spelling and composition in the final year of secondary school. Aftab et al. (2014) and Shih (2007) indicated a negative impact of preparing students for public examination on developing oral communication and other skills since most of the class time is devoted to the teaching of skills featured in the test. In the study context, another explanation that could be given for this ignorance of such activities is the limited time available to teach the whole curriculum. Students declared on different occasions during the group interviews that they often did not finish studying the textbooks in their earlier years of study at secondary school.

From the textbooks, the participant teachers also devoted their classroom activities to drilling grammar and vocabulary since this represented the focus of the SECE and ignored the other activities in the English language textbooks. For instance, S7 described T1's classroom as follows:

> *Although the teacher did not do as other teachers, highlighting the different versions of possible questions that might come in the public examination in each lesson, it is good that he highlights the important parts that usually come in the public examination* (Pre-examination group interview 2).

It is obvious from S7's description that such classroom activities were not limited to English language classes but were also used in other school subjects. Thus, such activities would be considered conventional in the final year and this no doubt reinforced the fact that school teachers had to follow that approach. In a question about whether the SECE affected her classroom, T2 stated:

> *Definitely, that happened, I became more aware of the important things in each lesson; so I pass by some activities and don't give them a lot of time. I just read them aloud and work together to translate them; so that students share the same idea and are confident that those activities are not important* (Interview T2).

Here, T2 is revealing that she seeks students' agreement that her selection of textbook tasks is appropriate. It is interesting to note that the students' satisfaction with that strategy was important for the teachers in such a washback effect. Those classroom

activities were favoured over the actual objectives of English education, not only by the participant teachers and students but also by ELT inspectors in the educational context. During the observation of the inspector's visits for T1, he did not give any comments regarding the type of classroom activities although the teachers' performance evaluation certificate is concerned with the existence of communicative activities as shown in Appendix 9 (T1's classroom observation 2). Most of the discussion was about the importance of following the teaching plan and assessment criteria for assessing final year secondary school students according to the government policy. This finding further reinforces the argument by Cheng et al. (2015) that traditional test-preparation activities are favoured internationally to prepare students for high-stakes examinations. Distinctively, in the Libyan context, the participant teachers applied these activities during the whole academic year, not just near the time of the examinations as reported in different contexts. These practices may reflect the test's value or highlight the power of the SECE in the Libyan educational context. Over 10 students (including S2, S3, S6, S10, S14 and S15) considered the classroom activities as effective for their preparation for the SECE and therefore gained considerable approval from the students.

The observations data also showed the students' most frequent questions were about the meanings of words and the ways in which questions might be formed in certain linguistic aspects. The participant teachers were observed directly giving the Arabic equivalents of the word within the reading texts of the textbooks. After finishing the curriculum, the teachers (e.g. T1) devoted classroom time to an explicit drilling of question-solving activities using the select-response questions from the previous public examinations. This result is congruent with Aftab et al. (2014) and Taqizadeh & Birjandi (2015) as the teachers made use of test format imitation and test-related material on the assumption that it would bring students success. The participant teachers and students agreed that extensive test-preparation activities should be carried out after finishing the English textbooks. Illustrating this, S5 stated:

> *At the end of the second semester, we were completely revising the questions of the previous examination with the teacher in the classroom, and he explained to us the techniques we should use in the exam so that we don't waste too much time in the exam* (Post-examination group interview).

The above quotation reveals that the participant Libyan school teachers also relied on developing their students' test-taking strategies. In the observations, T2 showed an example of the use of such test-preparation activities. She read questions from a previous English language exam on the SECE and its choices for students and then asked them to identify the correct answer. The students replied by reading their chosen options, either in the form of sentences, phrases in English or single letters such as a, b, c or d. On different occasions, the teachers asked students to give the reasons for their chosen answer, and students tended to give their reasons in Arabic. When wrong answers were given, the teachers answered the question directly and explained the meaning of the sentences in Arabic. In those activities, the students did not answer the questions individually, but the classroom interactions were mostly between the teacher and the students as a whole class. In the post-examination interview, some students explained that those activities enabled them to become familiar with the examination questions as they used previous versions of the English test. This finding corroborates the ideas of Barnes (2017) who suggested that a good test-preparation course may not actually reflect best practice in language teaching or espouse a communicative language approach, but it may ensure that students receive passing scores.

Overall, the fact that classroom content and activities were entirely geared in the direction of the SECE can be considered as undesirable in the sense that the participant teachers were engaged in teaching and practising the narrow aspects tested in the SECE rather than developing the students' language skills as proposed in the objectives of teaching the curriculum. Within the socio-cognitive framework, such a teaching approach and focus was identified as an attempt to provide students with the full executive resources of the SECE by focusing on the language knowledge (grammatical and lexical), which is considered to be the SECE construct and demands in the context validity, contrary to the curriculum designers' intentions.

7.1.3 Classroom assessment

Research data analysis indicates that the participant secondary school teachers tend to neglect not only the teaching practices prescribed in the curriculum but also the assessment criteria. This finding supports the conclusion of Safa & Jafari (2016) that the high-stakes examinations not only generate *teaching to the test* in the context but

also *testing to the test*. The teachers fully adopted the public examination format, content and techniques in the classroom-based testing. They declared that the main reason for that was to train students in what they would experience in the public examination. For example, one of the teachers mentioned:

> *I apply the same method as final exams in the classroom tests and the first and second school periodical tests so that students get used to such type of examination and do not get frightened of it in the final public exam* (Interview T1).

The other teacher also justified her practices as follows:

> *By doing that, the students would practise the public examination questions in a simple way since every time I include different parts of the curriculum in classroom tests and school periodical tests until I test them on the entire curriculum. Then students will have experienced the public examination questions that addressed the whole curriculum* (Interview T2).

The teachers' comments emphasise their commitment to implementing the assessment policy since it was an important aspect of their professional promotion and evaluation. This finding emphasises Wall & Alderson's (1993) argument that the teachers' commitment to the educational policy is a type of washback. The teachers also revealed their satisfaction with this practice as it helped in indicating the level of difficulty of the SECE for students. Therefore, the teachers considered the use of public examination content as a mock test for students, yet, such a practice had negative implications on students' learning behaviour and practices in the Libyan context. It forced the school students to neglect to develop their English language knowledge and skills as well as other academic skills, such as critical thinking.

However, the secondary school teachers' commitment to the assessment policy was not the only factor that made them use the select-response questions in their classroom testing. This study identified other factors that encouraged the participant teachers to adopt the public examination approach in their classroom testing, such as enhancing students' performance in classroom testing, eliminating cheating, teachers' lack of assessment literacy and self-protection. These factors are discussed in the following section.

7.1.3.1 Enhancing students' performance

Since the participant teachers believed that the select-response format was easier for both students to answer and for them to mark, as well as good preparation for the SECE (see 6.2.1.2), they employed this format in order to increase their cumulative grade towards the final score of the SECE. The teachers used the SECE format to ease the students' retention of information to pass the classroom tests and the school periodical tests. One of the teachers mentioned an example of this:

> *It is widely considered that teachers use the multiple-choice questions to make tests and examinations very easy for students; it is easier to mark and to ensure success for all students whatever their level, and I wanted them to get good marks in the classroom examination* (Interview T2).

This comment emphasises the importance for teachers to ensure that all their students pass classroom-based tests, especially those with lower abilities. The students' good performances in classroom tests and public examinations would ensure teachers avoided blame or embarrassment, which a poor performance by students would elicit in the educational and social contexts. It has been indicated in chapter four (the research methodology chapter) that the participant students (e.g. S4 and S14) had different educational backgrounds and had studied in different schools. They had moved to the school where this study was conducted because of the school teachers' good reputations in regards to their students' performances in the public examinations. S4 said:

> *I moved to this school because it has teachers for nearly all final year subjects, unlike the school where I studied my first and second year of secondary school. Students from this school always get a good score because the teachers have experience, and they have been teaching final year students for a long time. So, they are really good in preparing students for the final examination* (Pre-examination group interview 2*).*

The participant teachers also perceived the advantages of the SECE format given the time pressure on English language secondary school education. One of the teachers (T1) expressed this view as follows:

The second point in favour of using the multiple-choice technique is the investment of time, which we allow for students to answer classroom tests. The period time in secondary school is too short (Interview P1)*.*

This point of view is related to teachers' beliefs that the select-response technique reduces the memorisation of information required for the test (as discussed in 6.2.1.2), and the fact that the students do not need to spend time composing their answers, as they would do within comprehension or compositional responses.

One unanticipated finding was that T2's appreciation of select-response testing in increasing scores led her to rely on written tests to grade the students' coursework and neglect the coursework evaluation criteria prescribed by the Ministry. T2 replaced the marks allocated to English reading skills and speaking in the classroom activities with a written test on vocabulary and grammar, as students (e.g. S1, S3 and S7) described. S1 highlighted this as follows:

Since students can't speak and read, and as there are only two or three marks on speaking, pronunciation and classroom contribution in the marks' division of the coursework's evaluation, she [T2] told us that she would give us a spelling test or anything else to give us marks on that (Pre-examination group interview 1).

The described practices can also be attributed to the fact that the teachers filled their classroom time with their reading and explanations from the textbook as discussed in sections 7.1.1 and 7.1.2, which meant that there was no participation from students in the classroom discussion. Besides, the value placed on the final examination score could also have affected the teachers' decisions regarding such practices since such practices were not significant for the final overall score.

In addition, the teachers used other techniques to enhance the students' performances in their classroom testing. For example, they selected a specific part of the English textbook for each classroom test or the periodical test and offered clear explanations of which items would be tested. One of the teachers (T1) described that as follows:

My way to make the classroom tests easy for students is to divide the textbook into small chunks for testing. In every test, I choose a certain unit and ask students to focus on it for the classroom tests or the school periodical tests (Interview T1).

In the group interviews, the majority of students mentioned that their English language teachers' classroom tests were at a reasonable level for all students. For instance, in a discussion about assessment, students S8, S9 and S10 mentioned:

> *S9: {It is not so difficult} and goes with all students' levels in the class.*
> *S10: It is not difficult.*
> *S8: It means that even if someone didn't study hard for the exam, he/she will be able to answer it* (Pre-examination group interview 2).

The ease of classroom tests was considered "*reasonable*" since all students could increase their scores towards their overall summative result of the SECE which all teachers and all students wanted. They were not concerned about the nature of that achievement or the quality of the students' English learning outcomes, however, the implication of this is far-reaching since the assessment did not effectively provide a true distinction between students' performance, which ultimately affects the credibility of the SECE in the Libyan higher education context and in the social and economic environment. Besides, the ease of classroom tests did not motivate the students to improve their language competence or encourage effort since they would be able to perform well in the classroom tests using basic language knowledge as S8 pointed out above. Such assessment practices contravened the desired English language learning outcomes at secondary school level in the Libyan context.

7.1.3.2 Eliminating cheating

The participant teachers were also motivated to use the SECE technique in order to eliminate cheating in the classroom- and school-based testing. T1 described his approach as follows:

> *Personally, I started applying the public examination technique in the class tests; I prepare three different sets of questions for different groups about the same topic similar to what happened in the final exam. Then, I distribute the questions to students in which each two sitting together takes a different version of the questions. I tell them that by looking at [cheating from] the other colleague's paper, that will make them fail in the exam* (Interview T1)*.*

However, cheating was not highlighted as a significant issue in the school-based assessments. This is because the English language classroom test was considered as a reasonable and not difficult test (as discussed earlier in section 7.1.3.1). Hence,

cheating was not considered necessary to pass the classroom tests. Another possible explanation is that the recognition or acceptance of cheating in the school context threatens the school and its practitioners' credibility and would be damaging to their reputation both in educational and social contexts. In the public examination, the case is different – the public examination committee that oversees the SECE administration is from different schools across the region, and responsibility cannot be attributed to a single institution.

7.1.3.3 Teachers' lack of assessment literacy

The teachers' lack of language assessment literacy was also identified as a significant factor that motivated teachers to adopt the public examination format and content in their classroom tests. Berry & O'Sullivan (2016) defined assessment literacy as the required level of knowledge about assessment by individuals to interact with assessment in a meaningful way. In this study, the teachers admitted their lack of comprehensive knowledge of language assessment. Similar to observation reported by Gebril (2017) about the Egyptian teachers, the participant teacher reported that their teacher training programme was not effective in providing sufficient focus on language assessment literacy. For that, they relied on the traditional assessment method of summative examinations. For instance, the teachers mentioned that their professional development had fossilised at a certain level as they had been doing the same teaching and assessment practices for years. This situation is evident in the case of T1, answering a question about the need for training for professional development, who said:

> *Yes of course! With the development of teaching methods, we need to cope with the modern teaching and assessment methods in the world, especially in language education. We shouldn't depend only on what we have been teaching for a long time* (Interview T1)*.*

In the interview with one of the inspectors (P4), he stated:

> *The teachers can't prepare a task to test everything or to prepare different types of examinations to assess different types of language knowledge because they still follow the traditional way of preparing their assessment because of the lack of knowledge and experience in language assessment* (Interview P4)*.*

The above statements reflect the teachers' awareness of the need to develop assessment literacy in terms of the message of the integrated language curriculum, and pedagogy, and its relevance in global, contemporary educational contexts. The comments also imply that the participant secondary school teachers' assessment practices have been rooted in their professional experiences rather than being developed from their learning experiences, as Sheehan & Munro (2017) found in Europe. Although Alderson et al. (2017) and Brown & Bailey (2008) argued that the teacher training processes or levels of professional training had a significant effect on washback, the situation seems to be different in this example of the Libyan secondary school context. The head of the English language inspection unit (P1) noted:

> *We noticed that the newly graduated teachers have more knowledge about different types of assessment strategies and they could utilise them to prepare their own questions. However, the experienced teachers only follow the traditional methods of constructing their examination and they still have the same knowledge about assessment and teaching since they graduated* (Interview P1)*.*

The difference between the teachers was attributed to their level of teacher education, as T1 mentioned that he was trained in assessment for general education, but not in assessing language learning and outcomes. However, in spite of this difference, the washback was similar for both teachers. P2 in the above extract only indicates that the newly graduated teachers have the ability to use different assessment techniques, but they did not actually do that. Besides, lack of assessment literacy was not limited to the ELT teachers, but also to the wider education practitioners. The representative of the Examination Office (P2) admitted the need for training courses on assessment literacy to develop wider educational practices. He explained:

> *We need courses about assessment for all teachers, especially those who supervise the school examination. Those teachers usually face difficulties in implementing any changes in the assessment criteria in their work and become unable to explain it to other teachers in the school. For that, they should be trained to do their job in the right way* (Interview P2)*.*

In such scenarios, the participant teachers of English language were anxious about change affecting their students' attainment, which would affect the judgement about

their classroom performance and subsequent career. Thus, the participant teachers adopted the public examination questions to deflect any possible criticism or to avoid revealing any deficiency in students' learning outcomes. They considered the SECE system as a state-validated, ready-made examination that was safe for them to use in their context. Therefore, the adoption of the SECE in the classroom can be interpreted as undesirable washback as it constrains all practitioners' professional development and their creativity and spontaneity in educational assessment.

In addition, the teachers' lack of subject-specific knowledge integrated into the ELT textbooks generated a lack of confidence in preparing their own test questions. For instance, one of the teachers (T2) declared:

> *We sometimes get help from teachers of other subjects such as statistics, biology and chemistry to be able to answer the students' questions about certain* subject-specific *knowledge included in the English language curriculum* (Interview T2)*.*

Another example of that is the following statement by the head of the English inspection unit (P1):

> *The teachers had to teach things that they didn't understand, and these things were important for their classroom testing and the final examinations, so definitely they would have problems in preparing examinations in these aspects* (Interview P1)*.*

A similar connection between the teachers' subject knowledge and their assessment practice was observed by McSweeney (2014) and Qi (2005) in which they argued that the teachers' abilities in the language they teach determine the level of washback on classroom teaching and assessment practices. This connection could be the reason why the participant English language teachers tend only to use questions from the textbook lesson questions or those from previous public examinations, rather than preparing their own questions in their classroom testing. That is because the answers to these questions are given in the prescribed teachers' book or in the commercial SECE related materials.

7.1.3.4 Self-protection

Libya continues to be in a state of on-going conflict, and as a result, society is militarised and weapons are common. One of the significant findings that emerged

from this study is that participant teachers used the public examination format and content in their classroom testing to protect themselves from dispute or risk. The participant teachers revealed that marking select-response questions helped them to avoid arguments with the students, as well as with their parents, about their scores in their classroom tests. Such disputes could threaten teachers' safety, especially with the absence of effective state security forces and the widespread availability of weapons. One of the teachers (T1) explained:

> *I use the SECE approach for the accuracy of marking. For example, in English language, a letter can change the vocabulary meaning and can change answers from being right to wrong answer. When a student makes such mistakes in wh- open-ended questions, it is difficult to take a decision whether the answer should be considered as correct or not, because of the error or could be due to bad handwriting. Such judgement became more problematic for us now since they lead to arguments with students and their relatives. But here {in multiple-choice questions} there is a kind of clarity in deciding whether the answer is correct or not, which avoids such arguments* (Interview T1).

Although the participant teachers had not specifically experienced any aggressive behaviour, they experienced hostile behaviour and were aware of violent events happening to teachers at other schools and universities across the country. This finding enforces results that were identified by Pathan et al. (2016) in the same context, reporting that students' parents were seen to behave in a rude and hostile manner with school teachers and demanded forcefully that their failed children pass their courses. In this study, one of the inspectors (P3) commented as follows:

> *School teachers are facing real problems with the students' behaviour, and are anxious about that, especially at secondary schools. I witnessed a lot of events where students disrespect the teachers in classrooms and nothing can be done about that due to the situation of the country* (Interview P3).

This comment emphasises the challenges that the teachers face with student discipline, which could result in pressure and anxiety. Commenting on the issues of social challenge and safety, one of the teachers (T1) gloomily said that the school teachers had been left alone to face real difficulties on different levels within education and social contexts and have been accused of deficiencies in the public school sector.

The following extract from the head of the English inspection unit demonstrates the reasons for such a feeling:

> *My question: Do inspectors consider the contextual factors that affect students' motivation to learn or to attend the school and other contextual factors such as the teachers' security given ongoing conflict when evaluating the teachers' practices?*
>
> *P1: These issues are a part of the teachers' responsibilities. These issues are the responsibilities of the social support team who specialise in motivating students and supporting teachers by giving advice when they need in schools. But in this region, we haven't got those specialists at schools.*
>
> *My question: Who do you think is doing this work now?*
>
> *P1: I don't know, it could be teachers as each teacher is responsible for motivating students in his class and dealing with the challenges they face, but I have nothing to do with this, I have a plan and criteria from the Ministry which I follow when evaluating teaching practice* (Interview P1)*.*

The above interview underlines the focus of the inspection and its failure to consider the obstacles that could interfere with teachers' implementation of their practices. The practitioners showed a high level of commitment to the policies attached to the public examination, which would clearly have an effect on their professional evaluation. They insisted that teachers had to finish the curriculum and prepare students for public examination but at the same time, they ignored the challenges which could inhibit these practices. It also inhibits the implementation of CLT principles which were not significant in the public examination.

The findings from this section confirm my argument that final year secondary school English language teaching and assessment are orientated towards the assessment policy of the SECE. Thus, these practices were considered as washback of the SECE policy and approach.

7.2   Learning strategies

The research data analysis shows that the students' characteristics, such as attitudes and perceptions towards language education, and their motivation, had a major effect on their English learning strategies. The participant students adopted a measurement-

driven approach to English language learning in the school context. This learning approach was explicit in the practice of learning test discrete points such as comprehension texts and grammar instead of developing English language knowledge and skills more generally. Learning to the test is a common washback of high-stakes assessment reported by several research studies in different contexts (Damankesh & Babaii, 2015; Mogapi, 2016; Xie & Andrews, 2013). These studies highlight that learners tend to extensively focus on test-preparation material and activities. The findings of this study showed that the participant students' learning patterns are similar to those described by Damankesh & Babaii (2015) in Iran and Sukyadi & Mardiani (2011) in Indonesia. For example, Damankesh & Babaii (2015) reported that Iranian high school students placed their focus on the tested parts of the textbook from which is used to construct the final examination questions. They also practised testing through the review of papers of the previous examinations and using test preparation materials related to the textbook. Moreover, Damankesh & Babaii also highlighted that the Iranian high school final examinations directed students to develop different test-taking strategies with the aim of achieving higher scores. Similarly, the participant Libyan secondary school students prioritised translating the meaning of vocabulary from the taught parts in Arabic, engaging in test-preparation activities, adopting rote learning and various coping strategies as their English language learning strategies. These strategies are discussed below.

7.2.1 Translating the textbook

The participant students considered mastering the tested parts from the textbook by translating them into Arabic as their main learning strategy. This strategy aimed to understand the scientific, historical and cultural information, and grammar as that was the only knowledge they needed for the public examination of the English language subject. S8, for example, described his learning strategy as follows:

> *I focus on translating the texts because there are lots of new words that we don't know and because the teacher doesn't focus on the use of these new words in the class. We only know its meaning in the textbook for the examination* (Pre-examination group interview 2).

The above comment indicates the teaching approach of GTM described by Pathan et al. (2016) in the Libyan school language classrooms context. However, S8's comment

emphasises two reasons for using such learning strategy in her context. First, S8 refers to the challenging content of the ELT textbook by which they encountered difficulty in understanding many new words, especially the subject-specific knowledge introduced in English. Therefore, they focused on translating the textbook content to ensure comprehension for the classroom testing and public examination. Second, S8 identifies the teachers' practices as another factor for adopting this strategy as the teachers did not identify new words and how to write them and use them in context. This practice was previously associated with the time pressure of finishing the textbooks and the process of test preparation and performance (see 7.1.2). The following extract from the interviews reinforces the significance of the time effect on other students, such as S12, S7 and S6, in prioritising the translation of the textbook rather than focusing on developing their vocabulary:

> S9: She [T2] is speeding incredibly through lessons and classes.
>
> S7: It may be because she wants to finish the textbook quickly so that no one blames her for not completing it.
>
> S6: But it is difficult for us to understand all new information and vocabulary every day. There is a lot of new words in each lesson of the book. So, we only get the chance to translate the meaning of these words, not an opportunity to practise them in the classroom (Pre-examination group interview 2).

It may be inferred from S6's words that the intensity of exposure to the new vocabulary and its level of difficulty made it hard for the students to process these materials in the classroom context. Thus, the students focused on translating the lesson's vocabulary to be able to deal with it outside the classrooms in their preparation for the public examination. Some students (S13, S1, S2 and S11) also reflected on their experience as the reason why they took the decision to adopt this learning strategy:

> S13: What happened in the ninth grade exams (BECE) is that we only studied half of the book. So, students blamed the teacher for that as they did not get through the curriculum and understand all the information although she explained well the part we studied.
>
> S12: Yes, and many students weren't able to answer because they didn't know the meaning of words or what was required, because we didn't finish the curriculum at that time.

S11: It is possible we will face questions, which we have never thought about or seen before if we don't finish the curriculum.

S13: Because it depends on interpretation. Now, nobody understands the lessons and all students are translating (Pre-examination group interview 3).

Since the focus on understanding the content in detail in the classroom failed to prepare students for the public examination, they translated each test item as a technique to go through the whole textbook rather than developing vocabulary knowledge in the classroom to avoid that situation. Equally, when students were successful preparing for the BECE, a similar examination to the SECE, this also influenced their learning. They considered their prior experience of preparing for public examination as a certified and successful example of examination achievement. They also adopted the same test-preparation activities as discussed in the section below. This reinforced the argument of Elder & O'Loughlin (2003) and Stoneman (2006) that students' past learning experiences have a major influence on their choice of learning strategy and practices.

7.2.2 Engaging in test-preparation activities

Results from the study reveal that the participant students engaged in test preparation to develop their performance in classroom tests and public examinations. It is possible that the insufficient emphasis on the application of knowledge in educational systems makes individual participants value success in a test more than the construct of knowledge and understanding as Hyland (2011) stated. Therefore, the students engaged in activities in order to get a high overall grade rather than to engage in actual cognitive English language learning activities. The purpose and the nature of the SECE encouraged participants to focus on rote learning and revising examination content from the textbook over in-depth learning with the objective of developing English language competence. For example, S12 and S13 identified grammar as the only thing they needed for their classroom tests and the public examination. They said:

S12: I focus on doing the grammar exercise from the books because grammar is the basis of the teacher tests and public examination in school education.

S13: Yes, grammar is important because if we understand grammar we can get lots of marks because there are many questions about grammar in the public examination (Pre-examination group interview 3).

Students' learning focus in the above comment aimed to increase their familiarisation with the tested aspects of textbook content in order to enhance their performance and score more highly in the classroom tests and public examination. The students identified the drilling of grammar exercises as the most common test-preparation activity in and outside the classroom. Although it can be argued that the students' grammatical competence could be developed because of this language learning practice, however, learning linked with a test is unlikely to lead to a meaningful improvement in English language knowledge or to a development of communicative ability. In order for such positive washback to occur, Bailey & Masuhara (2012) argue that examinations should be matched to objectives and assist language learners to obtain and develop their knowledge and skills as part of the test.

Other test-preparation activities used by the participant students were similar to those observed in different Asian countries (Damankesh & Babaii, 2015; Aftab et al., 2014), such as answering and reviewing the exercises in the textbooks and revising previous versions of the public examination. For instance, S5 indicated as follows:

> *It is important to focus on the question, which might come in the final examination such as the questions of the lessons and units in the textbook. These questions usually come in the final examination in different forms. So we try to practise all these types of questions at home to prepare for the exam. The teacher told us that we would do these activities after we finish the textbook* (Pre-examination group interview 1).

In this account, S5 devoted substantial time to prepare for the public examination outside of the school classroom time. Again, this was necessary because students did not have sufficient time to practise these activities within the classroom. The need to practise outside of the classroom encouraged the students to rely on test-related material rather than using the school textbooks. The students also revealed that their teachers had advised them to use these materials since they perceived them as being relevant test content sources. Unique to the Libyan context, the classroom tests and public examination are constructed from the exact linguistic items of the prescribed textbook. Consequently, independent publications have emerged which specifically provide answers to the exercises and activities of the school coursebooks and workbooks. These publications are referred to as "answer-key" textbooks (such as

Almomtaz, Almurshed and Alnajah). Saukah (2015) argued that the proliferation of answer-key books for the public examination has a strong connection with the students' test-preparation learning strategy and test design in the international context. The students point out that they use answer-key textbooks to identify and practise the tested parts of the curriculum. For instance, S6 stated:

> *Now, the teacher does not focus on explaining the examination questions about each lesson in detail. He just highlights that this information could come in the examination and then continues teaching. But I try to find out how the lesson information can come in the examination. I will use answer-key textbooks and previous examination papers* (Pre-examination group interview 2).

The above comment indicates that answer-key textbooks allowed more opportunities for the students to practise the test-preparation activities than in the curriculum textbooks. However, answer-key textbooks have had a negative impact, minimising students' reliance on the school ELT-prescribed textbooks as a means of language learning in the context. They also adversely affected the language learning process in so far as students use answer-key textbooks in their classroom learning activities. In the classroom observation, the students were observed using these materials to answer the teachers' questions regarding the meaning of the vocabulary and to answer the lesson questions regarding reading comprehension paragraphs and texts (T1's classroom observation 1 and T2's classroom observation 3). In conversation with students after the classroom observation, the students also mentioned that they used the answer-key textbooks when doing homework. An explanation that could be given for the students' use of answer-key textbooks in their classroom activities is the importance of accuracy highlighted previously in this thesis by S12 (section 6.1.2) in school English language classrooms. For the participant students, it was important to handle these activities effectively in order to increase their scores towards the cumulative grade of the SECE. It is interesting to note that the students avoided engagement in different learning activities, such as doing homework and contributing to classroom discussion, without the use of answer-key textbooks. These learning practices were considered to be test-preparation activity since they aimed to promote the SECE's overall score. Therefore, such material had a potential negative impact on language-learning development since it inhibited the school students' exposure to

English language by narrowing the content of the syllabus and oversimplifying some concepts while undermining coursework activities.

In the post-examination group interview, the students suggested that such practice had assisted their target of passing the public examination with good scores. Weir (2005) argued that test-preparation activities could have a positive effect and develop candidates' linguistic and metalinguistic knowledge required for a test. Therefore, traditional test-preparation practices for high-stakes examinations are found to be desirable by assessment stakeholders in the international context (e.g. Damankesh & Babaii, 2015; Nikoopour & Farsani, 2012; Weili, 2010). However, if the test led students to develop test-taking strategies which promote performance without enhancing the ability being tested, then it can be considered as producing undesirable washback. What can be seen in the study context is that the SECE approach encouraged the secondary school students to engage in test-preparation activities while neglecting the development of language knowledge in the school education context.

7.2.3 Rote-learning practice

The data also revealed that the participant students adopted a memory-rote learning strategy in their English language classrooms. They concentrated on memorising the meaning of vocabulary and information as a key cognitive process to perform in the classroom tests and public examination. Such washback has been discussed in some Asian and MENA contexts (Aftab et al., 2014; Kilickaya, 2016; Pan, 2009). For example, Pan (2009) maintained that standardised tests restrict learning as students focus only on cognitive dimensions of learning tested discrete points of knowledge, and ignore real-life knowledge. In this study, S7 stated:

> *Students, and especially in SECE, are supposed to study the whole final year curriculum for public examinations and memorise information of all passages in order to get used to them before the final exam* (Pre-examination group interview 2).

This learning strategy was also found by Green (2007, 2013) in connection with the design of public examinations. Since the participants perceived the select-response questions required remembering information (discussed in 7.1.3.1), the students relied mainly on rote learning, perceiving it as compatible with the public examination

approach. The participant students claimed that select-response questions eased examinations and did not require too much effort. For example, in the post-examination group interview, S15 replied to a question about the SECE score reflecting language knowledge and ability as follows:

> *I don't think so, because this type of question made me memorise not understand. What I did is copy, paste, nothing out of my head* (Post-examination group interview).

The quotation above reveals that the students used a shadow-learning practice. This learning practice encouraged the participant secondary school students to memorise a large volume of information from the curriculum before the examination, but it did not encourage any deeper cognitive process. However, Patrick et al. (2016) also found that standardised tests encourage students to memorise the tested items of the ELT textbook for what has been described as *short-term retention* to be recalled during the examinations. This effect was reported in this research by S13 who believed:

> *I challenge students to remember anything after the SECE from those who take it now. We only memorise what we study and answer the examination questions. I mean we didn't understand anything about it* (Pre-examination group interview 3).

This comment emphasises that the Libyan public examinations system promoted short-term memory retention among the participant secondary school students in their English language classrooms. Particular to this study, rote-learning strategy has been encouraged by various language teaching related factors and examination policy and practice. For example, the students preferred this learning strategy due to the fact that they had difficulties in building a comprehensive knowledge of English during their earlier school education as discussed in the previous chapter (section 6.1.3). Since the students missed a lot of lessons from the structured English language syllabus throughout their earlier secondary education, it was difficult for them to develop a reliable knowledge of English. Consequently, memorising the target grammar and vocabulary was considered as the best learning strategy to prepare for the examination. The students also highlighted that the heavy focus on lecture-based teaching made it difficult to promote communicative activities in the English language classrooms. A similar finding was reported by Cheng (2005), attributing the Taiwanese

students' preferences for traditional learning strategies (e.g. reading texts and memorising grammatical rules, vocabulary and phrases) to the teachers' use of teacher-centred instruction in the language education context.

In addition, the assessment policy and practice used in the Libyan education system had a significant impact on the adoption of a rote-learning strategy among school students. Since the questions on the SECE are constructed from the exact linguistic resources of the students' textbooks, it encouraged the rote learning of the language input rather than the application and understanding of learning outcomes. The students' knowledge of the examination corpus encouraged them to focus on the tested grammar and technical information within the ELT textbook, rather than the development of knowledge or engagement with comprehension texts and activities. S6 highlighted that:

> *It is good to memorise all the material because it will be the same in the examination. Only a few words will be changed in the questions. For example, in true/false questions, they might change the verb from its original form or they might change a date of an event or any piece of information. So if we memorise these facts and information we will be able to answer. Even in grammar in the multiple choice, they will bring the same sentences from the curriculum* (Pre-examination group interview 2)*.*

The above comment emphasises the significance of the level of accordance between the teaching and examination content in the Libyan school system. This supports the argument by Weir (2005) that the knowledge of the task requirements facilitates goal-setting and monitoring as well as enabling students to choose the most appropriate learning strategies. Foster (2013) and Saukah (2015) contended that the close relationship between the curriculum and public examinations could encourage students to rely completely on memorising the tested items from the curriculum in their learning. However, this study found that such a relationship between the textbooks and public examination encouraged the participant students to contextualise an examination-learning environment, where they focus on memorising the exact test content they study and which will be examined rather than to be used for meaningful interaction. This impact is related to the fact that the SECE placed an overemphasis on rote learning and insufficient emphasis on the application of language knowledge as Hyland (2011) described. Besides, as the exact linguistic items of the prescribed

textbook were used to design the public examination items, the English examination questions were repeated over the years. Thus, the students focused on these questions in order to build sufficient test schema content as Weir (2005) described to enable them to enhance their performance in the examination. The repetition could also be a significant factor in the widespread cheating phenomenon in the public examination. Since the questions and the answers of the textbooks are the basis of the public examination, this would increase the possibility of cheating in the context.

Hughes (2003) highlighted that such a gap between testing techniques and the objectives of the course would always have the consequence of destructive washback. The examinations that focus on knowledge-based memorisation and the recall of information ignore many other qualities that are essential to students' success in language learning. In the Libyan school context, it could be worth assessing the application of the taught knowledge or the wider language knowledge integrated into the school curriculum outlines rather than repetition of the exact exercise content that has been performed before. That could encourage school students to employ different learning strategies to develop their English language knowledge and competence to perform in their classroom and public examinations.

7.2.4 Developing test-taking strategies

The Libyan secondary school students also developed test-taking strategies as preparation for the public examination. In spite of the extensive preparation that students did for the public examination, they were aware of the fact that it is not always possible for them to remember the answers to all questions. Such finding is linked to the argument by Green (2007) that there is a strong relationship between test design and developing test-taking strategies. Accordingly, the participant students tended to develop a coping strategy to enable them to respond to the examination question in cases when it is difficult to be confident of the accurate answer to the examination questions. S8 described that as follows:

> *This type of examination is tricky. It is very complicated when the answers could be the same with a change in a word or a letter, which makes lots of students confused. In this situation, we need to find another way to answer the questions so that we don't lose much time on one item as the exam is a bit long* (Pre-examination group interview 2).

In another group interview, S1 also explained:

> *We may come to a situation where we don't know what to write down, and here,*
> *away from our ability in English, we have a chance to succeed even if we don't*
> *know how to write* (Pre-examination group interview 1).

In this study, the students identified different test-taking techniques, such as guessing intelligently, the use of word collocations and sentences to identify a theme and addressing the test items. They used these techniques to answer the English classroom and school periodical tests, and the public examination. Only S11 highlighted his focus on developing his use of general language knowledge, such as considering the overall semantic context and eliminating inaccurate alternatives in questions, as a coping strategy for the SECE. In the pre-examination interviews, S11 mentioned that:

> *Sometimes, I try to understand the common words that I know to understand*
> *the questions if it is something I remembered or I know in general then I would*
> *be able to answer the question according to my common knowledge of English,*
> *especially in grammar* (Pre-examination group interview 3).

Compared with S1 in the previous paragraph, this distinction between the students' test-taking strategies could be related to their individual differences in their original English language proficiency. Therefore, such differences might be applied to the different types of language-learning strategies discussed in this chapter. In the case of S11, it can be argued that the SECE encouraged some students to use a few helpful test-taking strategies, such as employing other test items, the elimination of obvious inaccurate alternatives in the questions and guessing intelligently. These strategies could encourage the secondary school students to use their actual receptive English knowledge. However, Weir (2005) stated that such cognitive processing would likely bear little resemblance to the way texts are processed for information in real life. Hence, the use of any type of test-taking strategy that could improve a test score without developing measured knowledge is a type of negative washback. In addition, as the SECE is norm-referenced, it yielded a negative washback on the secondary school students; it affected their learning behaviour in that they were encouraged to develop their score using whatever means and ignore the development of their language knowledge and competence.

7.3   Chapter summary

This chapter has discussed the participants' experience of the washback of the public examination policy and practices in their classroom context. It focused on washback on classroom teaching approaches, activities and practice, and the students' learning strategies.

This chapter also emphasised the fact that the participant final year secondary school teachers experience pressure and anxiety because of the misaligned demands of language education policy in comparison with the policy of the public examination. So, due to the power of the public examination policy and the value of the SECE score in the Libyan education context, the ELT teachers and inspectors oriented their practices towards assessment policy. Therefore, I argued that the secondary school English language teaching practices signified the washback of assessment policy and practices in the secondary school education. For instance, the participant teachers used the teacher-centred approach and the GTM approaches to be able to superficially finish teaching the English textbooks as the public examination policy requires. The teachers also prioritised drilling grammar and raising awareness of the possible examination questions rather than developing the students' spelling, vocabulary knowledge and writing activities. The focus on teacher-centred and GTM approaches adversely affected the nature of the usual classroom interaction and activities and the teachers' responses to the students' demands in their English language classrooms. In classroom and periodical tests, the participant ELT teachers used the public examination format and questions to train students on the system in accordance with the policy requirement. However, they prioritised the use of summative tests rather than implementing the criteria of evaluating the students' coursework more broadly in order to increase the students' scores towards the final score of the SECE. Another key finding was that the ELT teachers have a lack of language assessment literacy, which encouraged them to use the public examination approach and content in their classrooms to protect themselves from possible criticism in assessing language-learning outcomes.

This study also highlighted the contextual challenges that the participant teachers and inspectors face during their work in the context of the ongoing conflict of Libya. These challenges had a significant impact on the participants' practices, especially since

these challenges are not considered in the education policy agenda. A key finding in this perspective was that of the final year secondary school teachers' use of the select-response format as a self-protection technique where liability and responsibilities are problematic. This helped them to avoid arguments with students and their parents regarding the marking of classroom tests, which could result in aggressive and intimidating behaviour since these scores are significant towards the SECE.

In the second part of the chapter, I argued that the participant students are largely passive in the English classroom. The well-intentioned classroom teaching approaches for the sake of the SECE led teachers to neglect the students' language learning demands in the final year of secondary school. Accordingly, the participant students utilised different learning strategies to cope with the educational practices. In particular, the students prioritised translating the taught lesson to be able to deal with it outside the classroom context as the teachers' focus was to finish the textbooks but not to ensure students' understanding of the lesson. They also utilised the rote-learning strategies since they had difficulties building the systematic language knowledge proposed in the English language curriculum by the Libyan education system. Since the English public examination is drawn from the exact exercises and vocabulary of the prescribed English language textbooks, the students considered rote-learning and test-taking strategies as effective to perform well in the classroom tests and public examination. I also argued that the students engaged in test preparation to take an active role in their demands in preparing for the public examination. As a consequence, the students tended to rely on the commercial test-related materials to expand their practices outside of class. This undermined the students' reliance on the prescribed school material as a means of English learning and yielded a negative washback effect on school students' exposure to the English language EFL context. They used these materials to do homework and take part in the classroom activities, which also adversely affected their engagement in the language-learning process and activities.

The next chapter will examine the SECE score validity according to the socio-cognitive framework for test validation and the washback effect.

## 8 Chapter Eight: washback on school language learning outcomes

This chapter discusses third research question in relation to washback on the students' language learning outcomes. It starts with a discussion of the statistical records of the SECE success rate from the first sitting in the Libyan context. Then, it analyses the effect of students' experiences of the classroom teaching and learning practice on their performance on public examination and their language learning outcomes according to Weir's socio-cognitive framework for test validity. The discussion will be based the students' score on the SECE and perceptions about their English learning development (scoring validity). Since scoring validity is affected by both contextual and cognitive validities (Weir, 2005), the chapter covers how these validities are linked to each other in the study context and how they affected students SECE score and English language learning. These findings offer more insight to support the growth of the students' language learning outcomes - to take an assets-based and improvement approach rather than to search for deficits and create a cycle of blame within the Libyan education context.

8.1 Students' learning outcomes

It should be highlighted that I was not able to get the records of the SECE before 2011 since these records have been reported as lost or damaged during the conflict. The general observation from the statistical records of the rate of Libyan students' successes in the SECE from the first sitting (Figure 5) showed an increase in the number of SECE candidates achieving high scores during the earlier years of the implementation of the reformed public examination system. The representative of the Examination Office reported that the best success rates in the SECE in the history of the Libyan education system were recorded after 2009, the year when the current public examination system was implemented. He claimed that the success rate has doubled in the recent years.

As can be seen from Figure 5, nearly 90% of the candidates passed the SECE from the first sitting in 2014/2015 when the participant students undertook the public examination. In that year, over 80% of the candidates got scores that were classified as *Excellent (with overall achievement 85% and above)* or *Very Good (with overall achievement 75% and above)* according to the reports published by the Ministry of Education.

Figure 5: The Libyan secondary school students' success rate in the first sitting from 2011 to 2017.
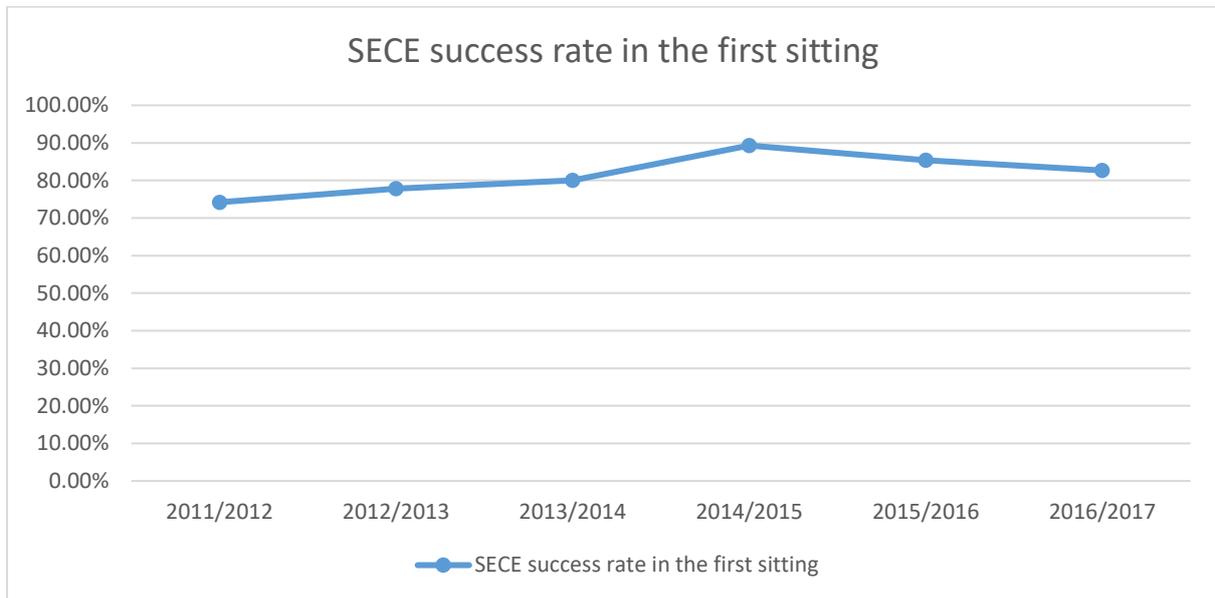
In the post-examination interview, the seven participant students mentioned that they successfully passed the SECE at the first sitting. They showed a high level of satisfaction with their examination results. Five of the students revealed that they got scores of *Excellent* while the other two had scores of *Very Good.* On the basis of the perceived good performance in the examinations, some students (S1, S4, S11 and S14) mentioned that they had expected their success in the SECE, not only in English but also in all other school subjects. For example, S11 said:

> *I was expecting these results, as we knew that we did well in the examination. We were focusing on the examination all the time during the year by studying what is important for the final exam and even in the classroom tests, most of the exam questions were similar to the final exam (The post-examination interview).*

About their success in English language, the students believed that these results came about because of their ELT teachers' classroom efforts as well as of their own efforts outside the classroom and in preparation for the public examination. However, although they acknowledged that they had achieved their targets, the students believed that neither their language knowledge nor their ability in English use had developed as much as their scores. They mentioned that they had not acquired

satisfactory knowledge that enabled them to use English to achieve simple speaking or writing tasks. The students commented on their language ability as follows:

> *S14: My problem is that I haven't learnt how to link sentences, I know vocabulary but I don't know how to link them together to form sentences for speaking.*
>
> *S1: Me too, I know only a few words but they are not enough to be able to speak well. And I would like to understand grammar, it is a problem to speak well about different things.*
>
> *S4: Now, my problem is in understanding long sentences, I can read but I can't understand the written words, and my English handwriting is good as well (*Post-examination group interview).

It is clear from the above interview extract that the students learning outcomes are relatively poor. According to the objectives of the final-year secondary school curriculum, students should have developed an adequate speaking ability to be able to engage in tasks such as solving puzzles and responding to suggestions, telling a news story, giving advice, telling a story from pictures, talking about books, exchanging information, giving instructions and opinions (see Appendix 1). I locate these skills in B1 level (Intermediate) according to the standard of the Common European Framework of Reference for languages. It is clear from the students' reflection on their speaking ability that those Libyan students were below the desirable level of English ability. From their reflection, they can only handle basic questions about personal details, as such questions require a single, simple sentence. Evidence accumulated through this study clearly indicates that the classroom educational practices were aimed at and focused on preparing students to succeed in the final examination of the SECE but not on developing English language competence. Therefore, the following part of this chapter examines the SECE score validity according to Weir's socio-cognitive framework for test validity.

8.2 Weir's (2005) socio-cognitive framework for test validation

The socio-cognitive framework for test validation consists of test-taker characteristics, associated with the context, cognitive, scoring, consequential and criterion-related validity.

8.2.1 Test-taker characteristics

The findings from this study showed that the psychological and experiential characteristics of the participant secondary school students had a strong impact on the improvement of their scores and had a strong effect on aspects of the context and cognitive validities.

8.2.1.1 Psychological characteristics

The participant students' motivation, as a psychological characteristic of a test-taker, was identified as a key factor in the students' successful practice towards their performance in the SECE. The data analysis revealed that the participant teachers and students were intensively motivated to engage in cramming practice in English teaching and learning, aiming to boost the students' scores. This motivation was linked to their negative perceptions and attitudes towards the value and the importance of the English language school curriculum and towards their perceptions of the aim of teaching English within Libyan secondary school education (discussed in section 6.1.1 and 6.1.4). These perceptions were a result of the traditional focus of the English public examination on assessing grammar and reading comprehension despite the integration of communicative skills in the English language textbooks. These perceptions diminished the students' interests to engage in "real" English language learning practice through the school English language material; instead, they were motivated to learn the curriculum only for the sake of the SECE. Therefore, the students, as well as their ELT teachers, set their main goal as preparing efficiently for the SECE and improving students' scores on the classroom assessment towards the final examination. The students and teachers' goals were also based on the value of the SECE scores for both teachers and students in the Libyan educational context. For the teachers, it was important to them that their students perform well in the SECE for their professional evaluation, to avoid social embarrassment and protect their reputation in the educational and social context. For students, the SECE score was valued for their future academic study. They were motivated to get high scores to get a scholarship, and to attend their chosen university education. The social value of attending an academic course, rather than vocational education, and the opportunity of employment in prestigious job available to graduates in the Libyan context motivated the students' good performance in the SECE. According to Weir (2005), these individual characteristics had a direct impact on the way that the participant

students process the task set up by the context validity and their cognitive or mental processing abilities. Consequently, the teachers and students focused intensively on test preparation activities. In the post-examination interview, the students described such a focus in practice as effective for their readiness to undertake the SECE before the examination, and as having a great effect on score improvement in general. Therefore, the participant students worked on improving their score rather than to develop language skills.

8.2.1.2 Experiential characteristics

The findings from this study suggest that the participant ELT teachers and students' experiences with public examinations and the education system were the most significant factors affecting English language teaching and learning practice in the study context. For instance, the teachers' successful experiences of preparing students for the SECE encouraged them to repeat their teaching and testing focus and activities in their classrooms. Their practices were considered as effective to promote the students' cognitive awareness of the necessary linguistic resources and mental processes for undertaking the English language exam in the SECE. In turn, that affected the participant students' motivations towards engaging in the classroom teaching and testing practices. It also had a positive effect on their perceptions of their readiness for undertaking the SECE and promoted their self-confidence. Similarly, the students' experiences of preparing for and undertaking similar examinations in the BECE had a direct impact on their performance and score in the SECE. Since students indicated their awareness of the examination requirements and criteria, that familiarity with the nature of the exam also increased their self-confidence in their chosen learning approach and test preparation activities, which, in turn, had a positive influence on students' performances in the SECE. Therefore, this study argues that the students had mastered the SECE resources and demands, in their classroom practices, according to their experience with the public examination focus and approach in the Libyan education system. The participant students narrowed their English language learning focus and limited their intellectual engagement with certain practices attached to the SECE rather than developing their language knowledge and ability.

The students' motivations and experiences seemed to promote their readiness for the SECE and minimise pressure or anxiety attached to the high-stakes examination. Therefore, it can be argued that such focus would develop their mental performance on the SECE tasks and scores. However, such motivation and practice could not facilitate the development of the students' English language knowledge and skills as proposed by the curriculum.

8.2.2 Context validity

O'Sullivan & Weir (2011) and Weir (2005) argued that there is an interactional relationship between the context and cognitive validities in the way they affect the students' performances, score validity and shaping construct validity within the socio-cognitive framework. The findings of this study showed that the aspects of the context validity directed the students' cognitive process of the cognitive validity. Hence, the context validity is discussed first in this chapter.

According to Weir (2005), the task context must be acceptable to both test-takers and testers as an appropriate situation to assess a specific language ability to achieve the context validity. In the case of the Libyan SECE, the lack of alignment between the focus of the English language curriculum (intended outcomes) and the focus of the SECE (task context) presents a critical issue in regard to the SECE's context validity. This will be discussed below in three sections, entitled task setting, task demands and task administration, which comprise the cognitive validity.

8.2.2.1 Task setting

Weir (2005) stated that clear knowledge of the requirements of the task facilitates goal setting, enabling students to choose the most appropriate learning strategies and to determine the type of targeted information. In this study, the discrepancy between the curriculum objectives of developing the students' communicative abilities and the focus of the SECE on assessing grammar and vocabulary influenced the participant students' and teachers' goal-setting in school language education. They considered the SECE as a norm-referenced examination since the test's score counted for more than language proficiency in the education context. Their understanding of the purpose of the test was translated to the need to focus on cramming practices in teaching and learning rather than developing language skills. Since the time devoted to teaching and learning the ELT textbooks was perceived as insufficient to explain and

understand it comprehensively, mastering the tested items to identify the targeted information was the purpose of the ELT classroom practice in the study context. The participant teachers and students focused on translating the tested items into Arabic in order to finish the textbooks to meet the purpose of the SECE. Being able to process the required knowledge from the curriculum in the SECE could have a positive effect on the students' readiness and performance in the exam, but not on the development of their English language knowledge and skills.

Moreover, the students also indicated that their knowledge of various SECE contextual elements had a positive influence on their performance, such as test familiarity, knowledge of response method and time constraints. For example, the students (e.g. S3, S11, S13 and S14) were aware of the need to mark the answer sheet clearly, to ensure efficient erasure of pencil marks when changing responses so that marks were not lost, as the examination is electronically marked. S13 explained in the following extract:

> *We have this information from our BECE, and we were told it is the same. It is only highlighting multiple-choice questions, matching, marking true/false question. Lots of students failed in the BECE because they didn't mark the answer sheet clearly* (Pre-examination group interview 3).

The students' experience of taking a similar examination in the BECE was the key source of their knowledge of the SECE contextual conditions. In the group interviews, the students also argued that the select-response items of the SECE contributed to the increase in their test score as its marking criteria primarily depends on the ability to convey these responses. They believed that such pencil marked responses protected them from losing marks because of their writing such as handwriting or spelling mistakes as in more traditional assessment. They also argued that the select-response format enabled them to use different coping strategies, such as guessing blindly (haphazard technique) or intelligently, using other test items, the elimination of obviously incorrect alternatives in the multiple-choice to answer the test questions. In this study, the teachers and students believed that coping strategies increased the chances to gain more marks in the current public examination design than the previous one. This perception is shared by other studies in Asia and the Middle East, which claim that the utilisation of test-taking strategies can help students achieve higher

scores in examinations (e.g. Damankesh & Babaii, 2015). However, the current public examination design had an undesirable influence on the development of the secondary school students English language knowledge and the ability. For instance, the SECE design was found to affect the ELT teachers and students' focus on developing English vocabulary spelling and writing skills in addition to the traditional neglect of listening and speaking in the Libyan schools. The teachers and students' beliefs regarding the purpose of the test and the aim of English language education, as discussed earlier in this section, seem to consider writing composition or ability as unimportant for the SECE. This effect was one of the significant factors that led the ELT inspectors (P3 and P4) and the representative of the Examination Office (P2) to hold a negative perception about the SECE format, especially since they believed assessing writing is crucial for language learning. In the interview with P1, as mentioned earlier in section 6.2.2, he declared that the SECE design has negatively affected the development of the secondary school students' writing skills, not only in English language but also in Arabic.

8.2.2.2 Task demands

The task demands also affected the students' performances in the SECE English examination. Weir (2005) argued that candidates' familiarity with the linguistic demands and content knowledge of a test could have a positive influence towards increasing their score on the test. That would enable them to choose the most appropriate learning strategies and determine the type of targeted information. In the study context, the fact that the linguistic resources of public examinations are the exact language lexis and grammar available in the prescribed textbooks enhanced the school students' familiarity with the SECE's content. This familiarity made participant teachers and students focusing only on the tested aspects in the SECE from the ELT textbook of the final year secondary school. They intensively concentrated on drilling and reinforcing grammar rules as test preparation throughout the academic year and memorising the scientific, historical or cultural information. The teachers were keen on the necessity of students' mastering the textbook lexis and specialised terminology by translating items into Arabic and giving examples in Arabic. In addition, the teachers and students spent a large amount of time on question-solving activities and administering mock examinations in the classroom-based testing. It could be the case, therefore, that these practices of teaching and learning supported the students'

knowledge of the linguistic demands of the SECE and positively affect their SECE scores, especially when the students were trained on these questions in the classroom-based tests throughout the academic year. This finding suggests that applying test-preparation practices before the examinations helped students achieve higher scores in examinations as discussed in different Asian and Middle East situation (e.g., Pour-Mohammadi & Zainol Abidin, 2011).

Likewise, it has been argued that the repetition of the examination input on language learning facilitates its comprehension and increases the students' proficiency when taking the examination (e.g. Chou, 2015). This study identified the repetition of the SECE questions over several years has helped the students to acquire sufficient schemata of the test content and enabled them to deploy appropriate skills and strategies to perform well in the SECE. For example, S8 mentioned:

> *As we expected, many of the examination questions were repeated, and we have studied them in the classroom or the teachers used them in the classroom and the school periodical tests* (The post-examination group interview).

S7 also mentioned that the teachers' focus contributed to improving their scores as they memorised all the tested aspects of the ELT textbooks after they had been translated and understood in Arabic. They stated that their teachers simplified the ELT textbook to enhance the students' interactions with the domain of knowledge relevant to performance in the criteria of the SECE. However, this practice cannot represent the authenticity of the situational (contextual) and interactional (cognitive) context of the test to be contextually valid (O'Sullivan, 2006). Weir (2005) stated that if candidates can predict the questions answers in advance, then their processing would be of an entirely different nature of authenticity (e.g. solely reliant on memory). Although Weir related this issue to the test security, however, the SECE resources and content were the main factors for the test-takers' predictions of the examination questions and answers in the case of the SECE. Therefore, such a finding suggests that the use of the examination content over time in the same format influenced the test-takers' characteristics and cognitive validity, and threatens the context validity. The SECE, however, is an inappropriate situation to assess the students' language abilities or encourage language teaching and learning practices aimed to develop the students' communicative competence. The research finding presented in this section has an implication on the socio-cognitive framework (more discussion in 9.3.3).

8.2.2.3 Task administration

Task administration is the final aspect of the context validity, which affected students' test scores and their learning strategies within the socio-cognitive framework. In the study context, cheating was identified as a significant factor that threatened the context validity of the SECE. During the pre-examination group interviews, some students used the word "help" to refer to cheating. S13 declared that:

> *Yes, because we may get help (cheating) from outside and that help (cheating) has an effective role in passing the examination and helps to get high marks* (Pre-examination group interview 3)*.*

The above quotation indicates that cheating in the Libyan public examination is socially accepted. This acceptance of cheating was related to different factors in the study context. First, there was intense pressure on students to succeed in the SECE created huge pressure throughout society to help students to succeed with high scores in order to enable them to attend university education because vocational education is not privileged in Libyan society. When the pressure to achieve a high score is intense, people sometimes resort to unlawful activity. In different parts of the world (e.g. India, China and the USA), many families were reported as resorting to cheating to help their adolescents to pass high-stakes examinations (Spann, 2015). Second, there are difficulties that individual schools could not resolve, such as the lack of teachers' availability which encourages cheating in the study context. It is possible that secondary school students do not study some school subjects, especially English language, throughout the academic year because of the lack of teachers' availability. This is a common issue in the rural areas in the south-west of Libya, where this study was conducted. An example of this was the subject of statistics, which was not taught for the participant students in the first semester or until the time of the data collection in March/April 2015. In spite of that, the students had to undertake a public examination in that subject. In this scenario, it seemed to be believed in the social context that cheating is the only way to "assist" those students to have a chance of success. In addition, with the pressure to pass the SECE, some students resorted to cheating by using threat or the force of arms. Students' relatives were also reported as forcefully accessing public examination sites in order to help students to cheat in order to pass. In the interview with one of the teachers (T1), he mentioned that he

decided not to supervise the public examinations due to potentially adverse circumstances. He explained that:

> *I have been offered to take charge of an examination site of the public examination, but I refused to do that. As you know, if I did the job properly, it would be a problem for me and my family as people will come and try to help their relatives and children* (Interview T1).

T1's decision can be attributed to the anxiety associated with the challenge of securing public examinations in the on-going conflict context of Libya. The cheating phenomenon in public examinations was significantly increased due to the widespread number of people carrying weapons, including students themselves. There was widely shared a photo on social media of a student with a gun on the table, believed to belong to a Libyan student in the SECE in 2014/2015. In such a context, it would be difficult for the exam administrative to implement discipline at the public examination site. Although examination security is an international phenomenon, the challenges are unique in an on-going conflict context, such as Libya where preventing cheating in public examinations could be life-threatening to the educational specialists. However, the significance of such challenges differs from one region to another even in the Libyan context.

Another important factor that leads to cheating is administering the same SECE examination in Libyan national schools and in schools belonging to the cultural affairs of Libyan embassies in different parts of the world. The time difference between different countries facilitated a new type of cheating in the public examination. For the first time, the public examination witnessed the phenomenon of the spread of examination questions on social media in 2014/2015 public examination. The SECE candidates in different parts of the world, especially Malaysia, uploaded their examination questions on social media, which enabled other students, who had not yet entered the examination, to get access to the examination questions. However, a recently official measure has been taken to prevent this by creating different questions for different international time zones. Still, the Ministry of Education witnessed a leakage of public examinations questions in certain subjects in 2016/2017. Similarly, these questions circulated through social media before the time of the examinations. These findings suggest that the on-going conflict context has resulted in widespread

corruption that undermines educational quality as an inevitable consequence of the overuse and misuse of high-stakes testing in context.

Overall, the research findings showed that the students' familiarity with the SECE criteria, response method and linguistic demands was reported as having a significant positive effect on the students' performance and scores in the SECE. However, the lack of alignment between the curriculum and examination focus makes it difficult for the SECE to indicate the students' levels of English language knowledge and skills. Since the exact language lexis and grammar of the textbooks are used to construct the public examination questions, that makes the Libyan SECE a repetitive practice rather than a discriminatory assessment practice. With the repetition of the examination's questions and the students' training on these questions throughout the academic year, the students build sufficient schemata of the test content, which enabled them to perform well in the test. Consequently, it would be difficult for such practice to represent a real context of testing or assessment, as Weir (2005) argued, or to mediate positive teaching and learning practices developing the students' language learning. Hence, the Libyan SECE can be considered as having a lack of context validity.

8.2.3 Cognitive validity

Cognitive validity is concerned with the test's ability to elicit cognitive processing of similar tasks in the real world. According to Weir (2005), the nature of context and cognitive validity is determined by the interaction between the demands of the task of a test and the executive resources of the candidate. In the study context, the demands of the Libyan SECE and the executive resources are the exact linguistic resources of the prescribed textbooks used in the language classroom without broadening the students' internal linguistic knowledge. This connection made the relationship between the context and cognitive validity overlap in this context more than has been described in the literature. As explained earlier in section 7.1.2., participant ELT teachers and students tended to cover a wide range of activities related to the SECE's task demands. These activities also represent the students' executive processing and resources of the SECE since the task demands and the task resources are alike in the Libyan SECE. Khalifa & Weir (2009) argued that what a task assesses is a central factor in establishing the test validity. As the Libyan SECE is a summative examination

system of content rather than assesses achievement, the students were required to recall the same task input (grammar and vocabulary) in the SECE as an assessment strategy. This assessment strategy leads the participant teachers and students to create a smaller language learning context focusing on developing the students cognitively to act effectively in the ELT assessment in the secondary school context. For that, the SECE cognitive validity is discussed in relation to developing the secondary school students' language knowledge and skills within the Libyan educational context.

Weir (2005) highlighted that the test-takers choose the most appropriate reading strategies once they had a clear idea of the required information in the text. As the participant students and ELT teachers had a clear knowledge of the SECE demands, the students' learning motivation and practice were geared towards learning discrete points of the textbooks. The participant students and teachers' classroom practices focused on mastering the prescribed textbooks rather than developing a comprehensive linguistic knowledge of English or their critical thinking and analytical skills. The teachers facilitated the students' learning of grammar, vocabulary and their recall of scientific, historical or cultural information from the reading comprehension texts of the textbooks as an executive resource of the SECE. These practices caused the adoption of cramming in teaching and learning practices. For example, the students adopted a rote-learning strategy as the proper cognitive learning practice to perform well in the test. They focused on learning grammar, translating the texts' lexis to Arabic and memorising their meanings as well as the relevant scientific information included in the final year ELT textbooks as the test executive resources. Those learning strategies were considered as the main cognitive process needed to perform well in the test. They also engaged in test preparation activities inside and outside classrooms, such as answering and revising previous versions of the public examination. In addition, the language classroom-based tests adopted the SECE format and content to train the students on how to perform in the final public examination. These activities were emphasised by the public examination policy and the teachers' and students' perceptions and experiences of its effectiveness in preparing for the Libyan public examination.

Therefore, it can be argued that the participant students were cognitively trained in the task of the SECE throughout the academic year, in order to develop their performance

in the examination sitting. Although the used learning strategies helped the students to prepare and perform on the SECE, however, this learning practice did not lead to meaningful improvement in their general language knowledge and ability, especially since the students' linguistic background knowledge was not considered in developing the SECE's executive resources. For that, it would be difficult for the students to transfer this knowledge to a real language use context. An example of this was the students' statements that they were unable to use the lexis they learnt from the school material outside the classroom context, as it is not related to the real context. Besides, the fact that the participant students understand the English language textbook material in Arabic could not develop their understanding of the English written texts in an authentic context.

Evidence accumulated through this study clearly indicates that the students' learning strategies were encouraged by the nature of the test and its design. All participant ELT practitioners and students reported that the SECE approach encouraged rote-learning of the language input (grammar and vocabulary) in words rather than the development of productive knowledge and skills (communicative competence). S7 and S6 said:

> *S6: For the final exam, we need to memorise the passage line by line because the questions of true and false will come from these texts.*
> *S7: The multiple-choice question also helps us to remember the answer, and then we answer correctly and get better marks* (Pre-examination group interview 2).

The reading comprehension information of the multiple-choice questions was described as "easy" since the choices are not identical and depend on scientific factors, especially when presenting questions without a text background. The perception that the select-response questions facilitate the cognitive process of remembering was also highlighted by the participant practitioners. According to Hughes (2003), if an examination focuses on assessing knowledge-based memorisation and the recall of information, it ignores many other qualities that are essential to student success in language learning. In this study, the SECE English examination was identified as encouraging the participant students to activate short-term memory to memorise certain items from the ELT textbooks specifically to be recalled in the examination. Accordingly, such assessment practices would not provide any effective feedback, which could help students to implement the metacognitive

strategies required for successful language learning. One of the participant students (S11) highlighted that he usually forgets the memorised linguistic aspects immediately after the examination. This comment represents the limitation of the students' language learning strategies and practices towards achieving the goal of developing their English communicative competence.

Moreover, the participant secondary school students and teachers neglected the development of English writing skills and spelling since they were not monitored in the SECE and were seen as unimportant as an executive process of the SECE. Thus, the students were also keen to develop coping strategies as a key executive process aimed at a good performance in the examination in order to get high scores rather than developing language knowledge and skills. In the post-examination interviews, S4 and S5 mentioned that they had no problem with the SECE questions items when they were able to translate the words into Arabic or when they remembered the grammatical rules required to answer the grammar questions. When students were not able to translate the words into Arabic, they used different test-taking techniques. The participant students mentioned that guessing was the most frequent coping strategy used in the SECE as they used guessing intelligently to answer the examination questions related to grammar. Guessing blindly was used to answer the true/false questions and multiple-choice questions when the students could not remember the meaning of the lexis in the questions. For example, S8 mentioned:

> *At the end of the exam time, if I can't find or get the correct answer for yes/no questions, I try to do my best by guessing according to the words where I know their meaning, or do that haphazardly and rely on my luck to get some of them right* (The post-examination group interview)*.

The participant students claimed that guessing had a significant effect on the students' SECE scores. This claim was based on their experience of this strategy in the classroom-based tests. Such claims challenge the cognitive validity of the Libyan SECE in school language education. Although guessing has been widely observed as a washback of standardised examinations in the global context (Damankesh & Babaii, 2015: Iran; Hsieh, 2017: Taiwan; Toksöz & Ertunç, 2017: Turkey; Yu & Jie, 2013: China), it is different in intensity in the case of the Libyan SECE since the exact linguistic resources had been previously known to students.

A further example of a test-taking strategy, which has been considered by a number of students in their classroom-based tests (e.g. S4, S3 and S7), was the use of the test items to answer questions related to grammar and identifying themes through words' collocations in sentences. The following extract from the pre-examination interview is an example of that:

> *S3: Sometimes I just look at any different group of words in the question and then search for these groups in the other questions. And if I find them I would try to find if there are any differences between the question and the sentence to answer the questions.*
>
> *S5: Yes, I did the same thing* (Pre-examination group interview 1).

S7 also said:

> *In some of the grammar questions, I was looking at the pattern of the question and sentences on the exam paper to see how the tense has been written and how I can form a similar sentence from the choices given in the question.* (Pre-examination group interview 2)

It is clear that the participant students' utilisation of these test-taking strategies was completely dependent on their remembering the meanings of the lexis, phrases and sentences of the examination's questions. These techniques were used to answer the classroom and periodical tests as well as the SECE. It seems that the participant students implemented some of the cognitive processes of the reading test identified by Khalifa & Weir (2009) and Weir & Khalifa (2008), such as word recognition, lexical access and establishing propositional meaning. It may be argued that the SECE encouraged the participant students to adopt a few useful test-taking strategies such as using other test items, the elimination of bad alternatives in the questions and guessing intelligently. It could be possible that these coping strategies enabled the students' cognitive abilities and mental readiness to activate their actual general knowledge in the examination. Although the students claimed the positive influence of such test-taking strategies on their scores, Hughes (2003) and Weir (2005) mentioned that the cognitive processing involved in determining an answer in this format would likely bear little resemblance to the way texts are processed for information in real life. In this case, the SECE can be perceived as having a negative influence on the

development of the participant secondary school students' language knowledge and ability.

According to this data, it can be inferred that the SECE focus and design did not elicit a cognitive process similar to those that could be implemented in the authentic context of the language use. The secondary school students' cognitive strategy in language learning was strictly limited to the memorising of the school textbook lexis context and developing test-taking strategies. These practices reflect the limitation of students' cognitive processes implemented to perform in the SECE. These findings indicate that the SECE does not fulfil the criteria of cognitive validity.

8.2.4 Scoring validity

Scoring validity is concerned with the test score's reflection of the actual level of proficiency in the language knowledge and use. The data analysis showed that there was an agreement between the participants' ELT practitioners and students' perceptions of different elements of the scoring validity aspects of the SECE. As mentioned in chapter four of this thesis, Khalifa & Weir (2009) and Weir (2005) highlighted that the scoring process of tests is usually influenced by different factors, such as the test type (objective or subjective) and the method of scoring (manually or electronically). In this study, the sufficiency in marking the select-response items of the SECE using the digital method was believed to increase the overall test stability. Since the examinations' answer sheets were electronically marked, this method was considered as a stable and consistent method of marking. The students acknowledged their satisfaction with that since it is free from marking errors or mistakes. Thus, all participants were satisfied with the marker reliability of the SECE. The ELT practitioners identified that as one of the main goals of introducing the public examination system in the Libyan education system. However, the ELT practitioners and students revealed the perception that the SECE score only reflects the students' performance in the exam but not their actual English language knowledge of ability. Different factors related to the exam design and administrations have been identified to threaten other aspects of scoring validity.

In terms of the score reflection of the students' abilities through the analysis of the examination items in the perspective of ease and difficulty, students mentioned that

the final exam of the SECE was not difficult, and the time was enough to answer the questions. In the post-examination interviews, S5 explained that:

> *The final exam wasn't that difficult. We faced lots of the final exam questions in classroom tests, or we came through them in previous versions of SECE exam for the recent years* (Post-examination interviews).

From the above quotation, the repetition of the examination questions over several versions in previous years, and the extensive preparation using these examination versions in the ELT classroom throughout the academic year was a significant factor in the ease of the SECE items. Hence, the SECE score can be considered as a reflection of students' improvement in a repetitive examination activity with the same characteristics. That made it difficult for the students' test scores to indicate their authentic language knowledge and abilities. The following extract from the students' interviews can also be considered as a representation of the actual performance indicated by SECE's score. In the extract, it should be noted that the students used the word "*understanding*" to refer to the actual cognitive process responsible for the ability of language use:

> *S9: I don't think the public examination score reflects our language knowledge and ability; because these types of questions make me memorise not understand, what I did is copy paste, nothing out of my head. It is for what students memorised from the curriculum that they get marks.*
> *S7: It reflects our level in what we memorised from the curriculum.*
> *S10: We memorise at home and then come to the exam to make copy paste.*
> *S8: It reflects memorising more than understanding.* (Pre-examination group interview 2)

The above comment indicates that the SECE measures the students' access to the school curriculum rather than their skill in a particular area. Such characteristics can also be linked to the internal consistency of the scoring validity. For practitioners, the SECE is inadequate for assessing students' learning outcomes and competence regarding the integrated language skills in the curriculum. As the focus of the ELT curriculum is on developing the students' language communicative competence, the public examination system does not provide the chance for students to express their productive Language knowledge and academic skills comprehensively. Thus, the

Libyan SECE constitute a technical, cognitive process required by the assessment policy (teaching tested parts of the curriculum and preparing students for the public examination), and whether the teachers and students achieved these policies. Consequently, the focus on testing specific linguistic knowledge (grammar and lexical) in the Libyan SECE cannot be considered as a sign of the students' abilities to use this knowledge in a real-life context. Hence, P1 did not only consider the Libyan SECE as a false indicator of the students' language abilities but also as a destructive activity towards achieving the goals of developing the school students' communicative competence in English language.

In addition, all participants of this study considered the select-response items as an inadequate measure of real language knowledge and ability as it does not provide comprehensive details or reference to their knowledge language and ability. That is because there are different factors that could increase the candidates' scores, apart from developing their language knowledge, such as guessing and cheating in the study context. For instance, in the post-examination interview, the students confirmed that there were several questions that they answered haphazardly on the test, and they thought that they got some of them right, which increased their marks on the test results. In the pre-examination interviews, students also said:

> *S3: Yes, the public examination design doesn't reflect our language ability. It is just an exam score because sometimes we answer haphazardly to questions and it comes right and I got marks on that, but in reality, I don't know the meaning of the answers to the questions.*
>
> *S2: Yes, and in that way, anyone can get high marks if they are lucky. If I got a lower mark in the school examination that doesn't mean that I don't know how to speak English.* (Pre-examination group interview 1)

In another group interview, S14 said:

> *The score of the public examination doesn't reflect our language ability because it doesn't show our level in writing since we only read and mark questions. So here [the SECE design] away from our writing ability [spelling and composition] in Arabic or English, everyone has a chance to succeed and move from one stage to another even if they don't know how to write.* (Pre-examination group interview 3)

In the above quotations, the students indicated the effect of guessing and test-taking strategies on scoring in the public examination design. They described it as a matter of luck, similar to the perception highlighted by the ELT practitioners. Weir (2005) argued that guessing might have a considerable but unknowable effect on test scores in such standardised tests. In the study context, such a character seems to have a negative implication on the learning motivation of the secondary school students. S14, for instance, said that it would be difficult to distinguish the hard-working students from others using this examination design. Therefore, he considered the SECE and classroom tests scores as meaningless in reflecting their English language knowledge and ability. Similarly, Hadri, (2015) stated that the Tunisian secondary school students held negative perception towards the inadequacy of the school assessment for providing a detailed report about their language performance and ability.

Moreover, cheating was also believed to have a significant direct effect on the students' scores in the public examination in this study. O'Sullivan & Weir (2011), in the re-conceptualisation of the socio-cognitive framework, highlighted that the context validity could directly impact on the scoring validity. That could reflect the practices of cheating, where the ready answers given to students would increase the students' scores, undermine their cognitive process. Consequently, this would also affect the factor of error of measurement within the score validity, which is concerned with the difference between the observed score and the corresponding true score of proficiency or knowledge. The study found that the secondary school-leavers English knowledge and ability is not satisfactory for either students or practitioners at different educational levels. The level of the participant students' language ability was identified as being below the objectives of the prescribed English language textbook for the final-year secondary school students. In the post-examination group interview, the participant students (S1, S11 and S5) admitted their weaknesses in English language and in their abilities to use English, although they passed the SECE with high scores. They declared that their level of English was not good enough to be able to use English for writing or for a long conversation. In the interview with the English language inspection unit (P1), he revealed that the ELT inspectors always get comments from educational practitioners in university education emphasising the weak level of the secondary school-leavers' knowledge in English language and general academic skills. He declared that it is natural to describe the secondary school-leavers as weak since the

judgement was based on the students' language skills and usage, which had never been considered in the general secondary school education. At the end of the interview, he called for alignment between the ELT curriculum and examination system in order to enhance the secondary school students' motivation in language learning and develop their language competence. The representative of the examination office (P2) also acknowledged the weakness of the school-leavers in subject knowledge and general academic skills. Whereas he argued that the lack of the subject knowledge is the responsibility of the teachers and inspectors, he traced back the weakness in the academic skills to the assessment policy of the early years of basic education. He stated:

> *The quality of subject knowledge learning outcomes is the inspectors' responsibility in the first place because they should supervisor the classroom and identify the students' weakness and advise teachers to focus on these points to develop their teaching practices and help students to overcomes difficulties lead to the weaknesses in their learning. However, the weakness of school leavers level of general academic skills, especially spelling and composition is not the responsibility of secondary school teachers. These weaknesses can be traced back to the transferring policy applied in early years of basic education. This assessment policy stated that pupils should move to the next school level even though their level is weak. When pupils pass the fourth level, teachers focus on teaching the scientific aspects of the curriculum, but not developing these skills at that level. Then, pupils keep moving from one level to another with a low level of academic skills, as they are not tested. For that, secondary school teachers should not be blamed for the problem.*

This comment reflects the neglect of assessing the development of the academic skills in the Libyan school education. Accordingly, practitioners, such the representative of the examination office, considered teaching the content of the public examination as important more than developing English language skills in the final year of secondary school. This washback seems to have a significant effect on the students' ability in the application of the information and knowledge learnt at school, especially in the English language subject. The comment also reflects the power of the assessment policy and practice on the practitioners' perceptions about their role in school education.

8.2.5 Consequential validity

This section discusses the study findings in relation to the extent to whether the intended goals of the implementation of the reformed SECE system in the educational context have been achieved to identify its washback. It will also discuss the unintended effect of the SECE's use and design in the educational context.

It was highlighted by the participant practitioners that the SECE system was introduced to avoid imperfections in the previous public examination approach from the perspective of its administration and process. For instance, the practitioners argued that the select-response items had developed the SECE's marking criteria and process. It eased the process of marking and publishing the results, especially for those who take the re-sit of public examinations. The practitioners believed that this objective of the reformed public examination system had been achieved. These achievements were considered as being of a great value to the practitioners and the Ministry of Education due to their experience of the complexity of these process and their concerns around the ethical issues of marking in the previous public examination design. The practitioners also believed that the Ministry of Education considered the increase in the number of secondary school-leavers, passing at the first exam sitting, as an indication of the efficiency of using this public examination system in the Libyan education context. Another aim of the reformed SECE design was to eliminate the cheating phenomenon in the public examinations. Since the cheating phenomenon was linked with complicated contextual factors, the examination design did not wholly eradicate cheating practices in public examination in the on-going conflict context of Libya. However, one of the participants (T1) highlighted that the examination technique was useful in eliminating some cheating in the normal situation such as the classroom-based tests.

Nevertheless, the findings from the study indicated that the SECE policy and practices had adversely affected the English language teaching and learning practices in the secondary school education. For instance, the lack of alignment between the focus of the English language curriculum planning and the public examination focus created tension among the language educators' perceptions and practices. In particular, the traditional focus of the English language public examinations yielded the perception that developing English language proficiency is not the aim of language education in

Libyan secondary school context. That is because the SECE has never assessed the English language skills of secondary education students, despite the integration of the communicative approach and the recent reforming of the public examination approach. Besides, due to the power of the public examination policy on the ELT practitioners' careers, the language educational practice in the final year of secondary school was driven by requirements of the public examination policy rather than the instructional practice required by the curriculum. Consequently, the examination policy was identified as the most significant factor affecting the English language school teachers' judgements of the suitability of their teaching and assessment strategies, such as *teaching to the test*, and the use of teacher-centred approach in their classrooms, rather than implementing CLT. The lack of alignment between the English language curriculum and its allocated time was also identified as a significant factor that affected the final-year secondary school teachers' and students' choices of their teaching and learning strategies and practices. The secondary school English language teachers taught to the test and used GTM in order to be able to fulfil the requirements of the public examination policy. The findings of the study indicated that the teaching and learning focus of the final year of secondary school is a washback of the traditional public examination policy and the reformed public design. However, the reformed public examination system influenced the intensity of this washback on perceptions and practices.

The participant students also held the perception that the school language education does not aim to develop their English communication ability. This perception emerged primarily because of the underutilisation of communicative activities in teaching practices in the English language classrooms, where the focus was purely on the SECE policy and practice. These practices influenced the secondary school students' adoption of measurement-driven learning strategies rather than the focus on developing real-life language knowledge and skills in their English education. Moreover, the participant students' negative perceptions about the English language curriculum content and the motivational strategies presented in the education policy significantly affected their adoption of *learning to the test* as a learning strategy in order to succeed in the SECE. The focus of the secondary school-leavers was on improving their score rather than developing their language ability.

As a result, those language teaching and learning practices were identified as having a positive effect on the participant students' score improvements in the classroom and the SECE but not on their English language learning outcomes. Therefore, the students felt that they had not got the chance to develop their English language ability during their secondary school education in this study. For that, their level of English language level and ability is relatively weak.

8.3   Chapter summary

This chapter examined the SECE's score validity and how the language related educational practice of the candidates of the SECE contributed to their examination scores and language learning outcomes according to the socio-cognitive framework for test validity.

In this chapter, I argued that the participant school-leavers' scores in English language in the SECE were lacking in scoring validity. Different factors were highlighted as threatening the SECE's scoring, context, and cognitive validities. From the perspective of the test-taker characteristics, the nature and value of the SECE's scores, the students' negative perceptions and attitudes towards English language education in the school context and towards the English language prescribed textbooks diminished their interest to engage in real English learning practice. Therefore, they were only motivated to study for the examination instead of developing language communicative competence. In addition, the students' experiences of preparing for and undertaking similar examinations of the BECE was the most significant factor that affected the students' psychological characteristics and guided their cognitive learning process and practices. In contextual validity, the lack of alignment between the focus of the English language curriculum and public examinations was a significant factor that threatened the context validity of the SECE. The focus of the SECE affected the participant students' beliefs that the examination score accounts for more than language proficiency in the Libyan education system. Therefore, the students engaged in test-preparation activities instead of developing their language competence. In addition, different features of the SECE's use and design (e.g. test familiarity, knowledge of response method, time constraint, the nature of the information and the content knowledge, and cheating) had a direct role in developing the students' scores in the SECE. It also encouraged the students to develop test-taking strategies (e.g. guessing

blindly or intelligently, using other test items and eliminating bad alternatives) which increased the SECE score but not their language competence. Cheating on public examinations was also identified as having a significant effect on the students' score on the SECE and threatening its score validity (discussed in section 8.2.2.3).

All the above contextual and test design facto

s, as well as students' characteristics, affected the test's cognitive validity. In their cognitive process, the participant students focused on developing the examination's executive resources rather than developing English language knowledge or skills. They focused on the translation of the lexis of the textbook and the memorisation of their meanings in Arabic as the main cognitive process needed for the test rather than developing a comprehensive linguistic competence of English language. Moreover, the students adopted memorisation of the input (curriculum) in words rather than activating the productive knowledge and skills as the suitable cognitive learning practice. The motive for that practice was the fact that the key linguistic resources needed for the SECE are the exact linguistic lexis and structure available in the school textbook. Hence, the participant students activated short-term memory to memorise the tested items of the curriculum for the examination. As a result, all the above language learning strategies and practices had a positive effect on the students' score improvements in the SECE but not on their English language learning outcomes or developing general academic skills.

Evidence accumulated through this study clearly indicates that participants' experience of the Libyan public examination format and content had a significant effect on their attitudes, perceptions and their choice of teaching and learning strategies and context. The findings of the study represent important insights for the socio-cognitive framework for test validity. This will be discussed in the next chapter of the research conclusion in the part of the contribution of the knowledge, section 9.3.3.

9 Chapter Nine: conclusion

Introduction

The conclusion of this thesis includes a brief revisit to the complete research study and its findings. In particular, the chapter summarises the research aims and design in connection with the key findings according to the research questions. It also discusses the study's contribution to knowledge, limitations, implications and suggestion for further research.

9.1 Summary of the study

This interpretative study investigated washback of the English language public examination of the Libyan SECE within secondary school education. It included washback on perceptions and attitudes, classroom teaching and learning strategies, and learning motivation and outcomes. A case study research strategy was used to conduct the study in a secondary school in the south-west of Libya (see 5.2 and 5.3). Weir's (2005) socio-cognitive framework for test validity was used as a conceptual framework to inform the research process of data generation and analysis (see 4.1 and 4.2). The participants of this study were two English language teachers and their inspectors, 15 final year secondary school students, a representative of the regional Examination Office of the Ministry of Education and the school head teacher (see 5.4). I utilised interviews as the primary source of data (see 5.5.1). One-to-one semi-structured interviews were conducted with ELT teachers and the practitioners while preparing students for the public examination. Three group interviews were conducted with three groups of students before undertaking the public examination and one interview with some of them after they undertook the examination. Three classroom observations for each teacher were employed to generate data about classroom practice for the study (see 5.5.2). In addition, document analysis and a research diary were also used as support to triangulate the data-generation tools and data sources for the study (see 5.5.3 and 5.5.4). Thematic analysis was employed to answer the research questions (see 5.8).

Retrieving the research questions from chapter one (see 1.3), this chapter discusses the main findings according to the research questions.

9.2 Study findings

The study research questions were as follows:

1) To what extent does the Libyan public examination of the SECE have an impact on motivation, perceptions and attitudes towards English language teaching within secondary school education?

2) To what extent does the public examination of the SECE influence English language teaching, assessment and learning strategies in secondary school classrooms in the on-going conflict context of Libya?

3) How does the Libyan public examination of the SECE influence the students' English learning outcomes within secondary school education?

9.2.1 Washback on perceptions and attitudes

The key finding of the study revealed that the lack of alignment in the planning of English language education created conflicting perceptions and attitudes among the participant ELT teachers and inspectors towards the SECE approach and the goal of ELT in schools within the education system (see 6.1.1). All participants held a positive attitude towards the public examination system in the SECE since it increased the efficiency of the SECE grading and publication of candidates' results and undermined cheating (see 6.2.1). These research findings emphasise the significance of improvement in administration processes in washback on the participant teachers' perceptions and attitudes towards examinations, as discussed in a MENA context or other developing countries reported by Taqizadeh & Birjandi (2015). In this study, that was more significant than has been reported due to the on-going Libyan conflict situation where education quality assurance is challenging for institutions and individuals. The participant teachers' attitudes in this research are different from those reported by Onaiba (2014) as the participants' positive perceptions were affected by the improved scores and by the satisfaction of the Libyan Ministry of Education with the increased number of school-leavers passing the SECE at the first sitting (see 6.2.1).

However, the participant ELT inspectors considered the SECE as unsuitable for assessing students' acquisition of English knowledge and skills from the school ELT textbooks. The neglect of assessing the language skills after curriculum reform in the 2000s in the public examinations produced perceptions among the participants that

developing language skills was not the real aim of teaching English in secondary school education (see 6.1.1). The traditional focus of the English language public examination on assessing grammar and vocabulary produced perceptions among the participants that secondary school students are required to promote secondary school students' knowledge of grammar and vocabulary included in the prescribed textbooks. Accordingly, this study suggests that the participant ELT teachers' perceptions about the traditional teaching approaches were influenced by the focus of public examination rather than being related to institutional and personal factors, as highlighted by the majority of research studies in the Libyan context (Alhodiry, 2016; Altaieb, 2013; Dalala, 2014; Hammadi, 2013; Omar, 2014; Rahimi & Alavi, 2017; Soliman, 2013) (see 1.1 and 7.1.1). The lack of significant attention paid to English language skills and use in the public examination and school classroom practices adversely affected perception among the students towards the value of English language education in school contexts. They perceived that learning English for communication is completely different from learning English at school (see 6.1.3). In addition, the participant ELT teachers and inspectors also considered the public examination approach as ineffective because it did not assess students' writing ability and skills, which were seen as important in language learning and acquisition (see 6.2.2).

Another significant finding was that the nature of the SECE, as a norm-referenced examination, and its score use in awarding scholarships and access to tertiary education within the Libyan education system, adversely affected students' motivation for developing English language competence in the school context. That motivated the participant students to study English to get high scores for their certificates rather than to develop their language ability. This study indicated that the excessive focus on increasing the SECE scores in the Libyan secondary school education system hindered the development of multiple language competencies and academic skills (see 6.1.4). Another important emerging theme was the effect of teachers and students' perceptions and attitudes towards the English language textbooks on the intensity of washback. This finding reinforces the importance of considering washback by textbooks highlighted by Bailey & Masuhara (2012).

9.2.2 Washback of English language teaching and assessment strategies in secondary school education in Libya

There were three findings that I considered as significant in this regard. Firstly, the lack of alignment between the focus of the English SECE and the objectives of the curriculum implied a conflict in the participant ELT teachers' and practitioners' perceptions towards their role within the language education practices (see 6.1.2). Whereas teachers were officially required to teach the curriculum communicatively, in practice, their teaching focused on the expected content of the public examination, especially as teaching time was not sufficient to complete the curriculum in the allocated time. This led to pressure on the participant teachers, because of the tension between pedagogical and ethical decisions as Spratt (2005) described. However, since the public examination and students' success had a powerful impact on teachers and inspectors' professional careers in the Libyan context, teaching the curriculum communicatively became less important for the participants than finishing the anticipated tested aspects from the curriculum in the context. Therefore, the participant teachers neglected CLT since it was ineffective in fulfilling their duty as final year secondary school teachers according to the SECE policy which is finishing the textbooks by whatever pedagogical mean (see 7.1.1). In addition, as the SECE was not designed to measure the curriculum outcomes associated with the development of the students' communication ability, the participant ELT teachers continued using their traditional teaching approaches with the focus on teaching grammar and vocabulary. This study suggests that the participant teachers' use of the traditional teacher-centred approach and a GTM approach in classrooms was deep-rooted in the traditional focus of the English language public examination. The challenges of disciplining students in such an on-going conflict context also encouraged the teachers to use these approaches to keep the classroom busy all the time to avoid conflict between students due to the socio-political tension in the context (see 7.3). This teaching strategy undermined any opportunity for extended classroom discussion and activities, transforming English language classrooms into a lecture environment and diminished all possibility of an integrated student-centred approach and the teachers' responses to students' needs in their English learning in the study context. In addition, the SECE approach led to the neglect of the activities of developing secondary school students' spelling and writing composition as writing was not considered as necessary

content or as a relevant response to the new public examination questions (see 7.1.2). This finding is incompatible with Cheng & Watanabe (2004) and Watanabe (1996) who argued that the effect of the test content and use is insignificant in determining the teaching methodology a teacher employs. In this study, the teachers' teaching approach and classroom activities were significantly affected by the test format and content (see 7.3).

Secondly, the teachers used the public examination approach in their classroom testing to secure their jobs as the assessment policy of final year secondary school students compelled teachers to use the public examination format, content and techniques in classroom-based testing (see 7.1.3). However, the use of select-response questions increased the value of summative examinations over other assessment criteria of the coursework towards the final score of the SECE. The teachers prioritised the use of summative tests rather than implementing the evaluation criteria for the students' coursework. This finding supports the conclusion of Safa & Jafari (2016) that the high-stakes examinations not only elicit *teaching to the test* in the context but also *testing to the test*. These findings also suggested that the participant teachers' assessment practices were rooted in their experiences rather than being rooted in their past learning experiences or levels of professional training, as Alderson et al. (2017), Brown & Bailey (2008), and Sheehan & Munro (2017) highlighted (see 7.1.3.3).

Thirdly, the unexpected finding was that the teachers also used the public examination approach to protect themselves from harm where self-protection was important given such an on-going conflict (see 7.1.3.4). The select-response questions helped the participant teachers to avoid arguments with students and their parents regarding the scores given in the classroom tests. Disputes over grading could result in violence that might threaten teachers' safety, especially since weapons are widespread in Libya and there is an absence of an effective state security force. Therefore, this study placed emphasis on the significance of the contextual challenges that the participant teachers and inspectors face during their work in an environment of on-going conflict, and the consequential nature of washback. These challenges increased the teachers' and inspectors' pressure and anxiety, as they were not considered in the agenda of school education policy.

This conclusion draws attention to the significance of the on-going conflict on teachers' educational practices. What is notable is that throughout this study the on-going conflict situation in Libya has emerged accumulatively of a significant effect on a wide range of the educational practices and in different ways.

The findings regarding washback on learning strategies and learning outcomes are summarised in one section as follows. That is to highlight how the former effect of public examination led to the later consequences.

9.2.3 Washback on students' English learning strategies and outcomes within secondary school education

The finding for this question is that the final year secondary school students' English learning strategy of *learning to the test* was closer to my expectation than other research questions. They showed classroom language-learning patterns similar to those described by Aftab et al. (2014) in Pakistan, Damankesh & Babaii (2015) in Iran, and Sukyadi & Mardiani (2011) in Indonesia, such as translating vocabulary to L1, answering and reviewing the exercises in textbooks and engaging in test-preparation activities, such as revising previous versions of the public examination (see 7.2). However, this study highlighted significant factors related to the nature and use of the SECE in the Libyan education system that have not been discussed. First, the fact that the SECE is constructed from the exact linguistic resources of the textbooks and the repetition of questions over the years encouraged the participant students to adopt rote learning of the language input rather than the application of the learning outcomes for communicative purposes (see 7.2.3). Such washback could only happen in a context where a high-stakes examination is aligned to specific discourse, such as the school textbooks in the Libyan context, or similar contexts, such as MENA countries (Kilickaya, 2016).

Second, the SECE as a norm-referenced score encouraged the participant students to develop a coping strategy to perform well in the examination to get high scores rather than developing language knowledge and skills (see 7.2.4). These strategies were guessing intelligently: the use of word collocations and sentences to identify a theme and address the test items. Although guessing has been widely observed as a washback of standardised examinations in the global context (China, Iran, Turkey and

Taiwan), it has more significance in the case of the Libyan SECE than has ever been reported since the test questions were previously mastered by the students.

In addition, the students' experiences of preparing for public examination in the BECE were also a significant factor in adopting those learning strategies (see 7.3). The students considered their success in the BECE, a similar examination to the SECE, as a certified and successful example of their accomplishment and their high level of achievement. Those who had different learning strategies that were not effective in their prior experience adopted those learning strategies as recommended by peer students and their teachers.

The final significant finding was that the participant secondary school students were largely passive in their English classroom practices because of the neglect of responding to their English learning needs in classrooms (see 7.2.1). The need to practise these activities beyond the classroom context led students to rely on commercial test-related materials, such as answer-key textbooks. These materials minimised the students' use of the prescribed textbooks as a means of English learning, which had a negative washback on students' exposure to English in the context where English is a foreign language, and undermined students' engagement in coursework activities. They used those answer materials to do homework and when participating in classroom activities which adversely affected students' language-learning development.

All the above language-learning strategies and practices had a positive effect on the participant secondary school students' score improvements in the classroom testing and SECE (see 8.2.3). Although the students successfully passed the SECE at the first sitting, they had not acquired satisfactory language knowledge and ability or led to meaningful improvement to use the language in a real context (see 8.1). Therefore, the SECE's lack of validity due to different aspects threatened its score validity such as test design and cheating in particular. The content of the SECE made the candidates' score a reflection of their improvement in a repetitive activity with the same characteristics (see 3.2.4).

To conclude the summary of findings section, this thesis revealed that the reformed public examination system achieved the washback intended by the Libyan Ministry of

Education of improving test administration and candidates' scores. However, it brought unintended negative washback on teaching and learning practices in English language education (see 8.2.5 and 8.3). The findings of this study suggest that the weaknesses in the students' English language-learning outcomes are caused by the public examination policy focus and practices within the school education system.

9.3 Contribution to knowledge

This study is important for its capacity to contribute new knowledge about the influence of the Libyan public examination policy and reformed system on English language education in the secondary school context. It will add to the academic knowledge about different factors that affect the nature of washback from under-researched context. This study is original in its topic (researching washback in the Libyan secondary school context) and its conceptual contribution of understanding washback in MENA context on the educational practices and the quality of education.

9.3.1 Language education research in the Libyan and context

This study is significant for its capacity to contribute new knowledge and debate in language education in the Libyan context. First, this thesis is the first study investigating ELT classroom practices in regard to the pedagogical and public examination policy as well as the role of ELT teachers, inspectors and other educational practitioners' roles within these policies. It is the first to argue that the lack of alignment between the objective of the English language curriculum and public examination is a significant factor inhibiting the focus on developing the school students' English communicative competence within the Libyan school context. The identified conflicting role of ELT teachers and inspectors created by the lack of alignment between the curriculum objective and public examination focus is a significantly contributed in the misalignment between ELT teachers' and other practitioners' priorities. This finding has pushed the debate forwards about the English language school teachers' neglect of CLT in their classrooms and its implications for the Libyan school-leaving students' English language ability.

Second, this thesis is also the first study to present empirical data regarding secondary school student learning outcomes in the Libyan context. It is also the first to include students as participants in researching Libyan teachers' practices towards CLT

integrated into the English language curriculum in 2000. Although final year basic or secondary school teachers were the participants of many Libyan research studies, these studies have not discussed the influence of the traditional teachers' practices on the students' perceptions and achievements or how these achievements influenced teachers' perceptions towards their classroom teaching approach. Finally, this thesis is the first study that discusses the validity of the public examination approach within English language education and highlights the challenges of ensuring the quality of school education on the basis of classroom testing and a public examination basis in the on-going conflict context of Libya.

9.3.2 The understanding of washback

This study contributed to the debates about the complexity of the washback phenomenon from a different under-researched context of MENA as several researchers recommended in the field (Sheehan & Munro, 2017; Shohamy, 2016; Tsagari & Cheng, 2017). In particular, the fact that the change in the public examination approach did not affect the general focus of the teaching and learning strategies can contribute and create significant steps forward to researching how a test design and test use simultaneously affect washback in context. It also added to the discussion about perceptions of the validity of the standardised assessment in context. This study accumulated evidence that the standardised test was perceived as valid in comparison with the focus of assessment policy and the focus of public examination. Such findings enforce the debate regarding the possibility of considering a test design as an initial sign of test validity of promoting constructive washback in context, as Green (2007, 2013) stated, and the demand of considering how perceptions and value within the educational context are influenced or affected by test design and test use (see 3.2.4.4).

This thesis also touches on the significance of the social context on the practices of education quality assurance through investigating washback on classroom testing and public examination practice in an on-going conflict context. This study can provide helpful discussion regarding the growing debate about the validity of automated scoring and the challenges associated with human scoring raised by Xi (2017), especially in an on-going conflict socially violated context.

Another noteworthy implication is the significance of understanding the connection between teaching materials and public examination on the nature and intensity of washback, especially in educational cultures where public examinations are constructed from the teaching materials, as in many MENA countries. This research study is also among the first to take into consideration the learning outcomes from the test-takers' perspectives in washback research studies.

9.3.3 The socio-cognitive framework for test validity

This study contributed to the socio-cognitive framework for test validity in different ways. Initially, it is the first to use the framework to investigate validation of general language public examination in a school education context. The provided critical evaluation of the operationalisation of the Libyan SECE reinforces the adaptability of the framework for researching washback and developing teaching and learning with specificity for context and the test type.

Second, the finding that students' experiences of public examinations affected their cognitive and contextual learning presents a significance insight for the framework. In the socio-cognitive framework for test validity, O'Sullivan & Weir (2011) and Weir (2005) hypothesised that test-takers' characteristics have a direct influence on the context and cognitive validities as shown in Figure 3 (section 4.2). They also hypothesised that the overlap could happen only between the cognitive and the context validities and the test-taker characteristics would only have a direct impact on both validities. However, in this study, the educational and experiential characteristics of the participant secondary school students have been initially shaped by the test-use in context. The participants' experiences of preparing and undertaking a similar public examination, such as the BECE, had a significant effect on their attitudes and perceptions, and their choice of teaching and learning strategies and context. In addition, as the SECE had an established repeated question format, the participant students' knowledge and experiences of the task setting and demands, as well as executive resources and processes, led them to adopt the same test-preparation activities and mental processes to perform in the SECE. In other words, the test-takers' characteristics were influenced by the previous experience of the cognitive validity and context validity issues of the test in the case of the BECE. This finding suggests that

the interaction and the overlap could exist between all the components of the socio-cognitive framework that affects the scoring validity as shown in the following Figure.

Figure 6: The interaction between the components of the socio-cognitive framework for test validity in the research context.
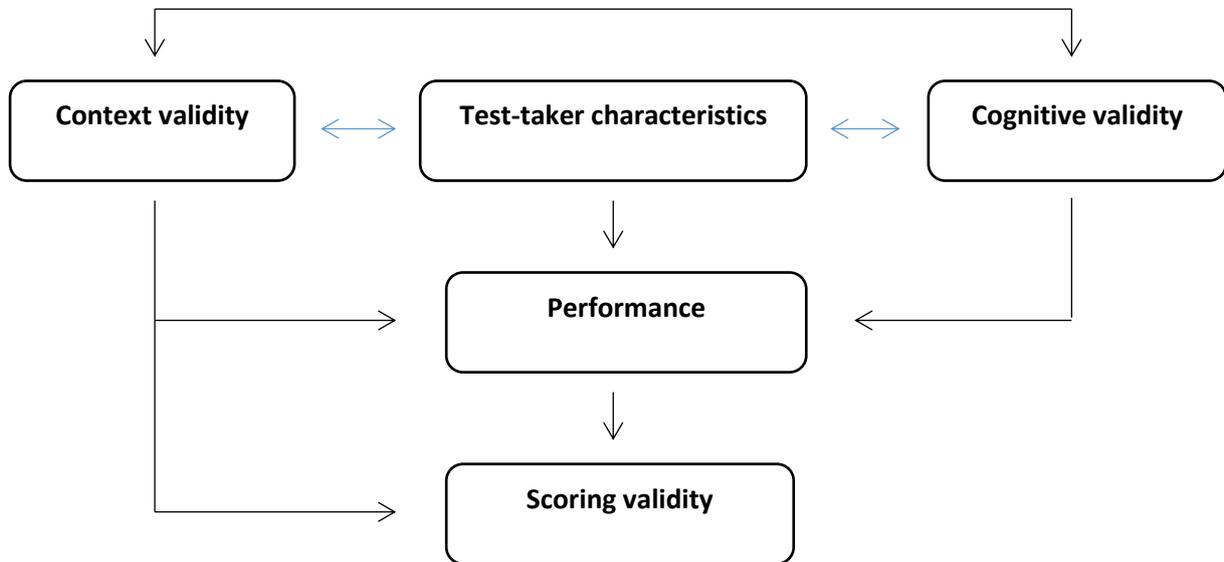
In the figure, the overlapping could happen between the test-taker characteristics, cognitive validity and context validity, especially in contexts where testing is aligned to specific textbook content, such as the SECE. This experience of the contextual and the successful cognitive process affected their learning practices, their performance in the SECE and the nature of washback on students' attitudes and perceptions towards the examination itself and the choice of the most appropriate learning behaviour and strategies.

9.4   Limitations of the study

The study was significant as it provided data about the perceptions and experiences of washback in the Libyan secondary school context. This interpretive qualitative research was informed by data triangulation and a detailed rich description of the research processes, my insider and an outsider perspective, and reflexivity to validate the data analysis (see 5.9.1.3 and 5.10). However, the study has a number of limitations that need to be acknowledged. First, the study sample was based on one school and a small number of participants, because of the crucial role of context in

constructing the educational practices and the nature of washback. Therefore, the case presented in this study may not be an accurate representation of other cohorts elsewhere in Libya. However, this study can be considered as an example of the perceptions, practices and experiences regarding the research phenomenon.

Second, since the data from this study were generated during a conflict situation in the research area, this required me to contact many people based on personal contacts to find alternative participants and get permission to facilitate access to school, and important documents related to the research phenomenon, which might affect the representation of participants in the research study. However, I was aware of these influences on the research data (see 5.9.1.2). I also took that decision since it was important to build trust with participants to be informative since I worked in a situation of on-going conflict with socio-political tension (see 5.4).

Third, my initial proposal to use a focus group for data generation was abandoned owing to the participant students' lack of understanding and experience of such discourse activity. Therefore, I adapted my strategy to group interviews instead of a focus group (see 5.5.1.2). However, this shift did not affect the generation of the relevant data as the researcher used the themes of the focus group to generate the group interview questions.

Finally, the translation of data from Arabic to English can be considered as another limitation as the process is not without drawbacks. However, a considerable amount of attention was paid to validating the translations through participant transcription validation and member-check, and I am confident that the translated data faithfully represented the meanings expressed by the participants (see 5.9.1.3).

9.5 Research implications

This research revealed the need for the Libyan educational authorities to take steps to face up to the problem in the planning of ELT within school education in order to develop the quality of English learning outcomes. In particular, the function of language education in a school context should be clear for the teachers and inspectors if it is to develop students' language ability in using English or to promote English language knowledge of vocabulary and grammar. Aligning the curriculum objectives of developing students' language skills to the focus of public examination on assessing

these skills could produce positive washback on ELT teachers' and students' perceptions and on their classroom practices and secondary school language-learning outcomes. Bailey & Masuhara (2012) argue that examinations should be matched to objectives and assist language learners to obtain and develop their knowledge and skills as part of the test.

Reforming the assessment policy within the education system can have a substantial effect on language education in particular. Language assessment should endeavour to assess the understanding and applications of language knowledge in context rather than repeating the exact exercises of the textbooks in the classroom tests and public examinations. That could enhance language-learning outcomes based on understanding and meaningful learning rather than rote-learning. This research also raised the awareness about the learning motivation imposed by the assessment policy in the rewards and requirements within Libyan education system, in which students working as score chasing engine in their language classrooms. Such policy should encourage more competitive skills development in the educational context to promote better learning outcomes.

More comprehensive support is needed for secondary school teachers and inspectors in their professional development to encourage a student-centred approach in order to deal with the pressure in teaching the curriculum. Such support should take into consideration the teachers' and inspectors' points of view rather than imposing instructions via a top-down policy implementation. School teachers and inspectors also need greater support to face up to the challenges they encounter in their classrooms, especially in the ongoing conflict of Libya. More radical and effective policies should be developed to protect teachers and other educational practitioners within their educational institutions and society.

9.6 Suggestion for further research

It has been highly recommended that collaborative research studies investigating washback of a certain test within the same educational context would build a truly comprehensive picture of the washback within a context (Cheng, 2008; Wall, 2012). Therefore, a replication of this study elsewhere in Libya would be of great value to get more insight and understanding of washback of the public examination within school

education, along with further research considering the washback effect on teachers and students at lower secondary school education levels to broaden our understanding of the scope of the effect of the teachers' perceptions and public examination policy and practice on language education in schools. It would be worth replicating this study on different school subjects (e.g. Arabic, science, maths, history, etc.) and the same populations. This could provide details regarding washback of the public examinations approach of select-response items, not only on English language education but also on general education.

Further studies should also include policymakers within the Libyan education system. Doing so would cross-reference research findings to build up a comprehensive picture of the washback operating in the Libyan secondary school context (Wall, 2012).

Although this research was the first to present empirical data regarding the secondary school student learning outcomes of the ELT, collecting more evidence regarding the students' learning outcomes would be of great benefit in the study context. Further research would benefit from collecting more evidence regarding the washback on students' learning outcomes. Future researchers could use different parameters of assessing language proficiency, such as CEFR, to observe changes in the school students' language learning development at different points of their secondary school education.

9.7 Reflection on my experience and own learning

This section presents reflections on my experience of the research programme and its effect on my own development from academic and professional perspectives.

I enjoyed doing my own research since I felt that I would be able to produce a contribution to the field with this research. I was passionate about the PGR skills courses which contributed in developing my research and academic skills that enabled me to successfully conduct this research project. I now feel much more confident in my ability to develop this area of research in my professional setting though sharing my interest and academic and research skills to other practitioners. From my engagement in the academic conferences, I became aware of the significance of such academic events to stretch the understanding of contextual and global educational development which is important for the Libyan context in particular. Libyan educators

and researchers need to organise conferences about education in order to create more opportunity for collaborative work to enhance the relationship between policy and practice in the education system.

Likewise, engaging with various literature relevant to the influence of testing on teaching and learning helped me to gain a better understanding of the fundamental role of assessment in education in general and in areas where teaching aims to develop practical skills such as language education. That has greatly enlightened me and provided me with a new lens to recognise that assessment is a holistic phenomenon to be practised inclusively within the situated educational and social contexts.

In addition, from learning about how to conceptualise the English language education within the Libyan education system, I am confident that I can take this knowledge and apply it to my career working with English language teacher education and training in Libya. This experience gave me a new identity of being a researcher besides being a teacher educator. This identity would add a lot to my practice since I have become aware of the value of research-led educational practice which would enable me to give more support to the pre-service and in-service teachers to deal with emerging educational and contextual challenges. I have also learnt the need to encourage the Libyan English language teachers to have co-operative role with students in order to overcome serious challenges in language education within the education system such as the insufficient time allocated for teaching. Finally, I have also learnt that it is important to support the Libyan English language teachers to gain an adequate level of language assessment literacy in our training programmes in order to give them the opportunity to choose to employ assessment strategies that is compatible to their context as assessment is situated practices.

This thesis has significantly stretched our knowledge and understanding of different factors and challenges that the Libyan school teachers encounter, not only the pressure they experience from the dilemma resulting from the lack of alignment between the objectives between teaching and assessment but also other contextual variables which has not been acknowledged in the Libyan context. This thesis has raised the awareness of the demand for updating our thinking and practice regarding teachers' education and training in the context of Libyan education.

10 References

Abidin, M. J. Z., Pour-Mohammadi, M., & Alzwari, H. (2012). EFL students' attitudes towards learning English language: The case of Libyan secondary school students. *Asian social science*, *8*(2), 119.

Abusrewel, F. (2014). *An investigation of the experiences of newly graduated English Language Teachers (ELT) in their first years in Libyan schools: a case study in post-conflict Tripoli* (Doctoral dissertation, University of Huddersfield).

Adnan, U., & Mahmood, M. A. (2014). Impact of Public Examination on Teaching of English: A Washback Perspective. *Journal of Education and Practice*, *5*(2), 132-139.

Aftab, A., Qureshi, S., & William, I. (2014). Investigating the washback effect of the Pakistani Intermediate English Examination. *International Journal of English and Literature*, *5*(7), 149-154.

Ahmed, A. (2010). *From curriculum development to classroom practice: a case study of Libyan context.*(MA dissertation, University of Birmingham).

Ainley, M. (2006). Connecting with learning: Motivation, affect and cognition in interest processes. *Educational Psychology Review*, *18*(4), 391-405.

Ajideh, P., & Mahmoudi, L. (2017). Iranian Pre-university Students' Perceptions towards the English Section of the Iranian National University Entrance Exam (INUEE): A Washback Study. *Journal of English Language Teaching and Learning*, *9*(20), 25-48.

Alderson, J. C. (2004). Foreword, In L. Cheng, Y. Watanabe, & Curtis, A. (Eds) *Washback in language testing: Research contexts and methods*, Mahwah, NJ: Lawrence Erlbaum Associates, ix-xii.

Alderson, J. C., Brunfaut, T., & Harding, L. (2017). Bridging assessment and learning: a view from second and foreign language assessment. *Assessment in Education: Principles, Policy & Practice*, *24*(3), 379-387.

Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of Washback. *Language Testing, 13*(3), 280-297.

Alderson, J. C., & Wall, D. (1993). Does Washback exist? *Applied Linguistics*,14(2),115 -129.

Aldridge, A., & Levine, K. (Eds.). (2001). *Surveying the social world*. Open University Press: Buckingham.

Al Heeti, A. G., & Brock, C. (1997). Vocational education and development: Key issues, with special reference to the Arab world. *International Journal of Educational Development*, *17*(4), 373-389.

Alhmali, J. (2007). Student attitudes in the context of the curriculum in Libyan education in middle and high schools (Doctoral dissertation, University of Glasgow).

Alhodiry, A. (2016). The Libyan EFL Teachers' Role in Developing Students' Motivation. *Procedia-Social and Behavioral Sciences*, *232*, 83-89.

Al-Issa, A., Al-Bulushi, A., & Al-Zadjali, R. (2017). Perceptions of the selection criteria of Omani English language teachers: Implications for policy reconsideration. *Journal of Language & Linguistics Studies*, *13*(2), 235-259.

Aloreibi, A., & Carey, M. D. (2017). English language teaching in Libya after Gaddafi. In *English Language Education Policy in the Middle East and North Africa* (pp. 93-114). Springer, Cham.

Altaieb, S. (2013). *Teachers' Perception of the English language Curriculum in Libyan Public Schools: An investigation and assessment of implementation process of English curriculum in Libyan public high schools.* (Doctoral Dissertation: University of Denver).

Al-Yaqoobi, W.; Al-Aradi, a. & Madera, E. (2010). Evaluating the impact of Bahrain National Examinations: implementing a research agenda, *the 36th Annual Conference of the International Association for Educational Assessment*, 22-27.

Andrews, S. (2004). Washback and Curriculum Innovation. In Cheng, L.; Watanabe, Y. with Curtis, A. (Eds.) *Washback in Language Testing: Research contexts and methods,* Mahwah, N J: Lawrence Erlbaum Associates, 37-50.

Andrews, S.; Fullilove, J.; & Wong, Y. (2002). Targeting Washback—a case-study. *System, 30*(2), 207-223.

Arksey, H. and Knight, P. (1999) *Interviewing for Social Scientists.* Sage: London.

Ary, D., Jacobs, L., Sorensen, C., & Walker, D. (2013). *Introduction to research in education.* Cengage Learning: USA

Asker, A. (2012). *Future self-guides and language learning engagement of English-major secondary school students in Libya: Understanding the interplay between possible selves and the L2 learning situation* (Doctoral dissertation, University of Birmingham).

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational researcher*, *36*(5), 258-267.

Bachman, L. F. (2015). Justifying the Use of Language Assessments: Linking Test Performance with Consequences. *JLTA Journal*, *18*, 3-22.

Bachman, L. F. (2013). Ongoing challenges in language assessment. *The companion to language assessment*. In A. J. Kunnan (Ed.), The companion to language assessment 1(3), 1586-1603.

Bachman, L. F. (2005). Building and Supporting a Case for Test Use. *Language Assessment Quarterly, 2*(1), 1-34.

Bachman, L. F., & Palmer, A. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press: England

Bachman, L. F., & Palmer, S. A. (1996). *Language testing in practice*, Oxford University Press: England

Bailey, K. M. (1999). *Washback in language testing*: Educational Testing Service Princeton, NJ.

Bailey, K. M. (1996). Working for Washback: A review of the Washback concept in language testing. *Language Testing, 13*(3), 257-279.

Bailey, K. M. & Masuhara, H. (2012). Language Testing Washback: The Role of Materials, *Applied Linguistics and Materials Development*, 303-318,

Bailey, K. M., Curtis, A., Nunan, D., & Fan, D. (2001). *Pursuing professional development: The self as source,* 63: Heinle & Heinle: Boston.

Baird, A. (2010). Beliefs and practice in teacher assessment. *Assessment in Education: Principles, Policy & Practice*, 17(1), 1–5.

Baleghizadeh, S. (2011). The Impact of Students' Training in Questioning the Author Technique on EFL Reading Comprehension. *ProcediaSocial and Behavioral Sciences*, 29, 1668-1676.

Barnes, M. (2017). Washback: Exploring what constitutes "good" teaching practices. *Journal of English for Academic Purposes, 30*, 1-12.

Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The qualitative report, 13*(4), 544-559.

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.

Bell, J. (2014). *Doing Your Research Project: A guide for first-time researchers*. McGraw-Hill Education: UK.

BERA. (2011). *Ethical Guidelines for Educational Research*.

Berg, B. L. (2009). *Qualitative research methods for the social sciences*. Allyn & Bacon: London.

Berry, V. & O'Sullivan, B. (2016). Assessment Literacy for Language Teachers. *TEA SIG PCE*, Avalible online at https://tea.iatefl.org/wp-content/uploads/2015/10/Vivien-Berry-and-Barry-OSullivan.pdf [last accessed 13/03/2018]

Biggs, B. (1995) Assumptions underlying new approaches to educational assessment. Curriculum Forum, 4(2), 1-22.

Biggs J., (2002). 'Aligning the curriculum to promote good learning', constructive alignment in action: imaginative curriculum symposium: *LTSN Generic Centre*. Vol 4

Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language teaching*, *36*(2), 81-109.

Borg, S. (2001). The research journal: A tool for promoting and understanding researcher development. *Language teaching research*, *5*(2), 156-177.

Booth, D. (2012). *Exploring the washback of the TOEIC in South Korea: A sociocultural perspective on student test activity* (Doctoral dissertation, Research Space Auckland).

Boyce, C., & Neale, P. (2006). *Conducting in-depth interviews: A guide for designing and conducting in-depth interviews for evaluation input* (pp. 3-7). Pathfinder International: Watertown

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77-101.

Brown, J. D., & Bailey, K. M. (2008). Language testing courses: what are they in 2007? Language Testing 25(3), 349 – 384.

Bryman, A. (2008). Why do researchers integrate combine quantitative and qualitative research: *Advances in mixed methods research.* , 87-100.

Burrows, C. (2004). Washback in Classroom-Based Assessment: A Study of the Washback Effect in the Australian Adult Migrant English Program. In Cheng, L.; Watanabe, Y.; with Curtis, A. (Eds). *Washback in language testing: Research contexts and methods*, Mahwah, N J: Lawrence Erlbaum Associates,113-146.

Chan, Y. C. (2008). Elementary school EFL teachers' beliefs and practices of multiple assessments. *Reflections on English Language Teaching*, *7*(1), 37-62.

Chen, S., & Tsai, Y. (2012). Research on English teaching and learning: Taiwan (2004–2009). *Language Teaching*, *45*(2), 180-201.

Cheng, L. (2013). Consequences, Impact, and Washback*, The Companion to Language Assessment, 9(68),*1130-1145.

Cheng, L. (2008). Washback, impact and consequences. In Shohamy, E.; & Hornberger, N. H. (Eds.), *Encyclopedia of language and education*: Language testing and assessment (2nd ed). New York: Springer Science, 349-364.

Cheng, L. (2005). *Changing language teaching through language testing: A Washback study,* Cambridge University Press: England

Cheng, L. (2004). The Washback Effect of a Public Examination Change on Teachers' Perceptions Toward Their Classroom Teaching. In Cheng, L.; Watanabe, Y.; With Curtis. A. (Eds). *Washback in language testing: Research contexts and methods*, Mahwah, N J: Lawrence Erlbaum,147-170.

Cheng, L. (1998). Impact of a public English examination change on students' perceptions and attitudes toward their English learning. *Studies in Educational Evaluation, 24*(3), 279-301.

Cheng, L. (1997). How does Washback influence teaching? Implications for Hong Kong. *Language and education*, *11*(1), 38-54.

Cheng, L., Sun, Y., & Ma, J. (2015). Review of washback research literature within Kane's argument-based validation framework. *Language Teaching*, *48*(4), 436-470.

Cheng, L., Klinger, D., Fox, J., Doe, C., Jin, Y., & Wu, J. (2014). Motivation and test anxiety in test performance across three testing contexts: The CAEL, CET, and GEPT. *TESOL Quarterly, 48*, 300–330.

Cheng, L., & Fox, J. (2013). Review of doctoral research in language assessment in Canada (2006–2011). *Language Teaching*, *46*(4), 518-544.

Cheng, L., Andrews, S., & 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.

Cheng, L., & Qi, L. (2006). Description and examination of the national matriculation English test. *Language Assessment Quarterly: An International Journal*, *3*(1), 53-70.

Cheng, L.; & Curtis, A. (2004). A Review of the Impact of Testing on Teaching and learning. In Cheng, L.; Watanabe, Y.; with Curtis, A. (Eds). *Washback in language testing: Research contexts and methods* (Eds.), Mahwah, N J: Lawrence Erlbaum, 3-17.

Cheng, L.; & Watanabe, Y.; with Curtis, A. (2004). *Washback in Language Testing: Research Contexts and Methods*. Mahwah, N J: Lawrence Erlbaum Associates.

Chou, M. H. (2015). Impacts of the Test of English Listening Comprehension on students' English learning expectations in Taiwan. *Language, Culture and Curriculum*, *28*(2), 191-208.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th Ed): Routledge: USA

Collins, H. (2010). *Tacit and Explicit Knowledge*, University of Chicago Press: USA

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching. Assessment.* Council of Europe: Estrasburgo.

Creswell, J. W. (2012). Collecting qualitative data. *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research: Pearson*, 204-35.

Creswell, J. W. (2007). *Qualitative inquiry & research design: choosing among five approaches*, Sage: London

Creswell, J. W. (2005). *Educational Research: Planning. Conducting, and Evaluating Quantitative and Quantitative and Qualitative Research, Enhanced Pearson eText with Loose-Leaf Version -- Access Card Package* (5th Edition): Pearson: USA

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed method approaches* (2nd ed.): Sage: London

Dalala, J. (2014). *Investigating the use of the self-assessment processes by Libyan EFL secondary school teachers in assessing students' written work* (Doctoral dissertation, University of Sunderland).

Damankesh, M., & Babaii, E. (2015). The washback effect of Iranian high school final examinations on students' test-taking and test-preparation strategies. *Studies in educational evaluation*, *45*, 62-69.

Davies, A. (1997). Demands of being professional in language testing, *Language Testing 14*, 328-339.

Denscombe, M. (2010). *The good research guide: for small-scale social research projects.* Maidenhead: McGraw-Hill Education.

Denscombe, M. (2003). *The good research guide.* Maidenhead: Open University Press: England

Denzin, N., & Lincoln,Y (Ed.). (2005). *The SAGE Handbook of Qualitative Research* (3rd ed.): Sage: London

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition.* Lawrence Erlbaum Associates Publishers: USA.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative and mixed methodologies*. Oxford University Press: England

Dörnyei, Z., & Ushioda, E. (2013). *Teaching and researching: Motivation*. Routledge: New York

Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge University Press: Cambridge

Earl, L. (2003). *Assessment as learning: Using classroom assessment to maximize students learning*. Sage: USA

Elder, C., & O'Loughlin, K. (2003). Investigating the relationship between intensive English language study and band score gain on IELTS. *International English Language Testing System (IELTS) Research Reports: Volume 4*, 153-207.

Erlandson, D.; Harris, E.; Skipper, B. &, Allen, S. (1993). *Doing naturalistic inquiry, a guide to methods*. Sage: London

Ferman, I. (2004). The Washback of an EFL national oral matriculation test to teaching and learning. In Cheng, L.; Watanabe, Y.; with Curtis, A. (Eds). *Washback in Language Testing: Research Contexts and Methods*. Mahwah, N J: Lawrence Erlbaum Associates, 191-210.

Finlay, L. (2003). The reflexive journey: mapping multiple routes. In L. Finlay & B. Gough (Eds.) *Reflexivity: A practical guide for researchers in health and social sciences*. Blackwell Science: Oxford.

Finlay, L. & Gough, B. (2003). Reflexivity: *A practical guide for researchers in health and social sciences* (Eds.). Blackwell Science: Oxford.

Flick, U. (2014). *An introduction to qualitative research*. Sage: London

Fulcher, G. (2013). *Practical language testing.* Routledge: Abingdon and New York

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, *9*(2), 113-132.

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'. *The guardian Weekly, Tuesday 11th October 2011,* Learning English (4), available at http://www.guardian.co.uk/education/2011/oct/11/why-more-language-test-cheating?INTCMP=SRCH. [Last accessed 13 Mar 2018].

Fulcher, G. (2004). Deluded by artifices? The Common European Framework and harmonization, *Language Assessment Quarterly*, 1, 253–66.

Fulcher, G. (2000). "The 'communicative' legacy in language testing." System 28(4), 1-15.

Fontana, A., & Frey, J. H. (2005). *The Sage handbook of qualitative research*: Sage: London

Froetscher, D. (2013). Investigating the washback of a standardized examination on classroom assessment. Available online at https://www.uibk.ac.at/news/teasig2011/programme/froetscher_doris.pdf [last accessed 13/03/2018]

Furaidah, A.; Saukah, A., & Widiati, U. (2015). Washback of English national examination in the Indonesian context. *TEFLIN Journal*, *26*(1), 36-58.

Gay, L. R., & Airasian, P. (2003). Education research. *Competencies for Analysis and Applications*, (7th ed.). Upper Saddle River, Pearson *Education.New Jersey.*

Gebril, A. (2017). Language teachers' conceptions of assessment: an Egyptian perspective. *Teacher Development*, *21*(1), 81-100.

Gebril, A., & Brown, G. T. (2014). The effect of high-stakes examination systems on teacher beliefs: Egyptian Teachers' Conceptions of Assessment. *Assessment in Education: Principles, Policy & Practice*, *21*(1), 16-33.

Gebril, A., & Eid, M. (2017). Test Preparation Beliefs and Practices in a High-Stakes Context: A Teacher's Perspective. *Language Assessment Quarterly*, *14*(4), 360-379.

Geranpayeh, A., & Taylor, L. (Eds.). (2013). *Examining Listening: Research and practice in assessing second language listening* (Vol. 35). Cambridge University Press: England

Ghorbani, M. R. (2008). The washback effect of the university entrance examination on Iranian English teachers' curricular planning and instruction. *The Iranian EFL Journal*, *2*, 60-87.

Given, L. M. (2008). *The Sage encyclopaedia of qualitative research methods*: Sage

Gomm, R., Hammersley, M., & Foster, P. (2000). Case study and generalization. *Case study method*, 98-115.

Gosa, C. M. C. (2004). *Investigating washback: a case study using student diaries* (Doctoral dissertation, University of Lancaster).

Gray, D. (2014) *Doing research in the real world* (3rd Ed). Sage: London

Gray, D., (2004). *Doing Research in the Real World: Qualitative And quantitative Approaches* (Edited). Sage: London

Green, A. (2013). Washback in language assessment. *International Journal of English Studies, 13*(2), 39-51.

Green, A. (2007). Washback to learning outcomes: a comparative study of IELTS preparation and university pre-sessional language courses. *Assessment in Education, 14*(1), 75-97.

Grix, J. (2004). *The foundations of research*. Palgrave Macmillan: London

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *ECTJ*, 29(2), 75-91.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of qualitative research*, 2(163-194).

Guest, G., Namey, E. E., & Mitchell, M. L. (2012). *Collecting qualitative data: A field manual for applied research*. Sage: London

Guilloteaux, M, & Dörnyei, Z. (2008). Motivating language learners: A classroom-oriented investigation of the effects of motivational strategies on student motivation. *TESOL Quarterly, 42*, 55–77.

Gunn, J., Al-Bataineh, A., & Al-Rub, M. A. (2016). Teachers' Perceptions of High-Stakes Testing. *International Journal of Teaching and Education*, *4*(2), 49-62.

Hammadi, N. K. B. (2013) *An exploration of the gendering mediating students' attitudes and experiences in the formal academic curriculum at secondary school level in Libya* (Doctoral dissertation, Nottingham Trent University).

Hammersley, M. (2008). *Questioning Qualitative Inquiry Critical Essays*, Sage: London

Hammersley, M. (2007). The issue of quality in qualitative research*. International Journal of Research & Method in Education, 30*(3), 287-305.

Hammersley, M., Foster, P. & Gomm, R. (2000). *Case Study Method: key issues, key texts*, Sage: London

Hamp-Lyons, L. (1998). Ethical test preparation practice: The case of the TOEFL. *TESOL Quarterly* 33(2), 329-337.

Hamp-Lyons, L. (1997). Washback, impact and validity: Ethical concerns. *Language Testing, 14*(3), 295-303.

Harlen, W, & Deakin Crick, R. (2003). Testing and motivation for learning. *Assessment in Education, 10*, 169–207.

Hartaty, W. (2017). The impact of the 2015 national examination policy in Indonesia on English language teaching. *Proceedings of ISELT FBS Universitas Negeri Padang*, *5*, 355-358.

Hayes, B., & Read, J. (2004). IELTS test preparation in New Zealand: Preparing students for the IELTS academic module. In Cheng, L.; Watanabe, Y.; with Curtis, A. (Eds) *Washback in language testing: Research contexts and methods*. Mahwah, N. J.: Lawrence Erlbaum Associates. 97-*111*.

Henn, M., Weinstein, M. & Foard, N. (2006). *A short ıntroduction to social research*. Sage: London.

Hennink, M. M.; Bailey, A., & Hutter, I. (2011). *Qualitative Research Methods*. Sage: London.

Hidri, S. (2016). Conceptions of assessment: Investigating what assessment means to secondary and university teachers. *Arab Journal of Applied Linguistics*, *1*(1), 19-43.

Hoque, M. E. (2011). *Washback of the Public Examination on Teaching and Learning English as a Foreign Language (EFL) at the Higher Secondary Level in Bangladesh* (Doctoral dissertation, Jahangirnagar University).

Hsieh, C. N. (2017). The Case of Taiwan: Perceptions of College Students About the Use of the TOEIC Tests as a Condition of Graduation. *ETS Research Report Series*.2017:1. 1-12.

Huang, S. C. (2012). Pushing learners to work through tests and marks: Motivating or demotivating? A case in a Taiwanese university. *Language Assessment Quarterly*, *9*(1), 60-77.

Hughes, A. (2003). *Testing for language teachers* (2nd ed.), Cambridge University Press: England.

Hughes, A. (1993). Backwash and TOEFL 2000. (manuscript. Reading: University of Reading).

Hughes, A. (1988). Introducing a needs-based test of English language proficiency into an English-medium university in Turkey. *Testing English for university study (ELT Document. 127), London: Modern English Publications*, *134-1*

Hyland, A. (2011). Entry to Higher Education in Ireland in the 21st Century. Discussion paper for the NCCA /HEA Seminar 2011. HEA: Dublin

Hyland, K. (2002). *Teaching and Researching Writing*: Pearson Education: England.

Kennedy, S.; & Lui, R. (2013). Washback of a high-stakes English test in China: Student and teacher perceptions. *Concordia Working Papers in Applied Linguistics*, 4(10), 22-29.

Khalifa, H., Weir, C. (2009). *Examining reading: Research and practice in assessing second language reading.* Cambridge University Press: England.

Khalifa, S. & Shabdin, A. (2016). Autonomy in Vocabulary Learning: Vocabulary Learning Strategies Teaching Programme for EFL Libyan Learners. *Arab World English Journal* 7(1). Pp. 140 – 159.

Kilickaya, F. (2016). Washback effects of a high-stakes exam on lower secondary school English teachers' practices in the classroom. *Lublin Studies in Modern Languages and Literature*, *40*(1), 116-134.

Kohn, A. (1993). *Punished by Rewards*, Houghton Mifflin Company: NewYork.

Krayem, H. (2013). *Language Learning Strategies of English Specialization Secondary School Students in Libya* (Doctoral dissertation, Universiti Pendidikan Sultan Idris).

Kvale, S. (2008). *Doing interviews.* Sage: London

Legard, R., Keegan, J., & Ward, K. (2003). In-depth interviews. *Qualitative research practice: A guide for social science students and researchers*, pp.138-169.

Li, H.; Zhong, Q.; & Suen, H. K. (2012). Students' Perceptions of the Impact of the College English Test. *Language Testing in Asia, 2*(3), 77-94.

Li, X. (1990). How powerful can a language test be? The MET in China, *Journal of Multilingual and Multicultural Development, 11*, 393-404.

Lincoln, Y.S. & Guba, E.G. (1985). *Naturalistic Inquiry*. Sage: London.

Madaus, G. & Clarke, M. (2001). The adverse impact of high stakes testing on minority students: Evidence from one hundred years of test data. In Orfield, G. & Kornhaber, M. L. (Eds.), *Raising standards or raising barriers? Inequality and high-stakes testing in public education*. The Century Foundation Press: New York. pp. 85-106.

Manjarrés, N. B. (2005). Washback of the foreign language test of the state examinations in Colombia: A case study. *The Arizona Working Papers in Second Language Acquisition and Teaching*, *12*, 1-19.

Markee, N. (1997) *Managing curricular innovation*. Cambridge University Press: England.

Marshall, C., & Rossman, G. B. (2006). *Designing qualitative research* (4th ed.). Sage: London.

Matsuda, P. K., & Silva, T. (Eds.). (2005). *Second language writing research: Perspectives on the process of knowledge construction.* Lawrence Erlbaum Associates: Mahwah, NJ.

Maxwell, J. A. (2012). *Qualitative Research Design: An Interactive Approach: An Interactive Approach.* Sage: London

Maxwell, J. A. (2005). Qualitative research design. *Applied social research methods series.*(41)

McNeill, P., & Chapman, S. (2005). *Research methods.* Routledge: London

McSweeney, K. (2014). *Assessment practices and their impact on home economics education in Ireland.* (Doctoral dissertation. The University of Stirling).

ME (Ministry of Education - Libya). (2012). 'Regulation (134) of (2012) regarding the re-organizational structure and specialisations of the Ministry of Education and its administrative department. Council of Ministers

ME (Ministry of Education - Libya). (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.

MHE (Ministry of Higher Education - Libya). (2010). 'Regulation (501) of (2010) regarding the Regulations of Higher Education": Former General People's Committee. Articles from 61 to 68.

Menken, K. (2008). High-Stakes Tests as de facto Language Education Policies. In *Encyclopedia of language and education* : Springer. 2529-2541.

Merriam, S. B. (2009). Qualitative research: A guide to design and implementation: Revised and expanded from *qualitative research and case study applications in education.* Jossey-Bass: San Franscisco.

Merriam, S. B. (1998). Qualitative Research and Case Study Applications in Education. Revised and Expanded from *Case Study Research in Education.* Jossey-Bass: San Francisco.

Messick, S. (1996). Validity and Washback in language testing. *Language testing, 13*(3), 241-256.

Messick, S. (1989). 'Validity' in R. L. Linn (ed.). Educational Measurement. Macmillan: New York.

Mewald, C. (2016). Planning for teaching and learning in EFL Effects of standardisation and standardised testing on the learning and teaching of EFL at lower secondary level. *Open Online Journal for Research and Education.*

Milanovic, M.; & Saville, N. (1996). Considering the Impact of Cambridge EFL Examinations: Manuscript Internal Report. Cambridge ESOL: England.

Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis: An expanded sourcebook (2nd ed.): Sage: London

Mohsen, A. (2014). Teaching English as a Foreign Language in Libya. *Scientific Research Journal*, Volume II, Issue XI,. 58 – 64.

Mogapi, M. (2016). Examinations Washback Effects: Challenges to the Criterion Referenced Assessment Model. *Online Submission*, *3*(3), 78-86.

Murray, J. C., Riazi, A. M., & Cross, J. L. (2012). Test candidates' attitudes and their relationship to demographic and experiential variables: The case of overseas trained teachers in NSW, Australia. *Language Testing*, *29*(4), 577-595.

Nikoopour, J. & Farsani, M.A. (2012). Depicting washback in Iranian high school classrooms: A descriptive study of EFL teachers' instructional behaviors as they relate to university entrance exam. *The Iranian EFL Journal*, 8 (2) pp. 9-34

Nunan, D. (1992). *Research methods in language learning.* Cambridge University Press: England.

Omar, S. (2013). The Relevance of Libyan Secondary School Materials to the Learners and their Needs. *Scientific Research Journal 1(5). 27 – 40.*

Omar, Y. Z. (2014). *Perceptions of selected Libyan English as a foreign language teachers regarding teaching of English in Libya* (Doctoral dissertation, University of Missouri-Columbia).

Onaiba, A. (2014). *Investigating the Washback effect of a revised EFL public examination on teachers' instructional practices, materials and curriculum* (Doctoral dissertation, University of Leicester).

Orafi, S. (2013). Effective Factors in the Implementation 0f ELT Curriculum Innovations. *Scientific Research Journal*, Volume I, Issue V, pp. 14 - 21

Orafi, S. & Borg, S. (2009). Intentions and realities in implementing communicative curriculum reform. *System, 37*(2), 243-253.

Ortlipp, M. (2008). Keeping and using reflective journals in the qualitative research process. *The qualitative report*, *13*(4), 695-705.

O'Sullivan, B. (2008). *Modelling Performance in Tests of Spoken Language*. Peter Lang: Bern

O'Sullivan, B. (2006). Issues in Testing Business English: The BEC Revision Project. *Studies in Language Testing*, 17. Cambridge University Press: England.

O'Sullivan, B. (2000). Exploring gender and oral proficiency interview performance. *System*, *28*(3), 373-386.

O'Sullivan, B., & British Council (2012). *Aptis test development approach*. Aptis technical Report, ATR-1). British Council: London, UK.

O'Sullivan, B., & Weir, C. J. (2011). Test development and validation. *Language testing: theories and practices.* Palgrave Macmillan: London.

Pan, Y. C. (2016). Traditional and Non-Traditional Test Preparation Practices: Learner Performance and Perspectives. *Electronic Journal of Foreign Language Teaching*, *13*(2). 170–183.

Pan, Y. C. (2011). Teacher washback from English certification exit requirements in Taiwan. *Asian Journal of English Language Teaching*, 21, 23-42.

Pan, Y.C. (2009). A review of Washback and its pedagogical implications*. VNU Journal of Science, Foreign Languages*, 25, 257-263.

 Pan, Y.C.; & Newfield, T. (2012). Tertiary EFL Proficiency Graduation Requirements in Taiwan: A Study of Washback on Learning. *Electronic Journal of Foreign Language Teaching, 9*(1), 108-122.

Pan, Y. C. (2014). Learner Washback Variability in Standardized Exit Tests. *Tesl-Ej, 18*(2), n2.

Pathan, M., & Marayi, Z. (2016). Teaching English as a Foreign Language in Libyan Schools: Issues and Challenges. *International Journal of English and Education, 5(2). pp. 19 - 39*

Patton, M. Q. (2005). *Qualitative research. Encyclopedia of statistics in behavioural*: USA

Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3 ed.). Sage: London

Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing*, 20, 26-56.

Pizarro, M. (2010). Exploring the Washback Effects of a High-Stakes English Test on the Teaching of English in Spanish Upper Secondary Schools, *Revista Alicantina de Estudios Ingleses* 23 (2010): 149-170.

Pour-Mohammadi, M. & Zainol Abidin, M.J. (2011). Attitudes towards teaching and learning test-taking strategies for reading comprehension tests: The case of Iranian EFL undergraduates. *Journal of Studies in Education*, 1, pp. 1-12

Punch, K. F. (2013). *Introduction to social research: Quantitative and qualitative approaches.* Sage: London.

Qi, L. (2005). Stakeholders' conflicting aims undermine the Washback function of a high-stake test. *Language Testing, 22*(2), 142-173.

Qi, L. (2004). Has a high-stakes test produced the intended changes? In Cheng, L.; Watanabe, Y.; with Curtis, A. (Eds). *Washback in language testing: Research contexts and methods*, Mahwah, N J: Lawrence Erlbaum, 171-190.

Rabe, M. (2003). Revisiting insiders and outsiders as social researchers. *African Sociology Review, 7*(2), 149-161.

Rahimi, F., Esfandiari, M. R., & Amini, M. (2016). An Overview of Studies Conducted on Washback, Impact and Validity. *Studies in Literature and Language*, *13*(4), 6-14.

Rahimi, M., & Alavi, J. (2017). The role of teaching experience in language teachers' perceptions of a top-down curriculum change. *The Curriculum Journal*, *28*(4), 479-503.

Rahimi, Z., & Nazhand, N. (2010). The washback effect of IELTS preparation courses to the learners: Iranian learners' perspectives on IELTS preparation courses. In *International Conference on e-Education, e-Business, e-Management, and e-Learning (*IEEE*).* 490-494.

Read, J.; & Hayes, B. (2003). The impact of the IELTS test on preparation for academic study in New Zealand. *IELTS Research Reports*, 5, 153-206.

Rea-Dickins, P.; & Scott, C. (2007). Washback from language tests on teaching, learning and policy: evidence from diverse settings. *Assessment in Education*, *14* (1), 1-7.

Regmi, K., Naidoo,Jennie .,& Pilkington,Paul (2010). Understanding the Processes of Translation and Transliteration in Qualitative Research. *International Journal of Qualitative Methods, 9*(1), pp.16-26.

Rhema, A., & Miliszewska, I. (2012). The potential of e-learning in assisting post-crisis countries in re-building their higher education systems: the case of Libya. *Issues in informing Science and information technology, volume 9*, 149-160.

Ritchie, J. & Lewis, J. (2003). Qualitative Research Practice. Sage: London

Robson, C. (2011). *Real World Research* (3 ed.). Wiley: UK

Robson, C. (2002). *The analysis of qualitative data*. Blackwell: London

Rose, M. (2015). *Education in North Africa since Independence*, British Council

Rowley, J. (2002). Using case studies in research. *Management Research News*, 25(1), pp.16-27

Rubin, H. J., & Rubin, I. (2005). *Qualitative interviewing: the art of hearing data*: Sage: London.

Ryan, J. (2011). Teaching and learning for international students: towards transcultural approach. *Journal of Teachers and Teaching: Theory and Practice*, *17*(6), 631-648.

Safa, M. A., & Jafari, F. (2016). The Washback Effects of High School Third Grade Exam on EFL Teachers' Methodology, Evaluation and Attitude. *Applied Research on English Language*, *5*(1), 77-98.

Saukah, A. (2015). *National Examination in Indonesia and Its Washback Effects.* In Developing Indigenous Models of English Language Teaching and Assessment. (Eds) Hamied, F.A., Yadnya, I.B. & Sosiowati,.A.G.. Udayana University Press. 143-160.

Saville, N. (2009). *Developing a model for investigating the impact of language assessment within educational contexts by a public examination provider.* (Doctoral dissertation, University of Bedfordshire).

Saville, N., & Hawkey, R. (2004). The IELTS impact study: Investigating washback on teaching materials. *Washback in language testing: Research contexts and methods*, 73-96.

Schwandt, T. A. (2003). Three epistemological stances for qualitative inquiry: Interpretativism, hermeneutics and social constructionism. In Denzin, N. and Lincoln, Y (Eds.), *The Landscape of Qualitative Research: Theories and issues.* (pp. 292-331). Thousand Oaks, CA: Sage.

Scott, C. (2007). Stakeholder perceptions of test impact. *Assessment in Education*, *14*(1), 27-49.

Seidman, I. (2006*). Interviewing as Qualitative Research: A Guide for Researchers in Education and the Social Sciences*. Teachers College Press: New York.

Shakuna, K.; Mohamad, N., & Ali, A. (2016). The Effect of School Administration and Educational Supervision on Teachers teaching performance: Training Programs as a Mediator Variable. *Asian Social Science*, *12*(10), 257.

Shaw, S. D., & Weir, C. J. (2007). *Examining writing: Research and practice in assessing second language writing* (Vol. 26). Cambridge University Press: England.

Sheehan, S. & Munro, S. (2017). Assessment: attitude, practices, needs. In: BAAL Language Learning and Teaching SIG 2017, University of Central Lancashire: Preston.

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for information, 22*(2), 63-75.

Shih, C.-M. (2010). The Washback of the General English Proficiency Test on university policies: A Taiwan case study. *Language Assessment Quarterly, 7*(3), 234-254.

Shih, C.-M. (2009). How Tests Change Teaching: A Model for Reference. *English Teaching: Practice and Critique, 8*(2), 188-206.

Shih, C.-M. (2007). A new Washback model of students' learning. *Canadian Modern Language Review/La Revue canadienne des langues vivantes, 64*(1), 135-161.

Shohamy, E. (2016). Critical language testing. In *Language testing and assessment* (pp. 1-15). Springer International Publishing: UK

Shohamy, E. (2013). The discourse of language testing as a tool for shaping national, global, and transnational identities. *Language and Intercultural Communication, 13*(2), 225-236.

Shohamy, E. (2011). Assessing multilingual competencies: Adopting construct valid assessment policies. *The Modern Language Journal, 95*(3), 418-429.

Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*, Pearson Educational: UK.

Shohamy, E. (1993). *The power of test: The impact of language testing on teaching and learning.* National Foreign Language Centre Occasional Papers: Washington

Shohamy, E.; Donitsa-Schmidt, S.; & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing, 13*(3), 298-317.

Silverman, D. (2010). *Qualitative research* (Ed.). Sage: London.

Silverman, D. (2001). *Interpreting qualitative data: Methods for analysing talk, text and interaction* (2nd ed.). Sage: London.

Silverman, D. (2000). Analysing talk and text. *Handbook of qualitative research*, 2, pp. 821-834.

Simons, H. (2009). *Case study research in practice.* Sage: London

Smyth, E., & Banks, J. (2012). High stakes testing and student perspectives on teaching and learning in the Republic of Ireland. *Educational Assessment, Evaluation and Accountability, 24*(4), 283–306.

Soliman, E. M. (2013). Libyan teachers' attitudes and believes regarding the use of EFL authentic materials within reading lessons at Universities levels in Libya. *International Journal of Learning and Development*, *3*(5), 121-129.

Spann, P. (2015). The Negative Effects of High-Stakes Testing. Available online at: http://www.luc.edu/media/lucedu/law/centers/childlaw/childed/pdfs/2015studentpapers/Spann.pdf [Last accessed 13/03/2018].

Spolsky, B. (1994). Policy Issues in Testing and Evaluation. *The Annals of the American Academy of Political and Social Science, 532*, 226-237.

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.

Stake, R. E. (2008). Qualitative case studies. In N. K. Denzin, & Lincoln, Y. S. (Eds.), *Strategies of qualitative inquiry* (pp. 119-149): Sage: USA

Stake, R. E. (2005). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 443– 466). Sage: USA

Stan, E (2012). The Role of Grades in Motivating Students to Learn in Romania, *Social and Behavioral Sciences* 69 (2012), 1998 – 2003.

Stoneman, B. W. H. (2006). *The impact of an exit English test on Hong Kong undergraduates: a study investigating the effects of test status on students' test preparation behaviours* (Doctoral dissertation, The Hong Kong Polytechnic University).

Sukyadi, D., & Mardiani, R. (2011). The washback effect of the English National Examination (ENE) on English teachers' classroom teaching and students' learning. *K@ta*, *13*(1), 96-111.

Taqizadeh, M., & Birjandi, P. (2015). What Drives High School English Teachers to Teach the Way They do? An Investigation of the Washback Effect of the University Entrance Examination in Iran. *Journal of Applied Linguistics and Language Research*, *2*(2), 8-20.

Tayeb, Y.; Abd Aziz, M.; Ismail, K. & Khan, A. (2014). The Washback Effect of the General Secondary English Examination (GSEE) on Teaching and Learning in Yamen, *Journal of Language Studies* 83 *Volume 14(3), pp.83-103.*

Taylor, L. (2006). The changing landscape of English: implications for language assessment, *ELT Journal, 60*(1), 51-60.

Taylor, L. (2005). Washback and impact. *ELT Journal, 59*(2), 154-155.

Thaidan, R. (2015). Washback in language testing. *Education Journal*, *4*(1), 5-8.

Thomas, G. (2011). A typology for the case study in social science following a review of definition, discourse, and structure. *Qualitative inquiry*, *17*(6), 511-521.

Toksoz, I., & Kilickaya, F. (2018). Review of Journal Articles on Washback in Language Testing in Turkey (2010-2017). *Lublin Studies in Modern Languages and Literature*, *41*(2), 184.

Toksöz, S., & Ertunç, A. (2017). Item Analysis of a Multiple-Choice Exam. *Advances in Language and Literary Studies*, *8*(6), 141-146.

Tongco, M. D. (2007). Purposive Sampling as a Tool for Informant Selection. *A Journal of Plant, People and Applied Research Ethnobotany Research and Applications*, 1-12.

Tsagari, D. (2017). Washback of a high-stakes English exam on teachers' perceptions and practices. *Selected papers on theoretical and applied linguistics*, *19*, 431-445.

Tsagari, D. (2009). *The complexity of test Washback. An empirical study.* Peter Lang: New York.

Tsagari, D. (2007). Review of Washback in Language Testing: What Has Been Done? What More Needs Doing?. *Online submission*. Available online at https://files.eric.ed.gov/fulltext/ED497709.pdf [Last Accessed 12/03/2018].

Tsagari D., Cheng L. (2017). Washback, Impact, and Consequences Revisited. In: Shohamy E., Or I., May S. (eds) *Language Testing and Assessment. Encyclopedia of Language and Education* (3rd ed.). Springer: Cham

Turner, C. E. (2006). Professionalism and high-stakes tests: Teachers' perspectives when dealing with educational change introduced through provincial exams. *TESL Canada Journal*, *23*(2), 54-76.

Volante, L., & Fazio, X. (2007). Exploring teachers candidates' assessment literacy: implication for teachers education reform and professional development. *Canadian Journal of Education*, *30*(3), 749-770.

Wall, D. (2012). Washback. In Fulcher, G. & Davidson, F. (Eds.), *The Routledge handbook of language testing*: Routledge: London. 79–92.

Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory* (Vol. 22). Cambridge University Press: UK

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. (1999). *The Impact of High-Stakes Examinations on Classroom Teaching: A Case Study Using Insights from Testing and Innovation Theory*. Department of Linguistics and Modern English Language: Lancaster University.

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.; & Horak, T. (2007). Using baseline studies in the investigation of test impact. *Assessment in Education, 14*(1), 99–116.

Wall, D., & 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 (TOEFL Monograph Series, MS-34). Princeton, NJ: ETS.

Wall, D., & Alderson, J. C. (1993). Examining washback: the Sri Lankan impact study. *Language testing*, *10*(1), 41-69.

Wang, J. (2010). *A Study of the Role of the'teacher Factor'in Washback* (Doctoral dissertation, McGill University Library).

Wang, J. (2008). Identifying the main impact on teacher classroom behaviour: Examining the effects of the CET, teacher beliefs, teacher knowledge bases and teacher proficiency. In *Language Testing and Research Colloquium*.

Wang, X. (2017). A Chinese EFL Teacher's Classroom Assessment Practices. *Language Assessment Quarterly*, *14*(4), 312-327.

Watanabe, Y. (2004). Methodology in Washback Studies, In Cheng, L.; Watanabe, Y.; with Curtis, A. (Eds) *Washback in language testing: Research contexts and methods*. Mahwah, N. J.: Lawrence Erlbaum Associates. 19-36.

Watanabe, Y. (2001). 'Does the university entrance examination motivate learners? A case study of learner interviews', in Akita Association of English Studies (ed.), *Trans-equator exchanges: A Collection of Academic Papers in Honour of Professor David Ingram,* Author, Adita, Japan, 100-110.

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. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing, 13,* 318-333.

Weili, W. (2010). Investigating the washback effect of the New CET 4 Listening Comprehension Subtest on language learners, Northwest Agriculture & Forestry University. *Chinese Journal of Applied Linguistics*, 33 (9), 28-39.

Weir, C. (2005) *Language Testing and Validation: An Evidence-based Approach.* Palgrave Macmillan: Basinogstoke.

Weir C. J. & O'Sullivan, B. (2011). Test development and validation. In B. O'Sullivan (ed.), *Language Testing: Theories and Practices.* Palgrave: Basingstoke.

Weir, C., & Khalifa, H. (2008). A cognitive processing approach toward defining reading comprehension. *Cambridge ESOL: Research Notes*, (31), 2-11.

Weir, C. & Shaw, S. (2006). Defining the constructs underpinning main suite writing tests: a socio-cognitive perspective. Research Notes, 26(3), 9-14.

Weir, C.; & Shaw, S. (2005). Establishing the validity of Cambridge ESOL writing tests: towards the implementation of a socio-cognitive model for test validation. *University of Cambridge ESOL Examinations Research Notes*,21, 10-14.

Weir, C. J.; & Milanovic, M. (2003). *Continuity and innovation: revising the Cambridge Proficiency in English Examination,* Cambridge University Press: England.

Wu, Z. (2014). Washback effect of the reformed CET-4 college English teaching and learning in China: Students' Perspectives. (Doctoral dissertation, Indiana University of Pennsylvania).

Xie Q. (2011). Is test takers' perception of assessment demand related to construct validity? *International Journal of Testing*, 11(4), 324–348.

Xie, Q. & Andrews, S. (2013). "Do test design and uses influence test preparation? Testing a model of Washback with Structural Equation Modeling." *Language Testing, 30(1).* 49-70.

Yang, H.C. (2013). *The Project of Washback effects on EFL teachers' beliefs and practices: When Grammatical testing meets Communicative testing*, (Doctoral dissertation, University of Stirling).

Yin, R. K. (2013). *Case study research: Design and method* (5th Ed.): Sage: London

Yin, R. K. (2011). *Applications of case study research.* Sage: London

Yin, R. K. (2009). Case Study Research: Design and Methods. Essential guide to qualitative methods in organizational research (Vol. 5). In *The Information Systems Research Challenge (Harvard Business School Research Colloquium). Sage: London.*

Yu, W., & Jie, B. (2013). Washback Effect on Second Language Acquisition Test: A Case Study of College English Entrance Exam. *International Academic Workshop on Social Science*, 864-868.

Zhang, S. (2016). Washback Effects of CET4 on College English Learning. In *Proceedings of The Fifth Northeast Asia International Symposium on Language, Literature and Translation* (pp. 779-784).

Zhan, Y., & Andrews, S. (2014). Washback effects from a high-stakes examination on out-of-class English learning: Insights from possible self-theories. *Assessment in Education: Principles, policy & practice*, *21*(1), 71-89.

Zhao, C. (2016). Motivational effects of standardized language assessment on Chinese young learners. *Cogent Education*, *3*(1), 1-11.

Zhao, X., Selman, R. L., & Haste, H. (2015). Academic stress in Chinese schools and a proposed preventive intervention program. *Cogent Education, 2*, 1–14.

11 Appendices

11.1 Appendix 1: Outlines and objectives of English language textbook in final year secondary school

Course Summary

	Reading Two lessons	Vocabulary Three lessons	Grammar
Unit 1 Puzzles and mysteries	The mystery of the Nazca Lines (Predicting content).	Certainty and uncertainty.	Subject and object questions. Talking about the past with *must, may, might* and *can't*.
Unit 2 Weather and climate	Hot and cold (Taking notes).	Adjective + preposition.	Adjectives with *so*, *enough* and *too*. Order of adjectives.
Unit 3 Facts and figures	Just a minute! (Scanning for specific information).	*Until, by* and future time phrases.	The future perfect and the future continuous. The infinitive with future meaning.
Unit 4 Great failures	Great failures (Reading to retell information).	Verb collocations.	How things could be different. The future in the past.
Unit 5 Literature	Two novel extracts (Identifying styles of writing).	Nouns and adjectives ending with *–ing*.	Adjectives, noun and question words followed by the infinitive. *–ing* or infinitive?
Unit 6 The world of sport	Fair play? (Identifying topic sentences).	Connecting words.	Verbs for reporting speech. Time phrases and questions in reported speech.
Unit 7 Health and first aid	The World Health Organization (Reading for specific information).	The body and first aid.	The passive – review. The passive – continuous tenses and *have* + object + past participle.
Unit 8 English in the world	English in the world (Understanding gist).	Review.	Review – sentence patterns. Review – the passive and conditionals.

	Speaking	Writing	Listening	Specialization
	One lesson	One lesson	One lesson	Four lessons
Unit 1 Puzzles and mysteries	Solving puzzles and responding to suggestions.	Presenting different points of view.	Listening for key information.	Solids, liquids and gases Heating and cooling Shapes Shapes, puzzles and mysteries
Unit 2 Weather and climate	Telling a news story.	Writing a news article.	Listening to a weather forecast.	Sinking and floating Icebergs Submarines and airships Weather science
Unit 3 Facts and figures	Giving advice.	Leaflets giving advice.	Listening for specific details and contrastive stress.	Interpreting data Pie charts and data Statistics Computer English
Unit 4 Great failures	Telling a story from pictures.	Writing a story.	Listening to complete notes.	Telephones Large numbers Inventions Safety in the lab
Unit 5 Literature	Talking about books.	A book review.	Listening for detail and consonant clusters.	Famous books in science The father of science fiction Science fiction novels Classification of plants
Unit 6 The world of sport	Exchanging information.	Longer sentences.	Functions of a conversation.	Lasers Scientific claims Compact discs Sports injuries
Unit 7 Health and first aid	Giving instructions.	Instructions.	Understanding information and instructions.	Malaria Distribution of disease Industrial processes What's happening?
Unit 8 English in the world	Giving opinions and comparing English with Arabic.	Comparing and contrasting.	Predicting content and listening for gist.	Scientific English A dictionary of science What I enjoyed A science quiz

246

11.2 Appendix 2: Samples of the SECE questions and answer sheet

Sample of the SECE questions

دولة ليبيا

دولة ليبيا
الحكومة الليبية
وزارة التربية والتعليم — إدارة الامتحانات

شهادة إتمام مرحلة التعليم الثانوي
ورقة أسئلة امتحان مادة

اللغة الإنجليزية

الدور الأول – للعام الدراسي 2015 - 2016

اسم الطالب:	
المنطقة:	رمز الطباعة: 04/20160622235141/120793
اللجنة:	
المدرسة:	
القسم: القسم العلمي	
المادة: اللغة الإنجليزية	
الزمن: ثلاث ساعات	
رقم الجلوس: 740892 رمز الامتحان: 26103820 تسلسل: 7/4	

إرشادات وتعليمات هامة

• على الطالب أن يتأكد من أن اسمه ورقم جلوسه مطبوعان من المنظومة على كل من ورقة الأسئلة وورقة الإجابة وأنهما فعلاً مخصصتان له.

• إذا استخدم الطالب ورقة إجابة أو ورقة أسئلة مخصصة لطالب غيره فلن تصحح إجابته وسيرسب في الامتحان.

اسم الطالب:

• على الطالب أن يستخدم قلم الرصاص من نوع HB فقط لكي يستطيع تغيير الإجابة إذا أخطأ وذلك بمسح الإجابة الخاطئة واستبدالها بالإجابة الصحيحة.

• في حالة استخدام الطالب قلم الحبر الجاف الأزرق أو الأسود لن يستطيع المسح ولا تغيير أية إجابة.

• في حالة استخدام الطالب قلم رصاص غير نوع HB أو قلم حبر غير قلم الحبر الجاف الأسود أو الأزرق لن تصحح ورقة إجابته.

• في كل سؤال، عليك أن تظلل الدائرة الخاصة بالإجابة الصحيحة بالكامل كما هو مبين في الشكل التالي، وعدم الاكتفاء بوضع علامة داخل الدائرة أو رسم دائرة أو مربع حولها أو وضع علامة أخرى خلاف التظليل الكامل للدائرة.

• جميع الأشكال التالية تمثل طرقاً غير صحيحة لتحديد الإجابة وتسبب في احتساب الاجابة خاطئة بالكامل:

• إذا أردت تغيير الإجابة بعد تظليلها فتأكد من مسحها أولاً ثم ظلل الإجابة الصحيحة.

تابع في خلف هذه الصفحة ...

إرشادات حول الإجابة عن أسئلة الصواب والخطأ والاختيار من متعدد والمزاوجة

- تتم الاجابة عن هذه الأسئلة في الواجهة الأمامية لنموذج الإجابة، وذلك باختيار إحدى الإجابات المتاحة وتظليل الدائرة الخاصة بها والموضحة برقم السؤال.

- الأسئلة من نوع الصواب والخطأ، يتم تظليل الدائرة (أ) عندما تكون الجملة صحيحة، وتظليل الدائرة (ب) عندما تكون الجملة خاطئة.

- الأسئلة من نوع الاختيار من متعدد، يتم تظليل الدائرة الخاصة بالإجابة الصحيحة، مع العلم بأن لكل سؤال إجابة صحيحة واحدة فقط.

- الأسئلة من نوع المزاوجة، عليك أن تختار ما يتوافق بين العمودين.

إرشادات حول الإجابة عن الأسئلة المقالية

- تتم الإجابة عن الأسئلة المقالية خلف ورقة الإجابة.

إرشادات هامة أخرى

- تمنع الكتابة أو وضع أية علامات في أجزاء ورقة الاجابة الموضحة بعبارة ((يرجى عدم الكتابة في هذه المنطقة))، كما تمنع الكتابة أو وضع أية علامات في هوامش ورقة الإجابة. أية مخالفة لهذه التعليمات قد تتسبب في عدم تصحيح ورقتك. وللأهمية نكرر عدم الكتابة أو وضع أية علامات بين أو على أو بالقرب من العلامات والإشارات التالية، وذلك لتأثيرها السلبي على عملية التصحيح الآلي:
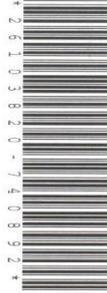
مع التمنيات للجميع بالتوفيق والنجاح إن شاء الله

أسئلة الصواب أو الخطأ:

Q 1) I asked him where his car was . It is a reported speech of " where is your car ? "

A) True

B) False

Q 2) Smallpox can permanently damage the skin and the eyes .

A) True

B) False

Q 3) Icebergs form mainly in the summer .

A) True

B) False

Q 4) Coal is an example of a non-renewable resource .

A) True

B) False

Q 5) He can't have been the murderer . He was out of town .

A) True

B) False

Q 6) Climate and weather mean the same .

A) True

B) False

Q 7) The air conditioning makes the air so cold that people get ill .The sentence is grammatically correct .

A) True

B) False

Q 8) Khalid is probably going to university next year . This sentence expresses certainty .

A) True

B) False

Q 9) She is going to have her tooth filled . The sentence is passive .

A) True

B) False

Q 10) To drive fast is dangerous .

A) True

B) False

Q 11) Most stories are written in the present .

A) True

B) False

249

Q 12) We make subject questions with do or did .

     A) True
     B) False

Q 13) Who you spoke to on the phone?

     A) True
     B) False

Q 14) In newspaper headlines the articles " a , an . the " and the auxiliaries " have and be " are usually left out .

     A) True
     B) False

Q 15) Mosquitoes are mainly found in hot countries .

     A) True
     B) False

Q 16) A psychologist is someone who studies the way human society works .

     A) True
     B) False

Q 17) A paragraph doesn't usually begin with a topic sentence .

     A) True
     B) False

Q 18) In formal English we sometimes use be + infinitive to talk about future arrangements .

     A) True
     B) False

Q 19) When solids are cooled they melt .

     A) True
     B) False

Q 20) Chess is very relaxed hobby .

     A) True
     B) False

Q 21) A fever is an illness causing a low temperature .

     A) True
     B) False

Q 22) Most of the volume of an iceberg lies below the surface of the water.

     A) True
     B) False

Q 23) I wish I had studied hard last year .

     A) True
     B) False

Q 24) " heavy , density , weight , volume "

The odd one out is volume .

A) True

B) False

Q 25) When liquids are cooled, they boil.

A) True

B) False

أسئلة الاختيار من متعدد:

Q 26) .... two of our best players were injured , we won the game.

A) Because

B) Although

C) However

D) So

Q 27) A .... is a four - sided figure in which only one pair of sides is parallel .

A) rhombus

B) trapezium

C) rectangle

D) parallelogram

Q 28) If he .... asleep , he wouldn't have missed the train .

A) was falling

B) hadn't fallen

C) fell

D) falls

Q 29) This time tomorrow , I .... to Garian .

A) drive

B) will be driving

C) will have driven

D) will drive

Q 30) He .... a better job if he improves his English .

A) would get

B) will get

C) is getting

D) got

Q 31) Can you give me something .... this terrible headache ?

A) for to stop

B) stopping

C) to stop

D) stopped

251

Q 32) By tomorrow , I .... a month writing this story .

    A) will spend

    B) should spend

    C) will be spending

    D) will have spent

Q 33) I will never forget you . The sentence expresses ....

    A) an offer

    B) a suggestion

    C) an obligation

    D) a promise

Q 34) Her hair .... by the new hairdresser now .

    A) is cutting

    B) is being cut

    C) will be cut

    D) cuts

Q 35) Has your daughter been .... against measles ?

    A) infected

    B) eradicated

    C) suffered

    D) vaccinated

Q 36) He was accused .... in the exam .

    A) for cheating

    B) cheated

    C) of cheating

    D) to cheat

Q 37) That text message wasn't recieved by my mother. This sentence means ....

    A) The text message didn't recieve my mother

    B) My mother didn't recieve that text message

    C) My mother will recieve that text message

    D) My mother recieved that text message

Q 38) Which gas is most commonly used in airships?

    A) nitrogen

    B) oxygen

    C) hydrogen

    D) helium

Q 39) This boy .... be your brother , you look very similar.

    A) mustn't

    B) can't

    C) must

    D) might

Q 40) Tennis has to be played on a court, ... volleyball can be played anywhere .

    A) so

    B) on the other hand

    C) whereas

    D) although

Q 41) We .... for five hours and we haven't had a break .

    A) worked

    B) have been working

    C) had worked

    D) had been working

Q 42) Yesterday Ali .... do some work , but he decided to do it later .

    A) goes to

    B) is going to

    C) would go to

    D) was going to

Q 43) " fog - snow - cloud - mist - smog " the odd one out is ....

    A) mist

    B) fog

    C) smog and cloud

    D) snow

Q 44) What would you wish for if you .... three wishes ?

    A) had

    B) will have

    C) have

    D) had had

Q 45) It's .... hot that you can't go out .

    A) enough

    B) too

    C) so

    D) very

Q 46) " I'm sorry I'm late . "

    A) Fatima told not to be late

    B) Fatima said she is late

    C) Fatima apologized for being late

    D) Fatima promised not to be late

Q 47) Charles Dickens wrote this letter . To make this sentence passive ....

    A) This letter was written by Charles Dickens

    B) This letter wrote Charles Dickens

    C) This letter has been written by Charles Dickens

    D) Charles Dickens has written this letter

253

Q 48) " How is your father ? " Ali asked Mohamed ....

   A) how was his father?

   B) how is his father

   C) how was his father doing?

   D) how his father was

Q 49) .... too fast is dangerous .

   A) Driving

   B) Drive

   C) Driven

   D) To drive

Q 50) He'll have arrived .... 5 p.m.

   A) by

   B) on

   C) until

   D) then

أسئلة المزاوجة (عليك ان تختار من جُمل العمود الأيسر ما يتناسب مع جُمل العمود الأيمن) :

Q 51)	Doses	B	A)	stopping a process developing
Q 52)	Ceremonies	C	B)	amounts of drug or a medicine that should be taken
Q 53)	Suppression	A	C)	formal social occasions
Q 54)	Setbacks	E	D)	information in the form of numbers
Q 55)	Statistics	D	E)	problems that cause a delay

Q 56)	afraid	E	A)	in something
Q 57)	worried	D	B)	from school
Q 58)	interested	A	C)	for something or someone
Q 59)	responsible	C	D)	about something or someone
Q 60)	absent	B	E)	of something or someone

***** انتهت الأسئلة *****

254

Sample of the SECE's questions and answer sheet

The source: Onaiba (2014).

11.3 Appendix 3: Table of the recent washback research studies discussed in this research

The author	Year of publication	Exam researched	Context	Participants	Methods of data generation
Zhan & Andrews	2014	CET-4	China	Students	Diary and interviews
Kennedy & Lui	2013	BMET	China	Students and teachers	Questionnaire and interviews
Xie & Andrews	2013	CET	China	Students	questionnaire
Yu & Jie	2013	CEEE	China	Students	Questionnaires and interviews
Xie	2011	CET	China	Students	Questionnaire
Weili	2010	CET- 4	China	Learners	Questionnaire
Hsieh	2017	TOEIC	Taiwan	Students	Survey and phone interviews
Pan	2016		Taiwan	Taiwanese university students	Interviews
Chou	2015	TELC	Taiwan	students	Questionnaire
Pan	2014	GEPT and TOEIC	Taiwan	Learners	Questionnaires and scores analysis
Pan & Newfield	2012	CSEPT	Taiwan	Students	questionnaire and interview
Huang	2012	classroom assessment	Taiwan	Teachers and students	Questionnaire and interview
Cheng et. al.	2011	(HKCEE)	Hong Kong	Students and parents	Survey
Adnan & Mahmood	2014	HSSC	Pakistan	Teachers	Questionnaire

Aftab et. al.	2014	HSSC	Pakistan	Teachers and students	Interviews
Hoque	2016	HSC	Bangladesh	Teachers and students	Questionnaire, Interviews and observation
Barnes	2017	TOEFL	Vietnam	Teachers	Interview and observation
Hartaty	2017	The public assessment policy	Indonesia	Teachers	Questionnaire
Furaidah et. al.	2015	ENE	Indonesia	Teachers	Interviews and observation
Sukyadiand & Mardiani	2011	ENE	Indonesia	Students	Multiple-qualitative case study
Ajideh & mahmoudi	2017	INUEE	Iran	Students	Questionnaires
Safa & Jafari	2016	Public exam on learning high school	Iran	Teachers	Questionnaire
Damankesh & Babii	2015	High school final exam	Iran	Students	Observation
Nikoopour & Farsani	2012	UEE	Iran	Teachers and students	questionnaire and observation
Pour-Mohammdi & Zainol Abdin	2011		Iran	Students	Questionnaire
Rahimi & Nazahard	2010	IELTS	Iran	Students	Questionnaire
Toksoz & Ertunc	2017	MCQ at state university	Turkey	Students	Questionnaires

Gebril	2017	High-stakes examination	Egypt	Pre and in-service teachers	Questionnaire
Gebril & Eid	2017	High-stakes examination	Egypt	Teachers	Questionnaire and interviews
Gebril & Brown	2014	High-stakes examination	Egypt	Pre and in-service Teachers	Questionnaire
Hadri	2016	High-stakes examination	Tunisia	School and university teachers	Questionnaire
Tayeb et. al.	2014	GSEE	Yemen	Teachers	Interviews and questionnaire
Al-Issa et. al.	2017	IELTS	Oman	Teachers	Interviews
Mewald	2016	Standardised test at lower secondary school level	Austria	Teachers and headteachers, pupils, parents	Surveys, interviews, observation and document research
Froetscher	2013	mature	Austria	Teachers	Survey and interviews
Tsagari	2017	FCE	Greek	Teachers	Interviews
Smyth & Banks	2012	Leaving Certificate (LC)	Ireland	Students	Group interviews
Stan	2012		Romania	Teachers	Questionnaire and focus-groups
Gunn et al	2016		USA	Teachers	Survey
Murry et. al.	2012		Australia	Learners	Questionnaire
Mogapi	2016	PSLE	Botswana	Teachers	Questionnaire

11.4 Appendix 4: The interview guide themes

Interviews Themes

Themes would be explored with participants (teachers, the headmaster, inspectors, the head of examination department and the head of the curriculum development department in the region of study) in the education system and context,

1 – Assessment in general education

- Role of assessment in education
- knowledge of assessment in education

2 - Language assessment knowledge

- Role of language assessment.
- Knowledge of language assessment.
- Pre-service or post-service training/support in language assessment.
- Need of training/support in language assessment
- View of personal practices in writing tests

3 - SECE English language examination

- Characteristics (format, clarity of instruction, difficulties and students' scores)
- Advantages and disadvantages
- Perceptions of a better practice

4 - SECE English language examination effect within the context

- Effect on teachers (anxiety, motivation and teaching strategies and techniques)
- Effect on students (anxiety, motivation of learning and learning strategies)
- Effect on teaching practices (teaching contents, methodology, general classrooms and formal assessments)
- Effect on beliefs and perceptions (traditional exams vs. current examination practices)

5 - SECE English language examination and outcomes

- Effect on learning outcomes
- Scores validity
- Peoples affect outcomes

6 - SECE English language examination score use

- Knowledge about scores use
- Perceptions of good scores

- The importance of scores

The Students Interview Themes

1 - SECE English language examination

- Characteristics (format, clarity of instruction, difficulties and students' scores)
- Advantages and disadvantages
- Perceptions of a better practice

2 - SECE English language examination effect within the context

- Effect on teachers' teaching strategies and techniques
- Effect on students (anxiety, motivation of learning and learning strategies)
- Effect on teaching practices (teaching contents, methodology, general classrooms and formal assessments)
- Effect on beliefs and perceptions

3 - SECE English language examination score use

- Knowledge about scores use
- Perceptions of good scores
- The importance of scores
- Peoples affect scores

11.5 Appendix 5: The interview guide questions

Teachers interview questions

Introductory questions

- Could you please tell me about your qualification?
- How long have you been teaching?
- Have you attended any in-service training during your work?

Teachers' interview main questions:-

- Which language skills and component do you focus more on in your teaching and testing? Why?
- Do you see any kind of relationship between the Secondary Education Certificate Examination, teaching materials, teaching objectives and learning outcomes? How?
- How do you help your students to get good score in the Secondary Education Certificate Examination?
- Do you think that assessment practices have an influence on the classroom behaviours, and practices or on teaching and learning techniques and strategies?
- What do you think could be an alternative practice for assessing students learning outcomes? Why?

Interview questions for (School Headmaster - Inspectors - Heads of Examination Department - Heads of Curriculum Development)

Introductory questions

- Could you please tell me about your qualification?
- What is your experience in education?
- Have you attended any kind of staff development sessions or courses during your work with the Ministry of education?

Main interview questions:-

- What do you think about the current assessment practices in secondary school education at school context? How? Why?
- Do you see any kind of relationship between the teaching and learning techniques and strategies of language education and language assessment in the school context? How?

- Do you see any kind of relationship between Secondary Education Certificate Examination, teaching materials, classroom behaviours, and practices? How?
- What do you think could be an alternative practice for assessing students learning outcomes? Why?

interview guide questions for students

Introductory questions

- Do you like English Language learning? Why?
- Do you think that learning English Language is more useful at school or at other English language teaching institution? Why?
- How are you doing on your English study? How? Why?
- What is your main focus in your study this year in general and in learning English in particular?
- Which outcomes are you more willing to get from your study of English language at school?

Main interview questions

- Which language skills and component does your teacher focus more in teaching and testing? How? Why?
- What is your opinion in the classroom exams and Secondary Education Certificate Examination? Why?
- What does the score you get / will get in classroom exams and Secondary Education Certificate Examination means for you? How? Why?
- What do you think of your preparations for Secondary Education Certificate Examination in perspective of information, materials, skills and techniques? How? Why?
- What are the consequences of your score in English on your certificate?

11.6 Appendix 6: The observation schedule

The Impact of Testing on Language Teaching and Learning Processes and its Outcomes in the Libyan Context

Date: .... Class: ……………………………

Time: .... lesson: ……………………………

Teacher: .... observation code: ………………

	The classroom activity	Utterance by teacher	Utterance by students	Notes
Classroom-based testing related themes in teaching	Test Taking Strategy			
	Test relevant information			
	Test preparation material			
	Mock quizzes and exams			
SECE related themes in teaching	Test-taking strategies			
	SECE relevant information			
	Test preparation material			
	Mock SECE quizzes and exams			
SECE related themes in learning	Students ask teachers SECE related questions			
	Students discuss the SECE			

11.7 Appendix 7: The consent form

University of Huddersfield
School of Education and Professional Development

Participant Consent Form

Title: The Impact of Testing on English Language teaching and Learning Process and its' Outcomes in the Libyan Context

I am Ahmed Ahmed, a PhD student at the School of Education and Professional Development at University of Huddersfield UK. My research aimed to identify the practitioners' beliefs and perceptions of the high-stake examination of SECE and its effect on their behaviour and practices in the context. It is also anticipating to examine the variables involves. I am grateful for your participation to this research project. It is important that you read, understand and sign the consent form.

Participant Identifier Number:

☐ I have been fully informed of the nature and the aim of this research

☐ I consent to taking part in it

☐ I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason.

☐ I give permission for my responses to be quoted after been anonymised.

☐ I understand that no person other than members of the research team and facilitators will have access to the information provided.

☐ I understand that my identity will be protected by the use of pseudonym in the report and that no written information that could lead to me being identified will be included in any report

☐ I agree to take part in the above study

If you are satisfied that you understand the information and are happy to take part in this project, please put a tick in the box aligned to each sentence and print and sign below.

Name of Participant:	Name of Researcher: ....
Signature of Participant:	Signature of Researcher: ....
Date:	Date: ....

(One copy to be retained by Participant / one copy to be retained by Researchers)

11.8 Appendix 8: The Official letters for research site access

University of HUDDERSFIELD

11 February 2015

Our Ref: IEO_AUTHABS/1367197/1

To whom it may concern

Mr Ahmed Ahmed, Student ID: 1367197, Date of Birth: 25/Sep/1981

Please note that the visa rules do not allow a student to officially extend a course end date if the student has been given a period of authorised absence from their course for the same reason. If you feel that you will not be able to finish your course on time you are advised to consider requesting a suspension instead. Alternatively if you are unable to finish on time you may be eligible to take a lower award

This is to confirm that Mr Ahmed Ahmed is enrolled on the PHD (Education) FT (Course code DX505) at the University of Huddersfield.

I have authorised a period of absence to be taken between the 2 March 2015 and the 1 May 2015 for the following reason:

Study visit/data collection outside the UK

Ahmed will be expected to complete his/her studies by 31/Mar/2018.

I trust that this information is of help.

Yours faithfully

Sharon Baines
International Compliance Officer
International Office
University of Huddersfield
Queensgate
Huddersfield
HD1 3DH

Tel: 01484 372048
Email: sharon.baines@hud.ac.uk

Please note that the visa rules do not allow a student to officially extend a course end date if the student has been given a period of authorised absence from their course for the same reason. If you feel that you will not be able to finish your course on time you are advised to consider requesting a suspension instead. Alternatively if you are unable to finish on time you may be eligible to take a lower award.

التاريخ : 11-02-2015	
الرقم الاشاري : 2015-02-13756	
رقم الملف : 5907	اسم الطالب : احمد ابوبكر مسعود احمد
رقم قرار الإيفاد: 293/2013	الدرجة العلمية : دكتوراة
بداية الصرف : 01/06/2014	التخصص : لغة انجليزية
نهاية الصرف : 31/05/2017	
اشهر المنحة : 36	

السيد / مدير إدارة البعثات الدراسية

بعد التحية ،،،

الموضوع / دراسة حقلية

تقدم إلينا المعني بطلب لموافقتكم على إجراء دراسة حقلية في ليبيا لجمع بعض البيانات والمعلومات المتعلقة بدراسته، وذلك حسب رسالة الأستاذ المشرف المرفقة.

الرجاء موافاتنا بالخصوص.

شاكرين حسن تعاونكم معنا.

والسلام عليكم ورحمة الله وبركاته ،،،

د. أحمد عبدالله الأويك
الملحق الثقافي المكلف

Printed On: 11/02/2015
Printed By: A5274

61-62 Ennismore Gardens, London, SW7 1NH T: +44(0)20 3006 9891 F: +44(0)20 7584 6961 W: culturalaffairs.libyanembassy.org

الحكومة الليبية المؤقتة
وزارة التربية والتعليم
شؤون التربية والتعليم

السادة مدراء المدراس الآتية

يطلب منكم التعاون مع السيد / احمد ابوبكر مسعود احمد

لإجراء عملية البحث العلمي المتعلق بالتأثير

الامتحانات الشهادة الثانوية علي طرق التدريس بالمدارس

"والسلام وعليكم ورحمة الله وبركاته"

صورة الصادر العام
صورة المتابعة
صورة المعني

11.9 Appendix 9: Sample of initial coding of the data

A: you talked about English as a language and subject, could you explain the difference between them?

S12: yes, there is difference, as language it is ok to talk with the colloquial language and so but as s subject there are many regulations you need to follow

A: what are these regulations?

S12: there are specific syllabi, examinations and marks etc.

A: how does that affect you or make difference for you?

S12: my focus here will depend on more than just learning English, I will think about how many classes am I going to take and how we will finish the syllabi, and how to get high marks to succeed, and here I have to say and do everything correctly to succeed, otherwise there is no benefit from learning English, but if we like it as a language and I would like to use it, to speak correctly or not is not a matter, the important thing is that the person I am taking to, be able to understand me

S13: moreover, as a subject, syllabi are not from the real life, in our syllabi they talk about Dinosaurs, ancient human being and fish, I mean there are things that we don't have in our real life context at all, but if you come to the colloquial language and speak with anyone, we find that we are going well with them in speaking English although I don't know too much grammar at school

A: Does that mean that you also like English as a language but not as a subject?

S13: yes

S12: teacher, our problem is that we can't find someone speak English with us on street, in the school or at home

S11: for me, I have been speaking English with an African man works for us in the garage, and I learnt from him many sentences in English as his language is English

A: ok! What about others?

S14: for me it is depend on the teacher

A: how it depends on the teacher

S14: it is not a matter of language and subject but it is depend on the teacher in the way he teach, it is different from one teacher to another, there is a teacher who can present and explain lesson and give information correctly in perfect way and other teachers who can't

A: is that mean, what really interest you is the teaching methodology of a teacher or the teacher personality in itself?

S14: both teacher, because the teacher personality reflects his teaching methodology and style in the class, and in teaching there is teachers who understand students' situation and help them to improve in language and there are teachers who only keen on syllabi and marks

Administrator
Perception of learning English – happy with formal but complicated in informal

Administrator
Perception of learning English – language school education attached with specific syllabi, examinations and marks
Seems to say exam are not part of leaning

Administrator
Student's focus in language education at school – more than just learning / process/ time/ syllabi/ high marks/ all with accuracy

Administrator
Student's focus in language education out of school – focus only on meaning either said it correct or not.

Administrator
Perception of learning English – authentic material effect

Administrator
Perception of learning English – Comparing syllabi context with their real life context

Administrator
Perception of learning English – level of competence in colloquial language higher

Administrator
Learning English – ability of use verses formal education

Administrator
Motivation - focus

Administrator
Learning English – effect of educational & social context

Administrator
Learning English – experience of informal learning

Administrator
Motivation – teaching and understanding

Administrator
Motivation – teachers' focus / improving language ability

Administrator
Motivation – teachers' focus / focus on finish syllabi and put marks
This shoes how much teachers' focus on examination could affect students motivation of learning English

A: I see, ok! Away from the teacher, what about the language itself or as a subject as how your colleagues classify it, do you have any interest in it or you like or want to learn it?

S14: I would like to learn it, but to some extent I haven't learnt it properly

A: So, that is means you like English

S14: as a language is better as a subject, teacher

A: why?

S14: as a subject is difficult but as a language is easy

A: ok1 have you tried to tell your teacher about this classification that you have, so that the teacher might emerge these together in the class?

S12: no we haven't, but as a subject even in designing of the book, if they brought something from our culture and customs, we will be more able to understand it than how we do now, but these customs put in the book aren't ours and we have difficulties in understanding them?

A: that is as what S13 mentioned it is a different context from ours'

S13: now, my problem is only in translating, I can read and write but I can't know the meaning, I am good in spelling as well

S11: a word doesn't last for a week before you forget it because you don't use it out of the book

S13: and there is problem in translating

S14: it is also because there is no consistency in learning English, you study it at school and after the school finish in the end of the year, that is it

A: is there anyone who tried to learn English somewhere else apart of school?

S14: yes, I tried last year, I attend a course for only two week but I didn't finish it as it was during the school academic year, and I wasn't able to cope with both of them, so I quiet

S12: the same as me

A: why did you go for the course to study English? What advantages could be available in the course but not in the school?

S14: at the course, the teacher speak English fluently as he is native speaker

A: what about others?

S15: me too, I went to study in the course but when schools started I quiet

A: the same situation then

S15: sure

Administrator
Motivation – reflects on practice / level of competence

Administrator
Motivation – depend of the focus /as a language is easy / as a subject is difficult
This refers to examination and marks and the process of learning

Administrator
Perception of learning English in context – the power of culture in learning

Administrator
Perception of learning English in context – material context and difficulties of understanding

Administrator
Learning English – issue of translation of meaning

Administrator
Learning English in context – practice / use

Administrator
Learning English in context – inconsistency on learning

Administrator
Experience of learning English out of school - yes

Administrator
Reasons of learning English out of school – teachers using English

S14: and also we go to courses because most teachers teach in the class with Arabic all the time but we used to have a teacher his Name is M in School E used to speak all the time in English....even the word that you don't know, he explain it for us in English and we never had a word in Arabic in the class, but then some circumstances happened to him and left the school and for that I left that school

S12: if the whole country can speak English, we will be ok but they don't

A: how?

S12: now if all those people living in this area speak English, we will be able to use it as I have a cousin who studied English here in Libya but he didn't get any usefulness, but when he travelled to Philippine he is 100% good now

A: And what do think will happen when he comes back?

S12: now he has been there for over than one year speaking in English, so even if he come back here he wouldn't forget it and he will sure find someone who will be able talk with him

A: have you tried to speak in English with each other in the class?

S11, S12: no

A: why?

S11: we know only simple and few words but not enough to be able to speak well

A: These simple words could be developed by using them and every day you could add new things on them, do you think so?

S13: but we don't study this in the book

A: do you think that learning English at courses could be better than schools?

S14: for me, at courses is better, especially during summer holiday of schools

Other: {at the course is better}

S12: because people there speak in English

S14: and in the school, we are limited to specific lessons but in the course with the use of English in all conversation, even if you don't understand something they will continue to explain it for you in English.....

A: what about examination system, do you think that it has an effect of studying English at school or at a course

Administrator
Experience of learning English – comparison between school teachers

Administrator
Experience of learning English – important decision related to teachers availability but not practice

Administrator
Perception of learning English in context – the power of social context in learning

Administrator
Perception of learning English in context – the power of social context in learning – relative experience

Administrator
Perceptions of vocabulary learning and using the language
That gives translating and learning words strong role in students learning process

Administrator
Perceptions of vocabulary learning and using the language – relation to the syllabi

Administrator
Perception of learning English out of school - at the course is better

Administrator
Reasons of learning English out of school – people using English

Administrator
Reasons of learning English out of school curriculum limitation

11.10 Appendix 10: Sample of certificate for evaluating Teacher of English Language Performance

Form, N9 Education

Ministry of Education
Administration of Education Inspectorate
Certificate for Evaluating
Teacher of English Language Performance

Area of Educational Inspection Office ....

School:- .... Name of Teacher:- ....

Nationality : - .... Qualification:- ....

Date of Graduation :- .... Date of Appointment: - ....

Specialization : - ....

Classes he/ she is in charge of : - ....

Number of classes per week : - ....

Date of Visit	Classes being visited	Lesson / Unit	Attendees	Absentees

Details on the visit	1st visit	2nd visit	3rd visit	4th visit	average	remarks
Intellectual performance and written lesson plan (15 marks)						
Follow up of written activities (7 marks)						
Use of audio - visual aids / materials (8 marks)						
Methods and procedures (12 marks)						
Application of syllabus and its effect on students progress a : Amount of syllabus being taught (7 marks) b : Student's competence (12 marks)						
Teacher's language ability a: Pronunciation (5 marks) b : Structure (5 marks) c : Fluency (5 marks)						
Teacher's character and conduct a: Appearance (5 marks) b : His / Her relation with students (5 marks)						
General activities related to the subject (5 marks)						
Teacher's cooperation with the inspector (9 marks)						
Total Marks 100 marks						
Average						

General remarks and recommendation :---------------------------------------
----

Grade : --
Name of Inspector : ---------------------------
Date : ---
Signature : --

Grading scheme :

Marks	Below 50	50 : 63	64: 77	78 : 89	90 : 100
Grade	Weak	Average	Good	Very Good	Excellent

Date :   /   /

Approved by head of unit

Name : ---
Signature: ---------------------------------------

11.11 Appendix 11: The booklet for English language teacher development
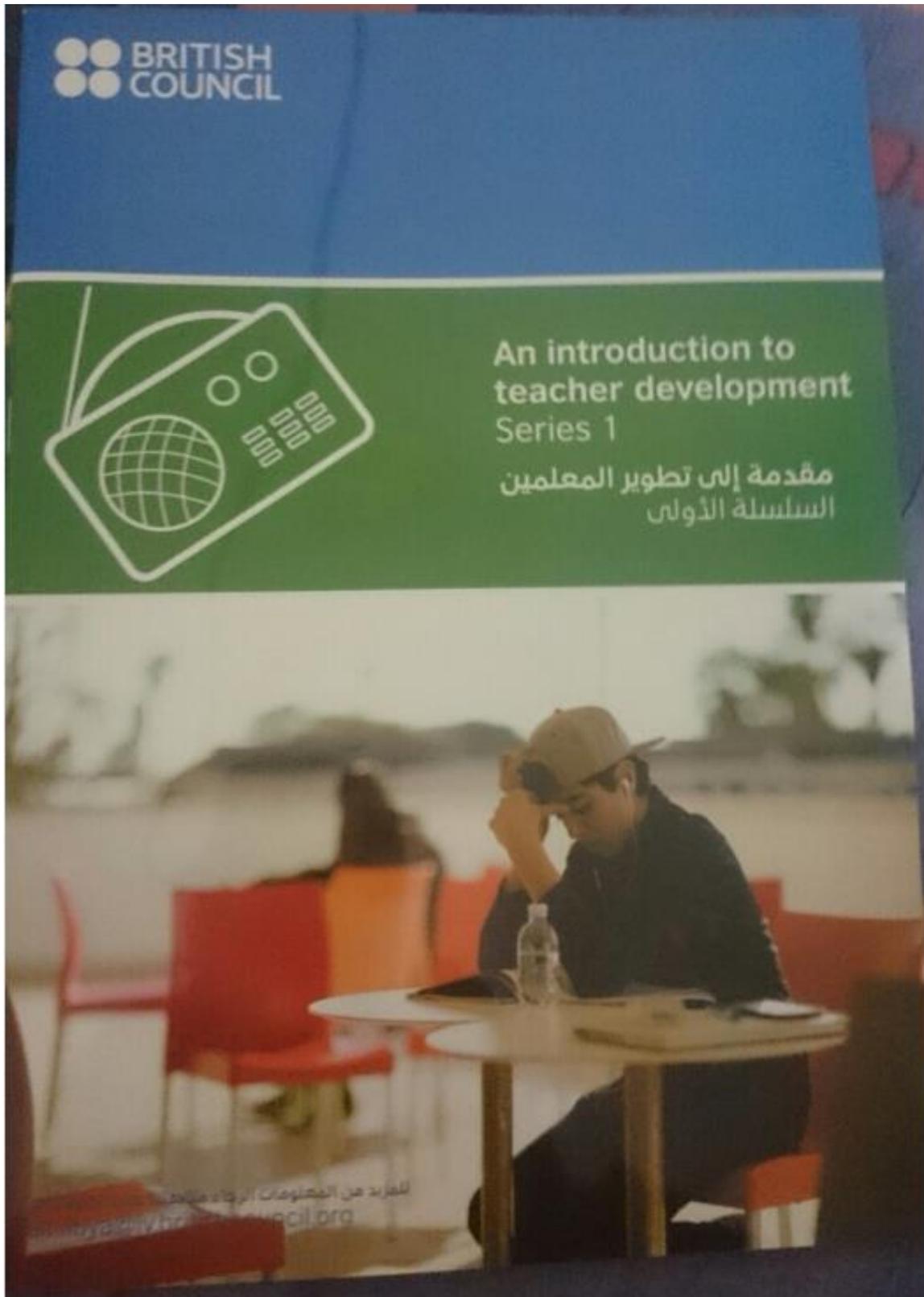
11.12   Appendix 12: The conference papers

Conference presentation

Ahmed, A. (2015a). The Impact of Secondary Education Certificate Examination on Language Teaching and Students Learning Outcomes in the Libyan Context. *Belgrade International Conference on Education*, 23 – 25 October 2015. Tomorrow People Organization: Serbia

Ahmed, A. (2015b). The Impact of Secondary Education Certificate Examination on Language Teaching in the Libyan Context: A Case Study. *PGR conference*, 13 November 2015. University of Huddersfield: UK

Ahmed, A. (2016). The impact of national examination on language learning practice in the Libyan context. *The conference of the School of Education and Professional Development: University of Huddersfield*, 7 April 2016, University of Huddersfield: UK

Poster

Ahmed, A. (2017). The Impact of Public Examination on Language Classroom Assessment. *Language Testing Forum 2017: Assessing languages in the UK*, 24 – 26 November 2017. University of Huddersfield: UK

Paper accepted

Ahmed, A. (2018). Quality of Education in an On-going Conflict Context**.** *The conference of the School of Education and Professional Development: University of Huddersfield*, 1 May 2018, University of Huddersfield: UK