

**Investigating the Washback of English Language Testing
Reform on Teaching English in Oman**

Safa Ibrahim Mohamed Al-Hinai

PhD

University of York

Department of Education

August 2023

Abstract

Washback research has gained significant interest among educationalists in recent years. This study examines the new English elective diploma examination in Oman, which assesses the English language proficiency of students seeking admission to tertiary education or the labour market. This research focuses on the intended consequences of the examination, as reported by Ministry of Education (MoE) decision-makers, teachers' perceptions of the examination, and the perceived effects on teaching practices. The study also identifies the extent of perceived washback effects associated with teacher-related factors. This study used various data-gathering methods, including a nationwide teacher survey, classroom observation and semi-structured interviews. Descriptive analysis was used to determine the prevalence of certain perceptions among teachers and inferential analysis was used to identify the factors contributing to specific teaching practices. Thematic analysis was used to understand the mechanics of the washback effect.

The study found that employing an examination to prompt pedagogical change does not always yield desired outcomes. The results indicate that the impact of the new English elective diploma examination on teaching practice is complex and inconsistent. There was no direct influence on teachers' teaching methods; rather, washback operated through other factors, such as teachers' beliefs, attitudes and responses to the test, as well as aspects such as gender and academic qualifications.

This study emphasises the need for further research on washback using appropriate methodologies and refinements to existing research tools. It proposes a conceptual framework of washback in teaching that contributes to understanding of the mechanisms of washback, which should be relevant not only in Oman but also to other Middle Eastern countries with similar exam-oriented contexts. The findings of this study may help improve English language teaching in Oman and serve as an impetus for curricular and examination reform.

Dedication

This thesis is dedicated to my late father, Ibrahim Mohamed Al-Hinai, who passed away on 23 September 2003. May Allah's mercy and forgiveness be upon his soul. Amen! Special thanks go to my mother, Raiya, who has encouraged and supported me to make this dream come true.

Acknowledgements

I am indebted to several people, without whom the completion of this thesis would have been impossible.

First and foremost, I would like to extend my greatest gratitude and my sincere appreciation to my supervisor, Dr Khaled El Ebyary, who has been of great support to me both academically and personally. His endless guidance, support and encouragement have been essential to the completion of this thesis. I would also like to extend my thanks to my thesis Advisory Panel, Dr Bill Soden, for his constructive feedback, ongoing advice and guidance throughout my four years of study at the University of York. I could not have wished for better supervision and support.

I would like to thank the Ministry of Education and the Ministry of Higher Education, Research and Innovation in Oman for supporting the study. My appreciation goes to all the principals and teachers who answered all my queries while I was carrying out my fieldwork. I also wish to acknowledge the staff and students of the Department of Education, University of York, for their generous assistance.

Last, but by no means least, I wish to thank my family for the faith and love that they have shown me during this long PhD journey. Thanks Mum and my dear brothers and sisters, for your encouragement, confidence and motivation, which enabled me to get through many challenging times.

Declaration

I hereby declare that no portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning. I further declare that this thesis is my own original work, except where reference is made in the text of the thesis to the work of others.

Table of Contents

Abstract	ii
Dedication	iii
Acknowledgements	iv
Declaration	v
Table of Contents	vi
List of Tables	x
List of Figures	xii
List of Abbreviations	xiii
Chapter 1. Introduction	1
1.1 Overview	1
1.2 Background.....	1
1.3 Rationale and Motivation for the Study	4
1.4 Significance of the Study.....	7
1.5 Aims and Objectives	10
1.6 Research Questions	12
1.7 Thesis Structure.....	13
Chapter 2. Education, English Language Instruction and Reform in Oman	14
2.1 Country Context.....	14
2.2 Educational Context in Oman.....	14
2.2.1 Development of education in Oman.....	14
2.2.2 English language curriculum development.....	17
2.2.3 Development of assessment	20
2.2.4 New English elective curriculum	23
2.2.5 English language teaching	27
2.3 Chapter Summary.....	31
Chapter 3. Literature Review	33
3.1 Introduction.....	33
3.2 Washback: Definition and Scope	33
3.2.1 Washback and terminology.....	36
3.3 Washback: Positive or Negative	38
3.3.1 Positive washback	38
3.3.2 Negative washback.....	40
3.4 Mechanisms of Washback	44
3.4.1 Alderson and Wall's (1993) hypotheses.....	44
3.4.2 Hughes' (1993) model of washback.....	47
3.4.3 Bailey's (1996) model	49

3.4.4 Green's (2007) model of washback.....	53
3.5 Characteristics of Washback.....	57
3.5.1 Washback dimensions	57
3.5.2 Aspects of learning and teaching potentially influenced by the examination	60
3.5.3 Factors mediating the washback process	60
3.6 Washback to Participants.....	62
3.6.1 Washback on teachers' perceptions.....	62
3.6.2 Perceptions of test writers.....	72
3.7 Washback Processes.....	78
3.7.1 Washback on teaching content.....	78
3.7.2 Washback on teaching methods	83
3.8 Factors Influencing Washback.....	88
3.8.1 Teacher characteristics.....	89
3.8.2 Educational background	92
3.9 Washback Studies in the Middle East	94
3.10 Gaps in the Existing Washback Literature	98
3.10.1 Research questions	100
3.10.2 Conceptual framework.....	101
3.11 Summary.....	106
Chapter 4. Research Methodology	107
4.1 Introduction.....	107
4.2 Research Paradigm.....	107
4.3 Methodology in Washback Studies.....	112
4.4 Research Design.....	114
4.4.1 Approaches to studying washback	114
4.4.2 Mixed-methods design.....	117
4.5 Research Phases	119
4.5.1 Phase one	119
4.5.2 Phase two	119
4.5.3 Phase three	120
4.6 Research Participants and Sampling Techniques.....	121
4.6.1 Description of the teachers.....	122
4.6.2 MoE decision-makers	125
4.7 Research Instruments	126
4.7.1 Questionnaire	126
4.7.2 Classroom observation.....	134
4.7.3 Interviews.....	140
4.8 Data Analysis.....	145

4.8.1 Quantitative data analysis	145
4.8.2 Qualitative data analysis	146
4.8.3 Observation data analysis	147
4.8.4 Interview data analysis	148
4.9 Validity and Reliability	150
4.10 My role as a researcher	154
4.11 Ethical Considerations.....	155
4.12 Summary.....	158
Chapter 5. Findings	160
5.1 Introduction.....	160
5.2 Teachers' Perceptions of the New Examination.....	161
5.2.1 Teachers' reactions to the new examination.....	162
5.2.2 Reasons for change.....	165
5.2.3 New examination format	168
5.2.4 Extra pressure placed on teaching by the new examination	171
5.2.5 Challenges in teaching.....	173
5.2.6 New English elective examination design	176
5.2.7 Decision-making concerning the new examination	178
5.2.8 Washback from the examination on classroom teaching	181
5.3 Intended Washback	199
5.4 Washback on Teachers' Classroom Practices	212
5.4.1 Analysis of classroom observations	213
5.4.2 Further analysis of the observations	226
5.5 The Washback Effect vis-à-vis Teacher Factors.....	234
5.5.1 Teachers' teaching practices	235
5.5.2 Teachers' classroom assessment practices	239
5.6 Conclusion	244
Chapter 6. Discussion	248
6.1 Introduction.....	248
6.2 Synthesis of Research Findings	248
6.2.1 Teachers' perceptions of the new examination.....	248
6.2.2 Intended washback from the new examination.....	259
6.2.3 The perceived washback effect on teachers' classroom practices	265
6.2.4 Extent to which teachers' characteristics influence the intensity of perceived washback effects	273
Chapter 7. Conclusion	277
7.1 Introduction.....	277
7.2 Summary of the Research.....	277

7.2.1 Research problem	277
7.2.2 Research procedures	279
7.2.3 Design and methodology	280
7.3 Key Findings.....	280
7.4 Study Limitations	282
7.5 Implications of the Study.....	283
7.6 Recommendations for Further Research	289
7.7 Personal Reflections.....	291
7.8 Concluding Remarks	292
References	293
Appendix A. Examination Specifications and Sample Items	324
Appendix B. Skills Covered in English Insights 3.....	328
Appendix C. English Elective Teachers at Grade 12 in 2018–2021	330
Appendix D. Teachers’ Questionnaire (Main Study)	331
Appendix E. Observation Instruments	340
Appendix F. Sample Observation Scheme and Field Notes – Teacher A	342
Appendix G. Sample Observation Scheme and Field Notes – Teacher B.....	344
Appendix H. Interview Schedule for Teachers (Main Study)	346
Appendix I. MoE Decision-Maker Interview Schedule (Main Study)	351
Appendix J. Sample Observation Transcript – Teacher A	355
Appendix K. Sample Observation Transcript – Teacher B	357
Appendix L. Coding Scheme for Teacher Interviews.....	361
Appendix M. Coding Scheme for MoE Decision-Makers’ Interviews	363
Appendix N. University of York Ethics Approval.....	364
Appendix O. Omani Ministry of Education Approval	369
Appendix P. Consent Forms.....	370
Appendix Q. Supplementary Tables	391

List of Tables

Table 2.1. Distribution of marks for continuous assessment and end-of-semester tests	25
Table 2.2. Key learning outcomes in the four main skills at grade 12 in the elective curriculum	26
Table 3.1. Proposed classification of Alderson and Wall's (1993) washback hypotheses	46
Table 4.1. Demographic characteristics of teachers in the survey sample	123
Table 4.2. Descriptive information for interviewed teachers	125
Table 5.1. Teachers' reactions to the new English elective diploma examination	162
Table 5.2. The reasons for the examination change as perceived by teachers	166
Table 5.3. Perceptions of changes in the English elective examination specifications	168
Table 5.4. Extra pressures on teaching from the examination	171
Table 5.5. Challenges in teaching the new English elective curriculum	174
Table 5.6. Teachers' perceptions of decision-making concerning the new examination	178
Table 5.7. Changes teachers would like to make to their teaching	182
Table 5.8. Teachers' practices for improving students' learning	184
Table 5.9. Teaching activities	189
Table 5.10. Role of mock examinations as perceived by teachers	192
Table 5.11. Assessment methods most frequently used by teachers	195
Table 5.12. Factors influencing classroom teaching	196
Table 5.13. Sample coding of data	200
Table 5.14. Percentage of time spent on different communicative modes (Teacher A)	214
Table 5.15. Average participant organisation as a percentage of total lesson time	215
Table 5.16. Content types as a percentage of total lesson time	218
Table 5.17. Student modality as a percentage of total lesson time	220
Table 5.18. Source material types used as a percentage of total lesson time	223
Table 5.19. References to the examination as a percentage of total class time	227
Table 5.20. Focus on examination strategies: Teacher B (cross-referenced with Teachers A and C)	228
Table 5.21. Example of "the importance of openings" (Teacher B, writing task)	229
Table 5.22. Focus on exam strategies: Teacher A (cross-referenced with Teachers B and C)	230
Table 5.23. Example: Identifying the meaning of new vocabulary (Teacher A, reading task)	230
Table 5.24. Focus on exam strategies: Teacher C (cross-referenced with Teachers A and B)	231
Table 5.25. Example: Making an outline (Teacher C, writing task)	231
Table 5.26. Teacher–student secondary interactions as a percentage of total class time	232
Table 5.27. Emphasising skills likely to be tested in the examination	236
Table 5.28. Skipping activities in the textbook	236
Table 5.29. Putting more emphasis on the integration of skills	237
Table 5.30. Giving feedback to students	239
Table 5.31. Preparing students for the real examination	240

Table 5.32. Identifying areas for re-teaching	241
Table 5.33. Getting students to pay attention in class	242
Table 5.34. Encouraging students to study regularly.....	243

List of Figures

Figure 2.1. Oman’s Basic Education System (adapted from the National Report on Quality Education in Oman, Ministry of Education, Oman, 2004).....	17
Figure 3.1. Hughes’ (1993) washback trichotomy	47
Figure 3.2. Bailey’s (1996, p. 264) basic model of washback	50
Figure 3.3. Green’s model of washback direction, variability and intensity (2007, p. 24)	55
Figure 3.4. Burrows’ (2004) set of models of washback responses	64
Figure 3.5. Proposed model of washback on teaching (adapted from Hughes, 1993, cited in Bailey, 1996)	105
Figure 4.1. Proposed sequential explanatory research design.....	120
Figure 5.1. Classroom observation: Seating arrangement (Teacher A).....	216
Figure 5.2. Classroom observation: Seating arrangement (Teacher C).....	217
Figure 5.3. Observed writing activity (Teacher B).....	222
Figure 5.4. Writing lesson combining material types in Teacher B’s classroom.....	225
Figure 5.5. Writing lesson combining material types in Teacher A’s classroom.....	225

List of Abbreviations

BAK	Beliefs, Assumptions and Knowledge
CLT	Communicative language teaching
COLT	Communicative Orientation of Language Teaching
EFL	English as a Foreign Language
ELCE	English Language Curriculum Department
ELT	English language teaching
ESL	English as a Second Language
GDEE	General Directorate for Educational Evaluation
HKCEE	Hong Kong Certificate of Education Examination
IELTS	International English Language Testing System
MDI	Measurement-driven Instruction
MoE	Ministry of Education
NCSI	National Center for Statistics and Information
NVivo	A qualitative data analysis (QDA) computer software package
OWTE	<i>Our World Through English</i>
SAH	Student Assessment Handbook
SIPTT	Specialized Centre for Professional Training of Teachers
SPSS	Statistical Package for the Social Sciences
SQA	Scottish Qualifications Authority
SQU	Sultan Qaboos University
TESOL	Teaching English to speakers of other languages
TOEFL	Test of English as a Foreign Language

Chapter 1. Introduction

1.1 Overview

Tests that can serve either as gatekeepers or as devices for educational reform have long influenced teachers and students in terms of their attitudes, motivations and practices in language classrooms (Pearson, 1988). However, such influence, commonly termed “washback” is a complex area of study, as previous investigations have revealed that it is difficult to predict and it is not clear how the impacts of tests will manifest themselves outside the research boundary, especially when intervening variables (i.e. beyond the test itself) in the given educational setting may determine or preclude the scope and nature of washback (Shohamy, 2001). Due to these complexities, empirical research is needed to examine the washback mechanism, to investigate whether testing causes positive or negative washback (or a combination of the two) and to address apparent contradictions in the existing literature (Alderson & Wall, 1993; Cheng, 1997; Spratt, 2005).

This study aimed to explore the perceived washback effects of a high-stakes English language examination¹ on the teaching and learning of English as a foreign language in Omani public schools. Specifically, it sought to study the impact of a new elective English diploma examination, administered in grade 12, by exploring teachers’ perceptions of the examination and its effects, as well as observing teaching practices in the classroom. It especially aimed to explore whether or not the washback on classroom practices intended by those implementing the reform, i.e. Ministry of Education (MoE) decision-makers, had been achieved.

In this chapter, I describe the background to the research, the rationale and motivation for undertaking the study and its aims and significance. Lastly, the structure of the thesis is outlined.

1.2 Background

During the last two decades, there has been a reconceptualisation of the link between teaching/learning on the one hand and high-stakes summative testing on the other due to a growing understanding of the processes that operate within educational systems.

¹ In this thesis, I refer to both “tests” and “examinations” depending on the name of the assessment and following the usage of the authors of the respective papers. There is not necessarily any functional distinction between the terms: the high-stakes assessment under study in this research is called the English elective examination, whereas other studies examine assessments that are termed tests but are likewise high-stakes.

Crucially, it has been recognised that tests can have a detrimental effect on students' learning, particularly if scores are used to provide or take away opportunities related to access to education, employment, or immigration, and they can also act as a "disciplinary tool" (Shohamy, 2001, p. 17). Certain demands associated with the test are imposed on teachers and students by those in authority such that teachers and students have no choice but to change their behaviours in line with these demands (Madaus, 1992; Spolsky, 1994). This influence, a phenomenon known as "washback" has emerged as a key area in language teaching and testing research, serving as an indicator of the impact of a test on the teaching and learning process (Alderson & Wall, 1993; Bachman & Palmer, 1996). The term "washback" is used to refer to circumstances in which teachers and students "do things they would not necessarily otherwise do because of the test" (Alderson & Wall, 1993, p. 117). Teachers demonstrate this concept in action when, for example, they focus on the content of a test to the detriment of other aspects of the curriculum or concentrate on certain types of questions because these are emphasised in a high-stakes public test.

Acknowledgment of the complex nature of the impact of washback on teaching and learning has led to the identification of two types, positive (or intended/beneficial) washback and negative (or unintended/harmful) washback. Positive washback occurs when tests are used to improve teaching practice and thereby promote the learning process, whereas negative washback, as suggested by Taylor (2005), is said to result when the design and use of a test is "based on a narrow definition of language ability, and so constrains the teaching and learning context" (p. 145). Consequently, washback is a phenomenon with great potential influence, positive or negative, and it has been seen as a strategy through which decision-makers can encourage educational innovation among teachers (Shohamy, 1993). Washback can have major implications for both students (whose test results can determine their access to important opportunities in education or employment) and teachers (whose professional development may depend on how successful they are in preparing their students for a test). Targeting positive washback therefore seems a highly worthwhile aim (Bailey, 1996; Cheng, 2005; Cheng & Curtis, 2004; Watanabe, 2004).

Given the highly examination-oriented education system in Oman, it is not surprising that its high-stakes tests are considered important by most teachers and students. English language education in Oman adopted the communicative language teaching (CLT) approach around two decades ago. However, several local studies have demonstrated that

these methods have not always been implemented as desired (2.2.5) with the use of traditional teacher-centred approaches persisting alongside a poor application of collaborative teaching and learning methods. Thus, a new elective diploma assessment system, based on a new curriculum, was introduced in the 2018–2019 school year aimed at addressing the negative washback on teaching and learning associated with the previous system.

Research on washback suggests that changing a particular test to improve teaching practices can work in certain settings and that positive washback may result from different factors. Tests that are seen as influential by teachers will clearly have an impact on how teachers operate in the classroom, as well as on their perceptions of their own teaching and worth (Cheng, 2005; Hamp-Lyons, 1997; Shohamy et al., 1996; Spratt, 2005; Watanabe, 2004). What teachers think and believe about the introduction of such an examination and how familiar they are with what it aims to assess are closely linked to what and how they teach. High-stakes tests may influence teachers to adjust their teaching methodology and adopt the CLT approach encouraged by the new test design (Bailey, 1996; Taylor, 2005). However, the design of an examination may lead to consequences that the test designers fail to anticipate and which may not always be desirable (Qi, 2005). It has also been found that while a new test may induce changes with respect to what is taught in the classroom, its influence on how teachers teach is much less clear (Andrews et al., 2002; Burrows, 2004; Qi, 2007; Wall & Alderson, 1993). Cheng (1997) has observed that this is determined by the level of the washback effect, since the impact of a test may be one of “form” alone, comprising a superficial change in the format of the teaching, or one of “substance”, with more substantial change brought about by the test.

Thus, research clearly indicates that washback is more complex than the simple conception of positive or negative and it is necessary to take into account aspects such as whether the effects are immediate or delayed, intended or unintended, or superficial or substantial. It is not easy to predict the washback of a test, or what washback will look like where there are other intervening variables (beyond the examination) that may facilitate or preclude its impact. The washback of testing on teaching practices can be affected by various factors, perhaps the most important being the teachers’ personal characteristics (educational background, training, gender, teaching experience, and so on). To improve the likelihood of positive washback, it is important to understand how washback operates and its impact. The stance in this study is that as the positive effects

of a new test on teaching practices depend on the adoption of particular teaching strategies, it is very important to evaluate teachers' actual day-to-day classroom teaching practices. Furthermore, given that the assessment of high-stakes examinations in Oman has so far failed to explore the reasons for the negative washback among local Omani teachers, understanding the mechanisms through which the washback of the new elective diploma examination is likely to operate is crucial.

Before embarking upon investigating the washback phenomenon in this context, it is important to discuss the importance and rationale for conducting this research and its aims and objectives.

1.3 Rationale and Motivation for the Study

Although a focus on language testing and teaching has gained momentum among researchers in Oman, this is the first empirical study to investigate the newly introduced examination based on the new curriculum. Existing studies have primarily focused on related topics, such as the “communicativeness of English tests” or “the predictive validity of the test”. A literature search only identified two studies on the washback of testing in the Omani context. One was an unpublished PhD thesis by Al-Lawatiyah (2002), which investigated washback on teaching and learning processes. The other, by Al-Hinai and Al-Jardani (2020), was only based on perspectives articulated in the existing washback literature and did not test them empirically. The review of literature provided by Al-Hinai and Al-Jardani (2020) concluded that there is a need to investigate this phenomenon empirically, especially addressing the characteristics of the teaching environment in Oman and associated variables, including local society. Moreover, the authors suggested that more effort is needed to promote beneficial washback in this context and that this will require careful diagnosis of the different existing contextual factors and their relationships with each other. Although the authors claimed that a primary focus in their research was to provide a clearer understanding of the interrelated factors operating in this context, it appears that the study presents only general claims about teaching and learning in Oman. To date, there is no empirical evidence or data to justify their main findings, specifically the existence of examination influence. This highlights the need for empirical research to explore the effect of the new high-stakes examination on Omani ELT stakeholders and practitioners to identify what kind of washback has taken place and determine its effect on teachers' instructional practices in the classroom.

Despite all the financial, professional and policy initiatives employed to improve teaching and learning in Oman, teachers' ability to improve their teaching and students' performance remain uncertain (2.2.5). Research has shown that teachers have the greatest influence on the academic achievement of students and that teacher characteristics are a strong predictor of educational outcomes (Denman & Al Mahrooqi, 2019; Organization for Economic Cooperation and Development [OECD], 2005). Indeed, the crucial role of teachers in washback has been acknowledged in various studies (Alderson & Hamp-Lyons, 1996; Bailey, 2005; Burrows, 2004; Cheng, 2000; Shohamy, 2007; Wall, 2005; Watanabe, 2004), as discussed in depth in 3.6.1. Therefore, the aim of this study was to examine the washback from the newly introduced General Education Diploma examination in English on grade 12 teachers' instructional practices in the classroom and specifically whether it has had a positive, negative or neutral impact on classroom teaching.

Previous studies in other contexts have identified the influence of tests on promoting certain teaching and learning processes as a vital element for successful educational change (see Alderson & Hamp-Lyons, 1996; Chapman & Snyder, 2000; Cheng, 2005; Stecher et al., 2004; Wall, 2005). However, existing studies have not sufficiently addressed the kind of influence on teaching and learning practices, the perspectives of teachers as critical agents in the washback process and the perceptions of teachers and MoE stakeholders concerning the nature of intended washback. Therefore, to what extent the new examination has promoted changes in teaching this context is unclear.

It is clear that washback on language testing requires investigation but it also relates to my roles as a teacher and an assessment policy actor. Throughout 17 years of working in the MoE, I have actively participated in implementing policies and perspectives that have influenced the trajectory of assessment reform in Oman, mainly focusing on examination reform and the assessment of students' learning, as well as teacher training in assessment and curricula. I have taken part in various initiatives to enhance teachers' experiences and practices in teaching and have seen first-hand the critical role of examination change and development processes. Thus, this researcher's work experience and knowledge about the context in which this study was conducted provided a comprehensive understanding of teachers' pedagogical and assessment practices. Moreover, I had fairly extensive experience with the test under investigation in this study and had thus formed an understanding of the potential positive impact and negative consequences prior to engaging in this research. In this

research, the role of researcher went beyond that of mere participant or observer in the educational system of Oman (for further discussion of this role, see 4.10).

Furthermore, my position in the MoE showed me a significant aspect of educational reform – translating a national examination reform at the MoE level into classroom-level practices. In this context, assessment reform or any other educational change regarding teaching and students' learning are formulated centrally, drawing on students' achievement levels and policy recommendations. However, there is often a gap between the intentions of the MoE stakeholders and policymakers and how teachers react and interpret such initiatives at the classroom level. My previous experience in teaching and investigations of the perceptions of teachers, students and supervisors concerning assessment policy since the introduction of basic education point to the potential disconnection between the perspectives of teachers and MoE personnel. I understand the centrality of how such a change is implemented in classroom teaching. Therefore, this study aimed to examine the influence of the implementation of the new requirements and format of the English elective diploma examination on instructional practices and to understand how teachers interpreted, adapted and enacted the intended washback of the new examination as proposed by the MoE stakeholders and policymakers. Through this research, I hope to bridge the gap between examination change policy and teachers' practices in the classroom, contributing to the practical evolution of the basic education reform processes.

In this thesis, I strive to contribute to advancing teaching practices, learning and assessment policies in my country. I feel an ethical obligation to examine how the examination reform has been interpreted and adopted in classroom. My experience as a teacher and examination policy member in the MoE has highlighted the process and procedures of teachers' involvement in assessment reform, especially in a highly centralised education system with limited opportunities to consider teachers' views. As a member of the MoE, I believe in the vital role examination reform can play in transforming the perceptions and practices of key stakeholders, particularly teachers, encouraging them to improve teaching and learning. This study investigated how teachers perceived and implemented the new examination requirements to foster teaching and learning. It ultimately aimed to support the efforts of curricular stakeholders in enhancing teaching and learning by encouraging more optimal teaching practices based on exploring teachers' input and the reality of practices in the classroom. The intention is that the study

findings contribute to policy decisions related to English language teaching and learning based on empirical evidence derived from the reform initiatives.

This study's main goal concerns examining the complex nature of washback and its relation to instructional practices. Many impact studies have been conducted in educational systems in which the reform of examinations is conducted by schools working independently (i.e. decentralised systems). However, few studies have investigated the apparent washback effects from a new examination reform developed by MoE stakeholders and policymakers within a centralised system like Oman and other similar exam-oriented contexts (such as Middle East countries). This study contributes to understanding how the requirements and format of a new examination affect teaching in the classroom from the perspectives of teachers in a context in which the implementation processes are controlled centrally through the MoE. The following section discusses the significance of the study and the implications of washback from testing on the teaching and learning process.

1.4 Significance of the Study

The washback of assessment has been the main focus of many studies in the arena of educational research. According to Cheng and Curtis (2004), test results in many educational systems are used to determine innovation and curriculum change. This was raised as a fundamental concern by a major evaluation report on the Omani education system, (MoE & World Bank, 2012), which argued that a rigorous assessment system “that provides students and their parents with realistic feedback from the early grades would help to set and maintain high standards” (p. 67). A recent study in Oman (Mohammed, 2019) explored the challenges that hinder Omani school leavers from attaining high marks in the English language placement test set by the Omani higher education sector. The study identified two main problematic factors: a) the exam-oriented teaching practices and rigid syllabus, and b) a lack of a lack of communication channels between grade12 and higher education teaching staff. Assessment reforms thus represent an important lever that may positively affect standards of achievement.

It is vitally important to explore the effect of this new high-stakes examination on Omani ELT stakeholders and practitioners to identify what kind of washback it has had and determine its effect on teachers' instructional practices in the classroom. Although this research provides insights into the washback effect in Oman, it could also contribute to a

broader understanding of the phenomenon beyond the immediate context. It also offers insights into English language teaching in grade 12 schools, which is based on a new communicative curriculum reform. As pointed out by Haladyna et al. (1991), one of the most important steps in evaluating the product of such a change is to understand “the nature of the processes or means that led to the end. It is not just that means are appraised in terms of the ends they lead to, but ends are appraised in terms of the means that produce them” (p. 6).

Early research into washback focused on the type of impact brought about by the introduction of a new examination (e.g. attitudes, curriculum, materials), but how change occurred in teaching was less clear. Bailey’s (1996) model, for example, recognises the relationships between participants, processes and products, but it provides little explanation on how these three components relate to each other or operate with respect to teaching and learning or how changes in teaching occur, if indeed they do. The mechanism for explaining the impact of testing on how teachers teach and why they teach the way they do has remained under-explored. Therefore, this study sought to address an empirical gap in previous research concerning the relationships between teachers’ perceptions and their practices in classroom teaching and the intended washback of the new examination. By addressing this gap, this study explored the washback of testing in terms of how examination reform has manifested or not in the teaching context (the mechanism of washback) or how a test can lead to changes in teaching. In addition, this study also investigated factors other than the test, such as teachers’ personal characteristics, which may help to broaden our understanding and address the concerns raised in earlier studies about the need for more in-depth research into the complex nature of test washback (e.g. Alderson & Wall, 1993; Bailey, 1996; Cheng, 2005; Shohamy et al., 1996; Wall, 2005). If a new examination is to have its intended washback, policymakers and examination specialists need to consider a range of factors that will affect its success or failure, including how its impact influences teachers’ attitudes and their behaviours in the classroom. This was highlighted by Alderson and Wall (1993), who argued that examination writers should pay more attention to the washback effect of their tests and should also be cautious about the oversimplified belief that “good” tests will automatically promote “good” impact. This could apply in relation to the MoE stakeholders (examination writers, curriculum developers, supervisors and trainers), and probably also to the general public in this context.

Moreover, changes in the English elective diploma examination were prompted by the perception that it focused solely on the development of reading and writing skills as discrete components of linguistic knowledge, neglecting other skills (MoE, 2017). Whether the changes made to the examination recently will yield the effects intended by the MoE stakeholders in improving English language teaching and learning outcomes across all schools in the whole country will depend largely on how teachers perceive and respond to this reform movement. The importance of teacher perceptions in shaping the impact of testing reform is well documented in language teaching and linguistics (e.g. Alderson & Hamp-Lyons, 1996; Bailey, 2005; Cheng, 2000; Kim & Isaacs, 2018; Shohamy, 2007; Wall, 2005; Watanabe, 2004; Winke, 2011). As Winke (2011) noted, teachers' perceptions of the testing process play a critical role in the successful utilisation of test results: "they administer tests, know their students and can see how the testing affects them, and they recognise – sometimes even decide – how the tests affect what is taught" (p. 633).

Nonetheless, in a country like Oman, which has a highly centralised testing system, the development of curricula and assessment is typically undertaken through a top-down strategy, which gives key stakeholders, such as teachers and students, little opportunity to express their opinions and concerns (Fan et al., 2020). In language testing, this group of stakeholders has historically been neglected in the process of reform (Shohamy, 2001). Consequently, in relation to the examination development process, little is yet known about the perceptions and reactions of English language teachers in post-basic schools (grades 11 and 12) in Oman to the introduction of the revised version of the examination.

It has been asserted that there tend to be discrepancies between the intentions behind an educational change and teachers' interpretation of that change (Andrews, 1995; Smith et al., 1994). One of the main issues that has been of concern to teachers is the impact the new elective test may have on teaching and learning. Teachers are responsible for covering all aspects of the curriculum, as determined by the Ministry. However, this has led to a gap between the MoE stakeholders' expectations and actual practice in the classroom, as shown by Al Balushi and Griffiths (2013). Supervisors and teachers have not been actively involved in the design of the new examination or the marking procedures; rather, the change in the examination specifications and materials has been imposed on them without any clear idea of the rationale behind it. The examination writers and curriculum developers in the MoE have not clearly stated their intentions in

any examination documentation released to EFL educators. More specifically, they have not defined the scope of the “impact” intended by the new examination, nor have they stated their views of how this could be achieved. Teachers implementing the change have felt pressurised to cover all the exam-related materials, especially with the change to include all four of the main language skills. The only points made in favour of the examination are that it has encouraged the integration of the four language skills and prompted students to work harder, since the changes are intended to help prepare students for international examinations, such as TOEFL and IELTS (MoE, 2018/2019).

Given the above, it remains unclear if the test reform process is likely to bring about the intended washback on the teaching and learning of English in grade 12. Thus, this study sought to investigate the perceptions of English language teachers in grade 12 post-basic schools in Oman concerning the new English elective diploma examination, particularly the new requirements and design. It examined the intended washback as perceived by the MoE committee, responsible for overseeing the introduction and implementation of the examination and the recent suite of reforms. In exploring the perceptions of teachers and the intentions of the MoE members, this study further contributes to the need for improved dialogue concerning language testing processes, particularly in the Omani context.

1.5 Aims and Objectives

Based on the above, this study investigated “Why has the intended washback on teaching occurred or not?” There has been no previous empirical research in the Omani context that can provide an answer to this question, which is of great importance to the students who take this examination, the teachers, school principals and parents, as well as other educators and researchers in the EFL field. By adopting a mixed methods design, this study investigated the complex nature of washback and its effects on teaching practices through empirical research in the Omani context. The main aim of this thesis is to explain how washback operates by understanding the perceptions and practices of teachers in response to the new examination requirements and format. There is a critical focus on how the perceived washback effects influence teaching and why change in teaching occurs or not. To this end, the study investigated current theories and research regarding washback in the field of language teaching and testing.

It is hoped that the thesis will provide useful findings to the MoE in Oman, helping support its goals and objectives in improving the assessment system but also, more

importantly, in developing and improving the teaching and learning of English in Oman. To guide this study and ensure a systemic approach, the key objectives were defined as follows:

Objective 1: To examine the washback effects of the new elective diploma examination through investigating the perceptions of teachers and how their views influence their behaviours. This involved exploring teachers' perceptions, experiences and attitudes with respect to the impact of the new examination on their teaching and the extent to which this impact was evident in their classroom teaching.

Objective 2: To investigate and understand the intended washback effect of the new elective diploma examination from the perspectives of MoE decision-makers in the study context. This entailed investigation of the rationale for and uses and objectives of the examination change. It involved exploring MoE stakeholders' perceptions, experiences and attitudes with respect to the intended washback and the extent to which it affected teachers' practices as they hoped.

Objective 3: To explore the perceived washback effects from the new English elective diploma examination on classroom teaching and the relationship between actual teaching and the intentions behind the test (as perceived by the MoE decision-makers).

Objective 4: To explore why the apparent washback effects of the elective diploma took the shape they did, the extent to which this was influenced by individual teacher characteristics and the reasons for the presence or absence of the intended washback on teaching as proposed by MoE decision-makers.

Objective 5: To apply the study findings to inform future assessment reform policies and practices and advance the existing research literature on the complexity of washback in Oman and other exam-oriented contexts. This required synthesising the study findings, their practical implications for teaching and learning in this context and the contributions they can make to the scholarly literature on washback.

By addressing these objectives, this thesis aims to provide valuable insights into the process of washback within the Omani context and enhance the teaching and learning landscape. To do so, the study set out to answer four main research questions, detailed in the following section.

1.6 Research Questions

To capture the reality, complexity and variation of classroom practice in the study context, it was necessary to gain insights into how grade 12 teachers perceived and reacted to the new examination and to identify how these perceptions of the new design and content of the new examination aligned with the intentions and expectations of the MoE decision-makers. Thus, the study investigated and recorded the perceived impact of the new examination on teaching practices by first exploring the views and attitudes of teachers and then examining the intended washback as proposed by the MoE decision-makers, establishing a general understanding of teachers' attitudes towards examination change and how these perceptions affect their practices in the classroom. This approach was chosen with a view to the data from the qualitative phase (RQ2 and RQ3) building and expanding on the data obtained from the first quantitative phase (RQ1), thus enhancing understanding of the washback effects of the new test. The qualitative data could help explore relationships between the teachers' perceptions of the examination change, what they actually did in the classroom and whether the new examination affected teaching in the ways intended by the MoE decision-makers. This was important as what might be positive for examination writers or curriculum developers might be negative for teachers as the different evaluators involved in washback have different objectives (Alderson, 1992).

By observing the examination reform, this study sought to identify differences between the washback intended by MoE decision-makers, i.e. top level down reform, and perceptions of washback as implemented by practitioners, i.e. teachers, from the bottom level up. Exploring teachers' perceptions of examination change could greatly enhance knowledge and understanding of how intended washback can influence teaching behaviours and in what areas. To this end, the following research questions were used to guide the study:

- RQ1. What are teachers' perceptions of the new English elective diploma examination?
- RQ2. What is the intended washback from the new English elective diploma examination according to MoE decision-makers?
- RQ3. What are the nature and scope of the apparent washback effects resulting from the new English elective diploma examination on teachers' classroom practices?

RQ4. How does the intensity of the apparent washback effects from the new English elective diploma examination differ according to teachers' personal characteristics?

1.7 Thesis Structure

This thesis comprises seven chapters. Following this introductory chapter, the next chapter describes the context of the research study. It presents an overview of the country and the development stages of the educational system in Oman. It also provides descriptions of the development phases of English language textbooks, assessment and teaching practices, and focuses in particular on the reform of the English language elective curriculum. Chapter 3 reviews the relevant literature and focuses on how it describes the washback phenomenon related to a new examination. The relevance of teachers' personal characteristics to washback is also examined, along with the influence that a new examination can have on classroom teaching and why teachers teach the way they teach. This chapter concludes with a conceptual framework which deconstructs the complexity of washback to explain how washback operates. This is followed in Chapter 4 by a description of the research procedures and methods used in this thesis. Chapter 5 presents the results of the data analysis in the main study and Chapter 5 discusses the research findings. The final chapter, Chapter 7, presents the overall conclusions with reference to the research questions, outlines the pedagogical implications of the results and offers some possible directions for future research.

Chapter 2. Education, English Language Instruction and Reform in Oman

This chapter provides an overview of the context of the Omani education generally and English language instruction more specifically. It describes the stages of development of the educational system as a whole and then addresses the development of the English language learning curriculum and approaches to teaching and learning. The latter sections address the new English elective diploma curriculum, including the reform of textbooks, assessment (continuous assessment and end-of-semester tests) and the teaching context.

2.1 Country Context

Oman is located in the south-eastern part of the Arabian Peninsula and covers an area of 309,500 km². It is an Islamic country and Arabic is its official language, but other languages are also spoken, including Lawati, Zedjali, Balushi, Swahili and Kamzari. This linguistic diversity reflects the multicultural fabric and social complexities of Oman, which includes people from many different ethnic groups, for example Arabs, Baluchi, Africans, and also South Asians (Indian, Sri Lankan, Pakistani, Bangladeshi). Oman has a total population of 4,982,568 million (Annual Educational Statistics Report, 2022–2023). Omanis constitute about 58% of this figure, at over 2,800,000 million, while expatriates, at over 2,000,000 million, make up about 42% of the total population (National Center for Statistics and Information [NCSI], 2023). The number of expatriates is large compared to the population of Omanis and English is widely used as a common means of communication, functioning as a lingua franca among people with different first languages.

As education is considered a main pillar of a country's progress and success, the following section provides a brief historical account of the key development phases of education in Oman.

2.2 Educational Context in Oman

2.2.1 Development of education in Oman

Oman's educational system has gone through different phases of development and a gradual evolution. The first phase of educational development lasted until the 1960s. This was confined to Islamic education in Quranic schools and mosque education groups, or so-called "Kuttabs", where the Holy Quran, Islamic principles and some basic principles

of reading and mathematics were the main subjects of study (Al-Nabhani, 2007; Al-Najar, 2016). According to Al-Hammami (1999), these “Kuttab” were the only means of educational provision in Oman in that period. They took place at teachers’ houses, mosques and even under the shade of trees in some areas (Ministry of Education [MoE], 2010). This phase of education sought to help students memorise the Holy Quran and prepare them for work in government areas, such as the judiciary and the official Zakat (charity) collection (MoE, 2010). There was a shift from non-formal education to a more formal system in 1969, when a total of three schools were established in Oman, all of which were for boys. These schools had 900 male students overall, who were taught by 30 teachers (MoE, 2005, 2010).

The second phase in the development of the Omani education system began when His Majesty Sultan Qaboos took power in 1970. This phase represented a turning-point for the spread of education at a more formal and effective level throughout Oman, as the new government concerned the development of a domestic work force vital for the social and economic progress of the country (MoE, 2010, 2016). One of the main changes during this phase was that education in Oman became centralised, reflecting the state political system, which was organised into macro, meso and micro administrative levels. The first Ministry of Education was established in 1970 and was responsible for school education across Oman in terms of national curricula, student assessment, educational evaluation and finance at the macro level (MoE, 2005). At the meso level, the MoE set up a directorate of education under its authority in each of the 11 governorates of the country (Muscat, Al-Dakhiliyah, Al-Batinah North, Al-Batinah South, Al-Buraimi, Dhofar, Al-Dhahirah, Al-Wusta, Musandam, Al-Sharqiyah North, Al-Sharqiyah South). The MoE had overall responsibility for these directorates, which had delegated responsibility for managing the schools located in their region, looking after teachers and students, delivering teacher training, and providing resources to and supervising all schools under their authority. At the micro level, schools were directed and managed by the school principals, who reported to their relative regional directive.

The government also stipulated that education should be compulsory and free for all Omanis across every section of society (Al-Hammami, 1999). Following these reforms, enrolment in education rapidly increased. The number of schools rose dramatically from only three in 1969 to 1,068 public schools and 530 private schools in 2016. The number of students attending school was over 540,000 in 2016, comprising almost equal numbers

of boys and girls (268,501 girls and 271,567 boys), with 518 students in special education schools and 101,860 students in private schools (MoE, 2016).

During the third phase, which began in the 1980s, the MoE made great efforts to achieve significant improvements in teaching and training by reforming curricula and adopting new pedagogical approaches. As part of this process, the former general education system was replaced by the revised basic education system. The previous system was viewed as no longer fit for purpose as it focused on memorising and drilling (Al-Kharusi & Atweh, 2008; Al-Nabhani, 2007; Al-Najar, 2016; Rassekh, 2004), was teacher-centred (Al-Najar, 2016) and did not adequately reflect the requirements of higher education institutions (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2010). The basic education reform was initiated as a response to the government's view that education serves as an economic driver for development, in line with Oman's Economic Vision 2020, which aimed to start the process of shifting the country from an oil-based economy to a knowledge-based economy.

This educational change has led to the division of the education system into a two-stage system. The first stage of education covers a 10-year period, divided into 2 cycles (see Figure 1.1): cycle one (grades 1–4) and cycle two (grades 5–10). Its main aims, according to the MoE, are to reduce the dropout rate among students, ensure students are prepared for higher education, emphasise a learner-centred approach that develops critical thinking and problem-solving skills, and provide equal opportunities for all (Al-Issa & Al-Bulushi, 2011; Al-Najar, 2016; MoE, 2016). To achieve these aims, changes were made to all aspects of the educational system, covering educational philosophy, school buildings and the regulatory framework for the curriculum at all grade levels, including teaching methods, ways and forms of assessment, and study plans (MoE, 2008).

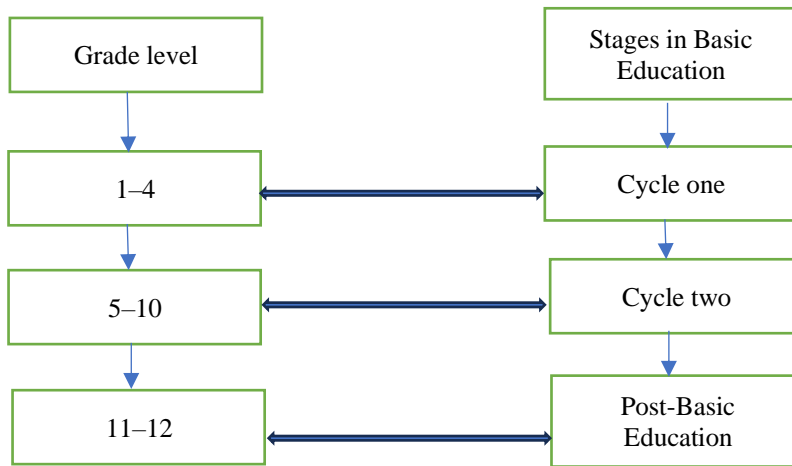


Figure 2.1. Oman’s Basic Education System (adapted from the National Report on Quality Education in Oman, Ministry of Education, Oman, 2004)

In 2007, the first batch of students to complete basic education (grades 1 to 10) moved on to the new two-year post-basic education system (grades 11 and 12), which replaced the previous form of secondary education. The post-basic education programme aims to complement and build upon the basic outcomes developed through the basic education cycles (grades 1–10). The outcomes of post-basic education focus on employment skills and academic strategies, such as further development of skills in critical thinking, creativity, leadership, communication, technological, management and social aspects. In grade 12, students sit the nationwide school-exit qualification, the Diploma of Post-Basic Education (equivalent of A level in the UK), held twice a year (at the end of semesters one and two). Successful candidates are eligible to pursue their studies in higher education institutions, enrol in professional or semi-professional training programmes, or enter the labour market (MoE, 2003).

Since the introduction of the basic education reform in 1998, the English language curriculum has undergone various development stages, which must be considered to understand fully the context of the reform of English language teaching in Oman.

2.2.2 English language curriculum development

Under the old general education system, English language teaching started at grade 4 (9 or 10 years old). According to Al-Issa and Al-Bulushi (2012), the first published textbooks used in schools were “*English for the Arab World*” and “*English for Oman*”, which were written and designed by Longman in the UK. The first in-house textbook, entitled “*Our World through English*” (OWTE), was published in 1992 (MoE, 1997). The decision to adopt an in-house built-in curriculum strategy was taken to better meet the needs of

teachers and learners and also to develop the skills of the staff in the curriculum department at the MoE so that they would be able to write materials themselves. Although the OWTE curriculum was seen as suitable at the time, it was not able to meet the changing needs arising from radical technological advancements and modern-day challenges as it continued to emphasise a teacher-centred approach rather than moving towards more learner-centred and communicative teaching approaches as suggested by educational and linguistic research. Moreover, as noted by Al-Toubi (1998), the OWTE curriculum was not suitable for addressing students' needs because it did not equip them to use the language in real situations and the curriculum lacked authentic materials and activities.

Consequently, as part of the basic education reforms, new English curricula entitled "*English for Me*" for grades 1–10 and "*Engage with English*" for grades 11–12 gradually replaced the previous OWTE. The new curriculum project started in the 1998/1999 academic year, with English language teaching being introduced from grade 1 instead of grade 4, as in general education. The curriculum for all grade levels consists of a coursebook, a skills book, a teacher's book and a resource pack for teachers containing visual aids, posters, activity cards, readers and audio CDs. The curriculum is centralised: all schools use the same English language coursebook and skills book and they are required to finish the prescribed objectives on time, with the textbooks forming the main source of teaching the English language.

The English language curriculum framework states that the basic education curriculum is based on a student-centred approach, embracing the principle of learning by doing (Al-Lamki, 2009). The framework considers individual differences among learners (Al-Lamki, 2009; MoE, 1999) by adopting mixed ability teaching techniques, such as group work. It focuses on developing skills such as communication, cooperative learning, problem solving, research investigation and innovation, rather than more traditional approaches, such as memorisation and depending on the textbook for knowledge (MoE, 2005). According to the document outlining the philosophical underpinnings of the curriculum, the teaching materials for each grade, mainly provided in the coursebook, are "based round a communicative and skills-based methodology and [encourage] active pupil participation and collaboration, rather than a teacher fronted and dominated classroom methodology" (p. 11). The textbook writers claim in the introduction that the aim is to use English that is based on topics and themes which are both interesting to

students and directly related to current affairs. Some of these topics in grades 1–4 include sophisticated communications, such as discussing likes and dislikes and describing the weather. In the early grades, the emphasis is on literacy and the development of receptive skills (listening and reading), aimed at providing a good attitudinal and linguistic foundation for developing language learning in later grades. This stage of learning is intended to develop positive attitudes towards English language learning by emphasising communicative and collaborative teaching and learning.

By the end of grade 10, the textbook states that there is an emphasis on the development of receptive and productive skills and on the use of English for communication purposes. According to the English Language Curriculum Framework (1999), grammar is taught both deductively and inductively in the curriculum to maintain a balance between the two approaches. Grades 11 and 12 continue to focus on the development of both linguistic and social skills in English (e.g. self-study and thinking strategies, entrepreneurial English and 21st-century skills). English is treated particularly as a means of communicating meaning in the outside world in terms of both its vocational potential and its possible use in further education (Al-Issa, 2010). All four main skills – reading, writing, listening and speaking – are thus given equal weight at all grade levels and English is also used to as a means of addressing international dimensions, helping students expand their understanding of different experiences in the outside world and of cultural diversity (English Language Curriculum Framework, 1999).

The curriculum for the post-basic education grade levels is organised on “a core plus electives” model, with students being given the opportunity to select various optional subjects alongside the core element. A range of subject areas relevant to the varying abilities, preferences and aspirations of students has been developed. This gives students the opportunity to study different specialisations (e.g. information technology, English language, science), or to select a subject area for its general interest. The curriculum emphasises the learning of key skills or fundamental competencies, such as numeracy and literacy, teamwork, attitudes and values, to enable students to operate effectively in a wide range of subject areas (MoE, 2008).

In grade 11, students should choose four elective subjects from the study plan in addition to the five compulsory subjects. After completing grade 11, they have the option of continuing with the same choice of elective subjects in grade 12 or substituting these with

other options. One of the choices offered is Elective English, which aimed to provide students in grade 12 with further opportunities for exposure to English and improve their competency level, upskilling them with the language skills required for work and academic studies. To this end, the English Language Curriculum Department (ELCD) at the MoE decided to use an imported coursebook entitled “Reading & Writing Targets”, published by Express Publishing in the UK. The curriculum was implemented for grade 11 in the 2006/2007 academic year and for grade 12 in the following academic year (2007/2008). The time allocated for the course was four lessons per week (40 minutes per lesson).

To fit the above reform of the English language learning curriculum, the MoE considered that there was a need to develop a new assessment system. Therefore, the assessment of students learning has undergone several stages of development, considered in the following paragraphs.

2.2.3 Development of assessment

Prior to the introduction of basic education, the common practice in Oman was to measure students’ achievement through high-stakes end-of-year examinations, which all students were required to pass to progress within the educational system. One of the main concerns of this system was to provide a test of key competencies and skills for entry into higher education institutions. However, as several studies have shown, these examinations almost exclusively assessed rote learning and memorisation (Scottish Qualifications Authority [SQA], 1996), while the examination-based system reduced learners’ motivation and affected their interest in acquiring the language (Al Sawafi, 2014; Al-Toubi, 1998). With the introduction of basic and post-basic education, new assessment policies and strategies were brought in, with the assistance of the SQA. The MoE stakeholders established working groups for all subject areas, including English language. These working groups, which consisted of teachers and supervisors, were trained in item writing to assess higher-order thinking skills and in classroom assessment practices. The test items produced by the committee members were incorporated into basic education assessment. The subject committees were also instructed on how to write a new examination specification for all grade levels that would allow the introduction of a wider range of assessment instruments, rather than just depending on high-stakes end-of-year examinations (Al Balushi & Griffiths, 2013).

In 2004/05, a new continuous assessment system was introduced. This was intended not only to provide a more accurate picture of students' attainments and needs, but also to achieve an alignment between classroom assessment and what had been taught and learned in class, thereby giving greater reliability and validity to the strategies and uses of the assessment system. This new system aimed to move away from over reliance on paper-and-pencil tests and reduce the proportion based on memorisation, as well as to emphasise autonomous learning in assessing students' level. More reinforcement of formative assessment principles and strategies was included in grades 1–4 and a combination of formative and summative assessment was included in grades 5–10 and 11–12. Teachers are now expected to assess students' skills and knowledge using a range of tools, such as quizzes, final tests, short written responses or oral tests, reports, projects, daily observations and self-assessment tasks.

The assessment documents were all developed by the General Directorate for Educational Evaluation (GDDE) in the MoE. The general assessment document was intended to provide a framework for the supervisors, teachers, trainers, curriculum officers and others involved in assessing student performance and sets out the main assessment policies for all subjects at all grades. The "Student Assessment Handbook" (SAH) outlines the assessment policies, continuous assessment arrangements, weightings and formal specifications for the end-of-semester examination. The learning outcomes are grouped under five main areas: Listening, Speaking, Reading, Writing, and Grammar & Vocabulary. A rating scale is provided for each learning outcome, following the standard "five levels of performance" model. Marks are awarded based on three methods of assessment: continuous assessment, class tests and end-of-semester assessments (one mid-year test and one end-of-year examination). Students in cycle one are assessed through quizzes and classroom observations, with no formal examinations. Students in grades 5–11 follow a system of continuous assessment. The examinations for grades 5–11 are written, administered and marked at the regional level; only grade 12 examinations are produced and scored centrally at the Ministry (Ministry of Education, Oman & World Bank, 2012). To increase the reliability of continuous assessment practices in the classroom, formal and informal moderation procedures for checking teachers' practices in marking students' attainment were introduced (MoE, 2008).

Under the SAH, end-of-semester assessments are administered twice a year, in January (the mid-year test) and June (the end-of-year examination), and the marks are aggregated.

The final large-scale public examination is produced by a central team and administered by MoE stakeholders under rigorous procedures. This examination represents 70% of the students' overall mark and they must pass it to be awarded the General Education Diploma. The results of continuous assessment and the mid-year test in January are collated into an overall result, marked out of 100. For all grade levels, scores below 50 are rated as a "fail" and results at or above 50 are rated as a "pass". The grade 12 end-of-year examination is considered high stakes because it is the principal formal instrument for "(1) certifying that students have satisfactorily completed post-basic education" and "(2) selecting students for jobs or for admission to a tertiary-level institution" (p. 13).

Although the SQA reform process in terms of increasing capacity within the Ministry, the MoE disbanded the working groups shortly after the project ended (MoE, 2008). Thus, the expertise built up in writing examination specifications and item writing was not properly cascaded or utilised. Moreover, despite the introduction of formal and informal moderation procedures, the move towards the new practices of continuous assessment put pressure on teachers, particularly those living and teaching less-populated areas of the country, to increase the scores of their students (Al Balushi & Griffiths, 2013).

Additionally, recent studies and reports have shown that the majority of grade 12 school leavers who have studied English language for 12 years and gone on to higher education lack the ability to communicate in the language appropriately and effectively in key language learning skills, whether in their social lives, academic studies, or at work (Al-Mahrooqi, 2012; Al-Mamari, 2012; Al-Mamari et al., 2018; Denman & Al-Mahrooqi, 2019; Moody, 2009; Oxford Business Group, 2013). For example, according to a report by the Oxford Business Group (2013) concerning new students who enrolled at Sultan Qaboos University (SQU), "just 14% of these students in 2011 achieved a pass mark in the English language test, and a considerable number perform less than satisfactorily in foundation programme placement tests". These findings are supported by Al-Mamari (2012), who found that more than 80% of students who completed grade 12 spend 6–12 months at SQU – the most prestigious tertiary institution in Oman – taking additional foundation course in English, because they failed to pass the exit test (equivalent to a score in the International English Language Testing System [IELTS] of 5.00). As the SQU generally admits higher achieving students, it is perhaps no surprise that the number who require intensive English language course increases to around 90% in some private universities and colleges in the country (Al-Mahrooqi et al., 2012).

The English language proficiency level of students also affects the awarding of scholarships granted yearly by the Omani government, which are only given to those well equipped to study overseas (Al-Issa, 2010; Al-Mashikhi et al., 2014; Al-Najar, 2016; Al-Seyabi & Tuzlukova, 2014; Ismail, 2011). Al-Mashikhi et al. (2014), in a study of grade 12 school leavers, revealed that more than 60% of students showed difficulties in comprehending class discussions and participating in class because they “are afraid of making mistakes” (p. 111). Also, high school leavers seeking employment opportunities with only a grade 12 diploma qualification often struggle to find suitable jobs that match their knowledge, skills and interests, especially in private institutions, due to their low ability in using the language (Al-Dhafiry, 2003; Al-Shaqsi, 2012). According to Al-Mahrooqi (2012), the major factors contributing to these poor English skills are “ineffective teachers, inadequate curricula, uninterested students, limited exposure to English outside the classroom, unsupportive parents, a poor school system, and peer-group discouragement” (p. 263).

Other researchers have attributed the low levels of English language proficiency in Oman to an examination-based system which encourages memorisation rather than actual learning, which seems to affect students’ motivation for learning (Al-Issa, 2007; Al-Issa & Al-Bulushi, 2010; Issan & Gomaa, 2010). Al-Issa (2007) reported that the English language examinations encourage a lot of reproduction of information and memorisation of information and there is a lack of focus on improving students’ learning or increasing their interest in learning. The study found that the main purpose for learning for most students was to memorise the content of textbook to pass the end-of-year examination and they had no clear idea of what they were expected to achieve at the end of their studies.

To address this situation and due to the importance accorded English language learning by the MoE, English language teaching has been recognised as a vital part of the curriculum. This was reflected in the English language curriculum development process, including the new English elective diploma curriculum, discussed below.

2.2.4 New English elective curriculum

In 2016, ten years after the implementation of the elective English curriculum based on “Reading & Writing Targets”, the MoE decided to change the elective coursebook. A report conducted by the MoE (2017) on the stakeholders’ (teachers and students)

perspectives on and experiences of the English elective diploma curriculum suggested that it was failing to prepare students to communicate in English because it did not include communicative language practice activities or cover the skills essential for success in the workplace or higher education. The MoE report further found that the linguistic demands of the coursebook did not match the expectations of Omani students (MoE, 2017). A committee was therefore formed to choose a new curriculum to reflect the needs of Omani students and improve their competency levels, providing them with the language skills required for the labour market and international examinations such as the Test of English as a Foreign Language (TOEFL) and International English Language Testing System (IELTS). The committee concluded its work by selecting a new imported curriculum entitled “*English Insights 2*”, published by Cengage National Geographic Learning. This was first implemented in the 2017/2018 academic year in grade 11, with “*English Insights 3*” being introduced in grade 12 in the 2018/2019 academic year.

Each stage of the new programme consists of three books (coursebook, workbook and teacher’s book) and comes with a set of accompanying materials (DVD, audio CD, digital copy of the three books and online resources). Unlike the previous textbook, the new coursebook aims to cover all four main skills (reading, writing, listening and speaking), test-taking strategies and study skills. The new materials provide students with the 21st-century skills needed to continue studying English after leaving school and to use the language as a tool for communication both inside and outside Oman. *English Insights* is a two-stage course designed in the UK for Omani students studying elective English in grades 11–12. Building on students’ knowledge of compulsory core English, *English Insights* helps students become confident speakers. According to the MoE (2017), some of the general aims of the new elective curriculum for grades 11–12 are as follows:

1. Strengthen students’ language skills, i.e. listening, speaking, reading and writing, in addition to grammar and vocabulary.
2. Develop students’ learning strategies, i.e. processes and actions to help students learn and use English effectively.
3. Develop students’ ability in academic writing in addition to the four basic types of writing.
4. Develop 21st-century skills, in particular creativity, ICT skills, problem solving and effective team-working.
5. Enhance students’ confidence in the use of English through presentations and projects.

6. Expose students to other cultures.
7. Develop students' ability and confidence in taking examinations.
8. Equip students for international examinations, such as IELTS, TOEFL, etc.
9. Integrate English with other subjects, in particular science, mathematics and technology.

Thus, the MoE, working with the higher education institutions in Oman, set out to revolutionise English language teaching over the course of a decade to meet the demands of the labour market. The new elective English language curriculum stresses CLT and is designed to focus on the development of 21st-century skills. The old “Reading & Writing Targets” curriculum was completely phased out across the Sultanate in 2016.

Following the introduction of the new elective curriculum in the academic years 2017/2018 and 2018/2019, the assessment document (SAH) was revised by the MoE to set out the new philosophy and aims of the elective diploma assessment and the skills and competencies required (MoE, 2022). The revised version of the SAH includes detailed guidelines and specific information about both formative and summative assessment techniques and procedures, marking, reporting and recording, examination specifications, weighting tables and glossaries. In the SAH (grades 11–12), the first versions of the new test for grade 11 were produced in January and June 2018, and those for grade 12 in January and June 2019. The topics assessed in the elective grade 12 examination include making notes, mind maps, dictionary skills, using visuals and summarising information. A comparison of the distributions of marks for both the continuous assessment and end-of-semester elements of the old and new grade 12 elective assessment systems is provided in Table 2.1.

Table 2.1. Distribution of marks for continuous assessment and end-of-semester tests

Elements	Grade 12 assessment – old elective			Grade 12 assessment – new elective		
	Continuous assessment	End-of semester test	Total	Continuous assessment	End-of-semester test	Total
Listening	–	–	–	–	20%	20%
Speaking	–	–	–	10%	–	10%
Reading	5%	20%	25%	5%	25%	30%
Writing	25%	50%	75%	15%	25%	40%
Total	30%	70%	100%	30%	70%	100%

In Table 2.1, the weightings in the right-hand columns reflect the key elements and content of the elective grade 12 course materials in *English Insights 3*. In both electives, 30% of the mark is awarded for continuous assessment, while the end-of-semester tests

account for 70% of the total mark. In the earlier system, only reading and writing were assessed, whereas under the new system, reading and writing are assessed using both methods, speaking is assessed through continuous assessment and listening is only assessed in the end-of-semester tests. Although reading and writing are both given significant weight in the examination, accounting for more than half of the marks, reading accounts for a much smaller percentage of the continuous marks compared to writing. Speaking is not included in either end-of-semester test for reasons of time and practicality (MoE, 2022). This may point to the reason for the lack of emphasis on developing oral skills in the classroom. It could also have implications for students' oral and interactive skills in higher education or when studying abroad, where these skills are essential. This reflects my experience working as an English language assessment officer in the MoE when I saw that common practice is for teachers to focus on the marks awarded for class participation as a main component in measuring students' speaking skills. The practice of "teaching to the test" in English language classrooms was also highlighted by the Ministry of Education and World Bank Report (2012, p. 31).

Building on knowledge developed in the core English curriculum, the English elective helps students become confident users of English, as defined by the Common European Framework of Reference (CEFR) B2 upper-intermediate band. The general outcomes assessed in the new elective curriculum for both continuous assessment and end-of-semester tests are outlined in Table 2.2.

Table 2.2. Key learning outcomes in the four main skills at grade 12 in the elective curriculum

Elements	Key learning outcomes
Listening	Can understand a variety of spoken texts
Speaking	Can interact with others
Reading	Can understand a variety of written texts
Writing	Can write and respond to formal letters and e-mails (interactive)
	Can write texts with the purpose of providing information (informative)
	Can write texts which express and justify opinions (evaluative)

(Source: SAH, MoE, 2022/2023)

As shown in Table 2.2, in the English elective curriculum for grade 12, the learning outcomes are grouped into four elements: listening, speaking, reading and writing, together with the requirements for students at this level. It should be noted that while the specifications for the new examination are based on the grade 12 curriculum, some of the outcomes of the new English elective in the textbook are not covered in the assessment specification (see Appendix A for the examination paper specifications and Appendix B for the textbook outcomes). With reference to writing skills, for example, the writing

outcomes in the textbook introduce different types of writing texts, such as analysing the essay question, describing data, summarising data, letters/reports and email writing. These are divided into four main types: evaluative, informative, narrative and interactive. In the new examination specification, the assessment of writing skills only includes two text types per semester rather than the four introduced in the textbook. Specifically, the writing skills covered for the first semester consist of an opinion text (evaluative) and description writing (informative), whereas the second semester consists of report writing (informative) and formal letter writing (interactive). This indicates that the narrative text type is not assessed in the final examination, even though it is included in the textbook. Moreover, the text types assessed in the examination are not aligned with the specifications for continuous assessment.

When these changes in the examination specifications were announced, some teachers had concerns about them, while others agreed with them (MoE, 2018/2019). Those teachers who were concerned argued that the content of the new curriculum was time consuming and it took a great deal of effort to prepare their students to cope with the demands of English at tertiary level. They believed that teaching should not emphasise material not tested in the final examination. Those who agreed with the changes believed that the new specifications would help address the worrying deficiency in student's English language proficiency after leaving grade 12. Al Mamari et al. (2018) showed that the poor performance of students in English language was due to the examination design, since the examinations were not aligned with the pedagogic developments in communicative teaching and learning and the input lacked authenticity. Hence, there is a general feeling of distrust regarding the new elective assessment reform and scepticism about its impact on teaching and learning. Thus, it seems the intended washback of the new examination system on teachers and their teaching approaches in the classroom is currently unattainable.

2.2.5 English language teaching

In Oman, English is taught as a foreign language in all public schools. There are 180 days of instruction per year and the teaching time has been extended from 4 hours per week to more than 6 hours per week. English is now delivered through five to six lessons a week in grades 1–4 and grades 5–10, with each lesson lasting 40 minutes. There are four lessons per week in grades 11 and 12, including Elective English, which totals around 3–4 hours of English language lessons per week.

The successful implementation of the basic education reform is largely dependent on the quality of the teaching staff. Taken together, the new vision and aims of the reforms introduced in Oman since the beginning in the 1980s represent a significant change in emphasis and constitute a major challenge for the teaching profession. The previous system was dominated by a traditional teacher-centred approach, but now the teacher's role has become much more complex and demanding. Teachers are expected to deploy appropriate classroom techniques and strategies moving away from whole class instruction, which rewarded rote learning, to a more student-centred approach emphasising autonomous and cooperative learning, problem-solving, critical thinking, creativity, synthesis and investigative research enquires. They have had to adapt to the new learning outcomes in the curriculum, which aim to reduce barriers and manage individual differences in students' learning, while also increasing student participation. Teachers are responsible for acquiring the new teaching strategies required to identify what their students know and can do and what actions they must take to improve students' learning.

To ensure the successful implementation of student-centred techniques in the classroom, class sizes in all grade levels were reduced significantly, with a maximum class size of 30 students being stipulated (Al-Issa & Al-Bulushi,2012). Several reforms have also focused on upgrading teachers' qualifications and supporting their professional development, with a view to improving the quality of provision and students' attainment. An in-service training programme was implemented to inform teachers of the substance of the new curriculum and the methods for teaching it. The MoE has also adopted a cascade model of in-service training for teachers in areas supporting the reforms. Supervisors are identified and brought to the Ministry for training, from where they go on in turn to train teachers in their region. Training for English teachers is currently offered in all the governorates.

The rapid expansion in school numbers at the beginning of the 1980s meant that many English language teachers were appointed with diploma qualifications. However, by 2008/2009, 83% of teachers who taught English in government schools had a Bachelor's (BA) degree in Education (MoE, 2009). In cooperation with the University of Leeds in the UK, the MoE also sought to upgrade the qualifications of Omani teachers from diploma level to a BA in Teaching English to Speakers of Other Languages (TESOL). The programme started in 1998 and ended in 2010, producing around 1,060 qualified

teachers (Al-Mahrooqi, 2012). Various other reforms have also focused on upgrading teachers' qualifications and supporting their professional development. The Main Training Centre, a department of the Directorate General of Human Resource Development, is responsible for implementing training programmes and supervising all training in the local centres in the 11 governorates.

Professional development plans were developed for all the parties involved, including teachers, supervisors, school administrators and other members of the MoE. In 2014, the MoE established the Specialized Centre for Professional Training of Teachers (SIPTT), which aims to provide professional learning for English language teachers and support the implementation of the English Programme (Al Shabibi & Silvennoinen, 2018). It has recently started training English teachers through two programmes: the English Programme and the New Teachers Programme (Jabri et al., 2018). These are long-term, accredited programmes, delivered both centrally and in some of the regions, lasting for one year (New Teachers Programme) or two years (English Programme). Completing the New Teachers Programme is a requirement for all new English teachers with one to three years' experience. The English Programme focuses on methodology and developing language proficiency and is aimed at all those interested in teaching English. The MoE also provides English teachers with electronic access to the international professional journal, *English Teaching Professional*, and encourages them to attend workshops and local and international conferences.

Although the MoE states that teachers are expected to change their classroom teaching approaches to fit the new system, there is evidence that many teachers still use traditional approaches and have experienced difficulties in making the hoped-for transition (Al Balushi & Griffiths, 2013; MoE & World Bank, 2012). Thus, teachers have continued to adopt a teacher-centred approach and have made relatively little use of collaborative teaching and learning methods. Other studies related to English language teaching in Oman have found that Omani teachers lack the required linguistic and methodological competencies (Al-Issa, 2010; Al-Issa & Al-Bulushi, 2011).

A review of teacher education in Oman conducted by Al Balushi and Griffiths (2013) reported that the educational experience of Omani teachers themselves – as school students or in higher education – was of “a teacher centred didactic approach to teaching and learning” (p. 168). Although the Ministry provided some in-service programmes on

student-centred teaching, these were usually short, developed independently of the teachers' own classrooms and did not encompass real-life settings in training. Additionally, the report also stated that the Ministry had largely failed to get across its message about the desirability of its reforms, including the move to a student-centred approach.

At the time of the so-called "Arab Spring" of 2011, one of the main requirements for English language teachers in Oman was to have an overall IELTS score of at least 6.0. However, many Omani teachers who had graduated from local or international institutions could not achieve the required IELTS score, despite having a degree in English language teaching, and consequently could no longer be employed (Al-Balushi, 2017). This had an impact on hundreds of male and female Omani teachers, who had to wait for many years without getting a job. In February 2011, during the "Omani Spring" period, thousands of Omanis, including those ELT teachers who were not employed, protested in the streets, complaining about the high salaries for certain sectors and demanding a solution to unemployment. One of the government responses was to employ around 50,000 Omanis within a month, including all the ELT teachers who had been waiting years to be employed.

However, MoE stakeholders were aware that those newly appointed teachers were not sufficiently qualified to start teaching English and needed up-to-date training in teaching skills and methodologies (Al-Balushi, 2017). Despite this training, Al-Balushi (2017) found little change in these teachers' understanding of the knowledge and skills required to teach English, and no noticeable impact on their classroom teaching after attending courses. There was recognition early on that it would be difficult to persuade teachers and students to change their practices in the classroom and adopt new materials and methods without a change in the assessment system (Al-Balushi, 2017; Al Balushi & Griffiths, 2013).

Moreover, teacher workload is among the contextual factors that may indirectly influence how teachers differently perceive the level of impact of examination reform on their teaching practices (New Zealand Consortium Report, 2013; World Bank Report, 2012). Several variables affect teacher workload, including the number of teaching lessons assigned, number of classes assigned, class size and administrative duties. Teachers are responsible for their direct teaching load alongside a mix of non-direct teaching

responsibilities that are part of being an effective teacher. In the Omani educational context, a student-learning week consists of five days, with up to eight lessons taught per day, comprising a total of 40 lessons per week (each lesson is 40 minutes in length) over a 180-day academic year. This means that a teacher will have a workload of 672 hours of teaching time per year. Within their teaching plan, teachers have various responsibilities: teaching a certain number of lessons, marking follow-up work, undertaking substitution classes as required for absent teachers, attending to administrative duties, attending meetings, and undertaking professional development opportunities. In addition, class size has an impact on teacher workload and an average class size is 26. Teachers need to mark students' work individually and provide feedback to each student and the parents. Teachers are also required to complete various administrative duties on a weekly and semester basis. These include attending assemblies each week, students' management duties (grounds, canteen, bus), activity classes, subject teachers' professional development meetings and others (such as e-portal reporting duties, parent meetings and report writing).

Furthermore, evaluation studies conducted in this context have reported that one of the main challenges that teachers face is the overcrowded nature of the curriculum and the short academic year in Oman. Research has found that the curriculum has too many outcomes to be covered in the time available bearing in mind the depth required for students to achieve learning outcomes through the different grade levels (see Al Shabibi & Silvennoinen, 2018; CFBT/NFER, 2012; New Zealand Consortium Report, 2013; World Bank Report, 2012).

Thus, there is an urgent need to gain further insights into the context of the elective diploma examination reform to establish whether it has been effective in helping teachers improve their teaching practices and whether it has triggered additional factors that could affect the educational system.

2.3 Chapter Summary

This chapter has provided detailed information on the Omani education system. It first discussed the stages of development of the education system in Oman and its evolution. It then outlined the reform of English language textbooks and assessment in Oman. It has also presented the general aims and objectives of the new English elective curriculum, as well as discussing the distribution of marks in the final examination and the continuous

assessment component of the new English elective. The key learning outcomes for the four main skills have also been highlighted in this chapter. The chapter concluded with a description of the English language teaching context in Oman.

The next chapter reviews the relevant literature and outlines the theoretical framework that underpins the study and contributes to the analysis of data and discussion of the results.

Chapter 3. Literature Review

3.1 Introduction

This chapter offers a thorough examination of the existing literature and discourse surrounding the washback effect, a comprehensive understanding of the washback concept and its related terminologies, and empirical findings, assumptions and perspectives that define how the process of the washback phenomenon works. Furthermore, this review aims to analyse patterns and find gaps in the existing literature, positioning this research and its contribution within this field of study. Critical works related to the washback phenomenon and its effect on teaching practices were sought using the SCOPUS database, key books related to this topic and Google Scholar. The NVivo qualitative data analysis software was used to help code and organise the reviewed literature.

This chapter begins by examining the concept and scope of washback, as well as discussing its various definitions. It also examines the types and characteristics of washback and the directions it can take, both positive and negative. Afterwards, the focus shifts towards examining the conceptual frameworks of washback that have influenced and guided research on washback both broadly and in this investigation specifically. The chapter also covers the relevant baseline theories on washback phenomena, examining the extensive range of existing empirical studies to help understand the research problem and relate it to previous work in this area. This study was interested specifically in exploring how tests can affect teachers' perceptions. Additionally, it delves into the intricacy of washback by examining the impact of tests on teaching methods. The other part of this study examines what teacher-related factors, in addition to the test itself, may influence teachers' practices in the classroom. This chapter ends with an examination of the research gaps and a recap of the key discoveries, highlighting implications for the remainder of this study.

3.2 Washback: Definition and Scope

“Washback” is a term that has been used by many education assessment experts over the past few decades (Alderson & Wall, 1993; Bailey, 1999); it is also sometimes known as “backwash” in general education (Hughes, 2003). Although the concept of washback is now widely accepted in the educational and applied linguistics literature, there have been relatively few empirical studies demonstrating its existence or how it functions. This issue

was emphasised more than two decades ago by Heaton (1990) and others (Alderson & Hamp-Lyons, 1996; Alderson & Wall, 1993; Shohamy, 1993; Wall, 1999) and has been even re-emphasised more recently by Alderson (2004) and Wall (2005). Alderson (2004), in the Foreword to Cheng et al.'s (2004) edited volume, *Washback in Language Testing*, was still only able to point to “a slow accumulation of empirical research” (p. ix).

According to Hamp-Lyons (1997), the concept of washback has been applied in various fields, such as general education, language education and language testing, to describe the interrelation between testing and teaching and learning. The idea of washback has been widely explored in language education and testing. Research from as far back as the early 1970s and 1980s (Kirkland, 1971; Madaus & McDonagh, 1979; Rudman et al., 1980) showed a clear understanding of the impact that standardised testing could have on both schools and students. Since their introduction in ancient China in 206 BC, examinations have been used as a means of selecting and certifying candidates for government service (Gipps, 1999), encouraging school performance, reducing nepotism and placing limits on corruption and patronage, allocating limited places within higher education, and controlling curricula (Eckstein & Noah, 1992). Examinations are still widely acknowledged in the field of applied linguistics as exerting a powerful influence on teaching and learning (Biggs, 1995; Cheng, 2002; Cheng & Curtis, 2004; Messick, 1996; Pearson, 1988; Rea-Dickins & Scott, 2007; Wall, 2005; Wall & Alderson, 1993).

The upcoming sub-section (3.2.1) discusses the debate in the scholarly discourse concerning whether the term “washback” can be substituted with “impact”. It should be noted that both terms have been used in this thesis to redefine the influence of testing on teachers. This relates to what Shohamy (1992) referred to as “the legislative power and authority of tests” and to the “policy tools; and societal signalling devices” described by Madaus and Clarke (2001). In the education literature, the term “washback” has been conceived, generally, as any effect or impact of the testing system on the teaching and learning process, and this can relate to teachers, learners, parents, administrators, textbook writers or classroom practices (Buck, 1988; Chen, 2002; Cheng, 2005; Gates, 1995; Hughes, 2003; Khaniya, 1990; Messick, 1996; Spratt, 2005; Taylor, 2005; Wall & Alderson, 1993; Wesdorp, 1982). According to Messick (1996), for example, washback refers to the extent to which a test causes teachers and learners “to do things they would not otherwise do that promote or inhibit language learning” (p. 241). Alderson and Wall (1993) highlighted that there is evidence that the washback effect from the introduction

of tests is linked to changes in teachers' attitudes and behaviours, which can have important educational consequences. Bailey (1996) discussed washback as a process that links testing to classroom teaching and students' learning, a view that is widely accepted and considered significant; however, few empirical studies have examined it in depth. Bailey further suggested that there are also concerns regarding the processes used to promote the positive washback and inhibit its negative effects, as well as the criteria used to differentiate between these two types of washback. This is precisely the context of this research, which addressed a new exam – the English elective diploma end-of-year examination – introduced to bring about a positive washback effect on teachers' attitudes and behaviours (see Chapter 2).

An important feature that merits specific consideration here is the relationship between washback generated by administering a new or revised examination and test validity. As Messick (1996) suggested, “for optimal positive washback there should be little if any difference between activities involved in learning the language and activities involved in preparing for the test” (p. 241). However, washback effects may only provide circumstantial evidence in terms of test validity, “in that a poor test may be associated with positive effects and a good test with negative effects because of other things that are done or not done in the educational system” (Messick, 1996, p. 242). Alderson and Wall (1993) have also argued that the complex nature of the washback phenomenon means it cannot be used as a standard for evaluating the validity of a test.

For Hamp-Lyons (1997), the consequences of tests:

...can be sought in classrooms, in teachers' and learners' behaviours, and in textbook materials. They can also be sought more widely by looking to the uses to which learners put the language outside classrooms, to their attitudes to the language and to the role of the language in their society. (p. 299).

Therefore, although the washback effect pertains to the impact of a test on classroom activities (e.g. teaching and learning), other elements of educational practice also contribute to the effect of testing on teaching and learning. Therefore, the term “washback” is used in this study to refer to the intended influence on the teaching of a new curriculum introduced by changes in the examination content and format. It is necessary to remember, however, that even if washback is employed to temper the power of testing and for the benefit of the educational process, with the ultimate aim of enhancing teaching practices and the quality of learning, unintended side effects may still

arise, since a successful implementation of examination reform is never a neutral process and is always accompanied by consequences. In this respect, key stakeholders should make such consequences “as constructive as possible, particularly for those who are assessed” (Stobart, 2003, p. 140). The following sections discuss the terminology used in relation to the washback phenomenon and how previous research has explained the way it operates.

3.2.1 Washback and terminology

Although the term washback is now widely acknowledged in language teaching and testing, there has been much debate in washback research about what Andrews (1994) has called “nomenclature” (p. 67). Many other terms have been used and sometimes equated with “washback”, including “backwash” (e.g. Hughes 2003; Prior & Lister, 2004), “impact” (Bachman & Palmer, 1996; Wall, 2005), “curriculum alignment” (Madaus, 1988; Smith, 1991), “measurement-driven instruction” ([MDI], Popham, 1987; Shohamy, 1992), “systemic validity” (Frederiksen & Collins, 1989), and “consequential validity” (Messick, 1996). As noted by Hsu (2010), there are often different aspects of the same phenomenon implicit in such terminology; thus, it is important to differentiate between the definition of washback used in this study and other similar terms found in the literature. However, discussing all the labels used to describe the phenomenon is beyond the scope of the thesis, so the focus here is on only the most common alternative terms in the literature, “backwash” and “impact”.

According to Prodromou (1995), some reference books do not refer to the phenomenon of washback at all (e.g. Richards et al., 1986), or describe it as not worth an entry (Seaton, 1982). There has long been a conceptual dispute in the literature over whether to refer to it as “backwash” or “washback” (Biggs, 1995; Hughes, 2003; Prior & Lister, 2004; Wesdorp, 1982), even though Alderson and Wall (1993) argued that there is no reason, either semantic or pragmatic, for favouring either term. While it is true, as Spolsky noted (1994), that the word “washback” hardly appears in standard language resources, the word “backwash” is found in some prominent dictionaries (e.g. Merriam-Webster, 2000; Oxford University Press, 2000). It is always used, however, with negative connotations when referring to consequences. Some researchers seem to accept both terms, arguing that “backwash” is commonly used in general education to describe this phenomenon, while the favoured term in the discipline of language testing and applied linguistics is “washback” (Alderson & Wall, 1993; Wall, 2012). Rea-Dickins and Scott (2007) contend

that this is the result of a lack of communication between language and education researchers. Alderson (2004) argued specifically that both terms can be used interchangeably to refer to the same phenomenon: “to clarify the distinction between the terms backwash and washback: there is none” (p. xi). Other researchers have also supported the interchangeable use of the two terms (e.g. Davis et al., 1999; Hughes, 2003). Nonetheless, most language testing and applied linguistics researchers appear to have favoured “washback” over “backwash” and this study follows the same approach, not least because if both terms are used synonymously, there will always be a risk that the negative connotations associated with the one term (“backwash”) could carry over to the other. Thus, “washback” is used throughout this thesis, rather than “backwash”.

There is also debate concerning whether the terms “impact” and “washback” refer to the same concept. Several researchers in language testing (Bachman & Palmer, 2010; Hamp-Lyons, 1997; McNamara, 2000; Shohamy, 2001) have distinguished between “impact” and “washback” as relating, respectively, to “macro” and “micro” levels of influence. The macro level refers to social and institutional factors, while the micro level relates to the individual stakeholders (primarily teachers and students). For Bachman (2000) and Shohamy (2001), washback is located under the umbrella term of impact. Thus, whereas washback is restricted to the limits of the classroom setting (Hamp-Lyons, 1997; McNamara, 2000), the concept of impact also encompasses the wider levels of the educational system and society, where decisions about school admission and employment are taken based on test scores (Bachman, 2000). In addition, the term “impact” embraces the effects of testing on individual participants beyond the context of the learning programme. Hawkey (2006) suggests that an impact study could examine the effects of a programme or test on school heads, parents, administrators of receiving institutions and high-stakes test providers.

Some scholars have placed the terms “washback” and “impact” within the theoretical framework of consequential validity. This framework considers the broader factors or evidence contributing to test validity, including the social implications of interpreting and utilising test scores (Messick, 1996). Although various researchers have used various terms to indicate the same phenomenon, this study did not strictly differentiate between test impact and washback. Rather, it is assumed here that both terms, washback and impact, refer to the same phenomenon. It is challenging to establish limits for the impact that tests can have due to the complexity of the washback effect and the varying factors

that can influence the washback process in a given context. It is not even possible to focus on the influence of testing only at the micro level and disregard any issues relating to the macro level. The interrelationships between the different levels of the phenomenon cannot be ignored, as will be discussed in greater depth later in this chapter.

According to Rea-Dickins and Scott (2007), it is better to consider the terms “washback” and “impact” as part of the same phenomenon in educational research. They argue that both terms refer to the effects or consequences of testing, even though they are often discussed separately in the literature. Andrews (2004) adopted a broad interpretation of the term “washback” as encompassing the various influences of tests not only on the classroom environment, but also on the educational system and its various stakeholders. In doing so, Andrews used the term “impact” in a non-technical sense, as a synonym for “consequence” and “influence”. Hawkey (2006) also acknowledged that “impact” has been used as a non-technical term and that even most researchers who differentiate between test impact and test washback still use both terms to refer to the same concept. An example of these researchers’ work is Wall’s (2005) study, “The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory”.

This discussion of washback has outlined the concept and the associated terminology. I turn now to how washback can occur.

3.3 Washback: Positive or Negative

The late 1980s saw several influential studies published on test impact. Some of these advocated the use of high-stakes testing to promote positive improvements in the educational system (Frederiksen & Collins, 1989; Popham, 1987), while others were either more cautious in their views of how tests might direct the instructional process (Airasian, 1988), or totally rejected the idea that testing can be beneficial (e.g. Fish, 1988; Madaus, 1988; Nobel & Smith, 1994), as had earlier works (Davies, 1968; Vernon, 1959).

3.3.1 Positive washback

Some researchers strongly believe that high-stakes testing can lead to beneficial change in language teaching, so-called “positive washback” (e.g. Airasian, 1988; Alderson & Wall, 1993; Fournier-Kowaleski, 2005; Pearson 1988; Popham 1987). In this scenario, school stakeholders, such as teachers and students, exhibit beneficial reactions and attitudes due to test influence and work willingly towards the purposes of the test. For

Fournier-Kowaleski (2005), discussing examinations designed to develop English language proficiency and language acquisition, positive washback would involve the incorporation in classroom teaching of activities connected to these aims, for example a conscious focus on task-based activities relating to real-world situations in which students are asked to use the target language, such as letter writing, reading and listening to factual information. As Pearson (1988) put it, good tests will not only promote effective teaching and prompt productive learning, but also “be more or less directly useable as teaching-learning activities. Similarly, good teaching-learning tasks will be more or less directly useable for testing purposes” (p. 107). To Pearson, the washback effect of a test is determined by its potential positive and negative influences on teaching. He believed that test washback would be positive if it managed to bring about the entire spectrum of desired changes and negative if its effects did not match the curriculum objectives to which it was linked.

Popham (1987) suggested using high-stakes tests to bring about beneficial changes in the educational process and proposed the concept of measurement-driven instruction (MDI) to support this idea. Popham categorised examinations into two main types. The first type comprised examinations with significant consequences for the individuals taking them, for example determining whether the student can advance to the next grade level. The second type comprises examinations the scores of which are seen as a measure of effective teaching and school progress. For example, nationwide achievement tests are reported by the media on a school or district basis. For Popham, high-stakes tests allow teachers to focus a significant portion of their teaching content on the objectives assessed by such tests and therefore serve as “a powerful curricular magnet” (p. 680). Popham rejected the contention that MDI leads to curricular stagnation and reductionism, or that it lowers the aspirations of students and constrains teachers’ creativity. He argued that these criticisms are specious because they lack both analytic rigour and empirical evidence. Frederiksen and Collins (1989) also supported the argument for the positive potential of high-stakes testing. Their position was based on the premise that the educational system adjusts its curricular and classroom teaching practices to achieve educational goals and that one of these goals is maximising the scores in any tests that are used to evaluate outcomes. Thus, for example, if the reading component in the examination assesses phonics, this will become the curriculum focus in terms of reading skills.

Despite the potential benefits of a test in guiding instruction, some researchers remain sceptical. Airasian (1988) challenged the validity of MDI, arguing that various factors can affect the process and lead to inconsistent outcomes. A test's impact on teachers' classroom practices depends on several factors, including the test's importance, difficulty and the quality of the teaching content it evaluates. Airasian compared high-stakes and high-standards testing programmes in Europe, such as English O and A levels, to low-stakes and low-standards examinations in the United States (US), concluding that the most significant impact on instruction occurs when high standards and high stakes are present. He suggested that although it is important to consider whether a test measures key skills at a lower or higher level of cognitive achievement, the inclusion of higher-level operations may not necessarily improve learning because there are other factors that relate to the quality of teaching in an educational system. According to Airasian (1988), those who drive MDI programmes are typically state-level educational administrators, legislators and psychometricians. These individuals may be unaware of the limitations and obstacles that teachers face in actual classroom environments. It is thus important that stakeholders explain what they mean when they use terms like higher-level cognitive outcomes and consider carefully whether the existing teaching resources give adequate guidance on the development of the desired outcomes and knowledge. Although Airasian's (1988) arguments show the complex nature of test impact on teaching, they lack empirical support.

3.3.2 Negative washback

Some researchers have criticised examinations for causing negative effects on teaching and learning, so-called "negative washback" (Alderson & Wall, 1993, p. 115; see also, Davies, 1968; Fournier-Kowaleski, 2005; Vernon, 1956). Vernon (1956) commented that teachers tend to focus solely on the skills and activities related to the examination format and content and claimed that examinations "dominate and distort the whole curriculum" (p. 198). For Davies (1968), examinations often have a bad effect because they "have become teaching devices; work is directed upon what are – in effect if not in fact – past examination papers, and consequently becomes narrow and uninspired" (p. 125). Alderson and Wall (1993) referred to the undesirable influence on learning and teaching activities of a particularly "poor" test (p. 117).

Typically, when referring to a test as "poor", it suggests that the teachers and students are not motivated to teach or learn the material that is expected to be tested at a particular

grade level. These tests do not establish a connection between teaching and learning methods and the curriculum objectives they are supposed to be associated with. According to Fournier-Kowaleski (2005), negative washback results from including exam-related activities in classroom teaching that hinder or fail to promote the learning of the language or the development of essential skills, such as reading, listening, writing and speaking. This might involve, for example, a conscious focus on vocabulary tasks to prepare students for specific questions in the test at the expense of other skills. In similar terms, Prodromou (1995) referred to negative washback as “overt backwash”, which involves “teaching to the test”. For Prodromou, overt backwash occurs when teachers tend to do a lot of revision of past examination papers or replicate the task types that will appear in the real examination drawing on past papers and textbook exercises. Many language instructors are likely familiar with the examples he provides of washback, such as using a “text + questions” approach, incorporating language snippets that are not authentic and focusing on skills that are easier to evaluate in terms of administration and marking, such as reading rather than speaking.

Fish (1988) found that one negative effect of tests on teachers’ behaviour was caused by poor student performance and reported that experienced teachers showed a lower level of anxiety and accountability pressure than teachers who lacked teaching experience. Nobel and Smith (1994) noted that tests can sometimes have an unintended negative impact that goes against the intentions of those seeking to enhance education. They also observed that focusing solely on test-taking strategies and the test format may result in higher test scores but not necessarily a deeper understanding of the subject matter.

Additionally, an extensive investigation carried out in elementary schools by Smith (1991) identified some serious negative impacts of testing on teaching practices. The study revealed that the curriculum had been drastically narrowed and teachers had less ability to create, modify, or innovate in their teaching. The teachers involved felt themselves pressed for time, pressurised by their principals’ expectations and in danger of losing control of their local curriculum. Most of them gave in to these pressures, but those who resisted knew there would be a price to pay for such resistance.

Madaus (1988) is another example of an educator critical of the MDI phenomenon. He predicted only negative effects on teaching, the curriculum and learning if testing were used as “the engine, or primary motivating power of the educational process” (p. 84). For

Madaus, MDI leads to cramming or a narrowing of the curriculum, concentrates on skills that are more related to the test format and “constrains the spontaneity of teachers and students, and demeans the abilities of professional teachers” (p. 85).

Other researchers have supported Madaus’s concerns about the potential deleterious effects of examinations. Haladyna et al. (1991) proposed the concept of “test score pollution” to refer to “factors affecting the truthfulness of a test score interpretation. Specifically, pollution increases or decreases test performance without connection to the construct represented by the test” (p. 4). Hence, the pressure to improve students’ test scores results in practices which pollute the interpretation of test scores. Three main factors were found to contribute to test score pollution: (1) the methods schools use to prepare students for testing, (2) the administration of the test, and (3) external factors beyond the school's control, such as student anxiety and motivation levels, as well as their proficiency in the language used for the examination. Only the first of these sources of polluting practices is relevant to this study. Haladyna et al. (1991) discussed a number of activities aimed at preparing students for a test, which they placed on a scale ranging from “ethical” through to “highly unethical” (p. 4). An example they gave of ethical test preparation practices was motivating students to perform well in the examination by discussing the importance of tests with them, while an example of unethical practice was presenting materials in the classroom which directly matched the test. The researchers commented that unethical or highly unethical practices which inflate test scores are actually common in schools. They also found that in schools where the test scores were rising, teachers were more likely to feel pressure from parents and the community than in schools where the test results were stable or falling.

Other researchers have shown directly opposing views concerning the washback effect. Wiseman (1961), for example, limited the concept of washback to “the deleterious effects of examinations” (p. 159), and criticised the paid coaches outside schools who prepared students for examinations the results of which were published in public. He argued that such special preparation classes were not worth the time, effort, or money because the students were practising examination requirements rather than focusing on learning. According to Heyneman (1987), however, many proponents of high-stakes testing view coaching for tests not as a drawback but rather as a virtue. In this regard, the study suggests that the increased number of coaching classes may prepare students for examinations, but what students learn is more about test-taking strategies and information

about the test itself rather than language learning. Thus, if tests fail to reflect what they are intended to accomplish, they will generate a negative washback effect on both teaching practices and students' learning, as suggested by Pearson (1988).

To summarise, the above discussion has reviewed key studies which have focused on the influence of high-stakes testing on teaching. Some of these studies discuss the washback effect of specific tests in specific situations (Heyneman, 1987; Smith, 1991; Wiseman, 1961), but the majority aim either to review the literature for general tendencies (Alderson & Wall, 1993; Pearson, 1988) or to discuss particular issues relating to the educational system (Airasian, 1988; Haladyna et al., 1991; Madaus, 1988; Popham, 1987). What all of these studies have in common is an acknowledgment of the existence of the phenomenon and the importance of washback effects in relation to teaching and learning. Where they differ is in their view of the nature of this influence: some researchers believe that tests can have a positive washback effect (Frederiksen & Collins, 1989; Popham, 1987), while others believe that the washback from testing is more likely to be negative. There is considerable support in the literature for the idea that washback from testing will always be detrimental (see Black & William, 1998; Crooks, 1988; Shohamy, 1997). Thus, given the setting of this study, which introduced a new examination as part of an intentional attempt to promote the principles and practices of the CLT approach, the way in which a test is designed and used, particularly if its content and procedures promote good teaching practices, can determine whether the test's washback is positive or negative (see Bailey, 1999; Barnes, 2017; Taylor, 2005).

To fully understand the effects of testing on education, research must go beyond simply labelling the impact positive or negative. It is crucial to examine the various complex factors that play a significant role in a test's washback on teaching and learning. Additionally, the educational system in which the test is implemented should be taken into consideration. If the test is associated with important consequences for teaching and learning, the context of that test should be examined. The extent to which the washback of testing is positive or negative will largely depend on how washback operates in practice and the educational system within which it is introduced. To promote positive effects on teachers' classroom behaviour and students' learning practices, it is essential to understand the concept of washback and how it operates. Therefore, the following section examines how prior studies have elucidated how washback functions.

3.4 Mechanisms of Washback

While the work of certain key researchers serves as a basis for explaining the mechanisms of the washback phenomenon (e.g. Shohamy, 1992; Wood, 1996), it was Alderson and Wall (1993) who first referred to the complex nature of washback in their paper, “Does washback exist?”. The complexity of the washback phenomenon was a focus for research in the following years and more than a decade later, researchers (e.g. Burrows, 2004; Cheng, 2005; Watanabe, 2004) were still emphasising that washback is far from direct and simple. Alderson (2004) reasserted that “the phenomenon is a hugely complex matter, and very far from being a simple case of tests having negative impact on teaching” (p. ix). If we consider washback to be a process that can positively impact teaching and learning and accept that it can be both positive and negative, it is worth striving towards a positive outcome. This has been suggested by various authors, such as Pearson (1988), Buck (1988), Shohamy (1992), Watanabe (2004), and Cheng and Curtis (2004), among others. Hence, the issue becomes more methodological in nature, that is, “how to bring about this positive movement” (Cheng & Curtis, 2004, p. 8). Alderson (2004) argues that the research focus now is not about whether or not washback exists, but rather: “What does washback look like?”, “What brings washback about?”, “Why does washback exist?”, and “Why does washback have the effects it does have?” (p. xi). Various researchers have sought to discuss the complex process of washback, putting forward several different models to illustrate the way washback might function (Alderson & Wall, 1993; Bailey, 1996; Green, 2007; Hughes, 1993; Shih, 2007; Watanabe, 2004; Zhan, 2009). This section reviews and critically reflects on three seminal models for washback phenomenon, namely Alderson and Wall’s (1993) 15 washback hypotheses, Hughes’ (1993) trichotomy of backwash model, and Bailey’s (1996) basic model of backwash. Additionally, this review includes two other washback models related to this study: Green’s (2007) model of washback (covering direction, variability and intensity) and Watanabe’s (2004) multi-dimensional washback model.

3.4.1 Alderson and Wall’s (1993) hypotheses

Alderson and Wall (1993) proposed 15 washback hypotheses, which marked a major advancement in the rigorous study of the washback phenomenon. The hypotheses proposed by these authors served as the basis for clarifying the washback concept and driving future research. In their empirical study conducted in Sri Lanka, the authors used an ethnographic approach to obtain accounts of participants’ attitudes and perceptions concerning the phenomenon through interviews, as well as conducting test analysis and

classroom observation in seven schools based in different parts of Sri Lanka over the course of two years.

Alderson and Wall suggested that the phenomenon is not as straightforward and immediate as it may initially appear. Showing their discontent with other researchers' claims that good tests are sufficient in themselves to bring about change, these researchers took the discussion further by examining both the existence of the washback effect and the ways in which it might function. They expanded this latter aspect of the concept of washback by identifying different possibilities. For Alderson and Wall (1993), in line with some of the views discussed in the previous section on terminology, washback was restricted to the micro level of the classroom and did not encompass the concept of "impact" operating at the macro level. They strongly emphasised that intended washback effects, whether positive or negative, are "independent" of test quality and any claims of a direct relationship between the test in question and its influence are unjustified (p. 118). Rather, it is more accurate to assume that a test's impact on teaching is generally determined by many factors. Thus, a well-designed test can have a positive or negative impact on the behaviour of both teachers and students and equally, a poorly designed test may lead to beneficial washback if it is used to enhance student learning. In view of this complex interaction, Alderson and Wall (1993) proposed a list of 15 hypotheses relating to the washback phenomenon, each representing a possible factor that might affect the way in which washback operates. One way of looking at these hypotheses, as suggested here, is to divide the washback effects into different areas, with one area geared towards teaching, another to learning, another towards attitudes, and the others classified as washback on test stakes and test stakeholders. Table 3.1 illustrates the suggested classification.

On the basis of the hypotheses, one might suggest that Alderson and Wall (1993) paid considerable attention to the existence and importance of washback and introduced a rather comprehensive definition of the concept. Their hypotheses comprise a broad definition of washback, covering the various possible relationships and potential consequences that can lead to a washback effect beyond the test itself. They also refer to possible factors relating to the test under investigation. The authors separated their hypotheses about what teachers teach, why teachers teach, and teacher attitudes to identify the three main separate aspects of a test's influence. They also emphasised the issue of washback variance. This variance explains that washback is not linear or simple

in that its effect is not the same on all stakeholders. In other words, washback on teaching might have different effects from one teacher to another. This idea is reflected in more than one hypothesis (e.g. hypotheses 14 and 15 related to test stakeholders).

Table 3.1. Proposed classification of Alderson and Wall's (1993) washback hypotheses

Hypothesis	Aspect of impact
1. A test will influence teaching.	Teaching
2. A test will influence what teachers teach.	
3. A test will influence how teachers teach.	
4. A test will influence the rate and sequence of teaching.	
5. A test will influence the degree and depth of teaching.	
6. A test will influence learning.	Learning
7. A test will influence what learners learn.	
8. A test will influence how learners learn.	
9. A test will influence the rate and sequence of learning.	
10. A test will influence the degree and depth of learning.	Attitudes
11. A test will influence attitudes to the content, method, etc. of teaching and learning.	
12. Tests that have important consequences will have washback.	Test stakes
13. Tests that do not have important consequences will have no washback.	
14. Tests will have washback on all learners and teachers.	Test stakeholders
15. Tests will have washback effects for some learners and some teachers, but not for others.	

(Alderson & Wall, 1993, pp. 120–121)

However, although references to the complex and non-linear nature of the phenomenon are emphasised throughout these hypotheses, “the test” is still the main factor that triggers the influences suggested in each hypothesis. Thus, although Alderson and Wall’s 15 hypotheses provide a clear expression of the possible influences of examinations on teaching and learning, further examination of how these hypotheses might interact or work against each other in practice is still called for. Alderson and Wall (1993) themselves indicated that their hypotheses were not intended to be exhaustive, suggesting that further research was required to increase the “specification of the Washback Hypothesis” (p. 127); indeed, their hypotheses can help in developing other hypotheses, not only focusing on areas related to teaching and learning, but in relation to such issues as “motivation and performance” and “innovation and change in the educational setting” (p. 127). Additionally, for Alderson and Wall, the possibility of a washback effect on teaching and learning very much aligns with linking washback only to the test and its consequences. This is clear in the stated hypothesis that “tests that have important consequences will have washback”. Cheng (2005) identified this notion as “washback intensity”, which refers to “the degree of the washback effect in an area, or in a number of areas of teaching and learning affected by examination” (p. 33). This means that if “important

consequences” are associated with the test score, the test will likely have washback, be it positive or negative, at both the classroom and educational context levels.

3.4.2 Hughes’ (1993) model of washback

In an unpublished paper, Hughes (1993, as cited by Bailey, 1996), described washback as a process by means of which actions are taken to improve learning outcomes, such as the construct being measured or language proficiency. Hughes (1993) synthesised Alderson and Wall’s hypotheses and questioned their view that it was necessary to be more precise about what constituted washback. In his empirical study (1988), evidence of the beneficial washback effects of the new Bogazici proficiency test conducted in Turkey was based on the rise in English language level standards since the introduction of the test and the positive attitudes held by the university academic staff surveyed. He suggested that defining washback as “the test’s effect on any aspect of teaching and learning” (1988, as cited in Alderson and Wall, 1993, p. 125) was adequate and introduced his own way of categorising the influence that testing might have on teaching and learning. More specifically, Hughes proposed a trichotomy (see Figure 3.1) that serves as a framework to illustrate how washback works in teaching and learning processes. Hughes highlighted three main components that are affected by a test: the participants, the process and the product. These three components in Hughes’ model have been used as guides for other washback models, particularly those of Bailey (1996), Markee (1997), Cheng (2005) and Tsagari (2006).

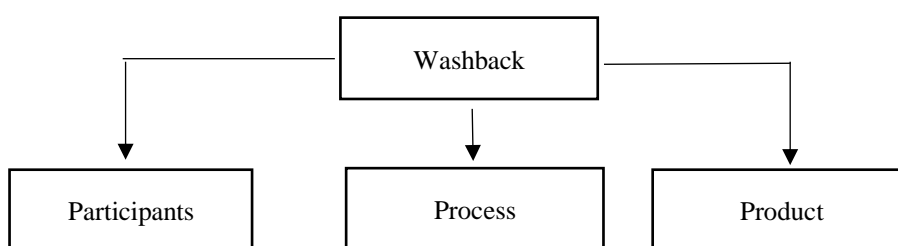


Figure 3.1. Hughes’ (1993) washback trichotomy

In Hughes’ view, participants include “all of those whose perceptions and attitudes towards their work may be affected by a test” (1993, as cited in Bailey, 1999, p. 9). This encompasses teachers, learners, administrators, materials developers, researchers and any other pertinent stakeholders. The term “process” refers to “any actions taken by the participants which may contribute to the process of learning” (1993, as cited in Bailey, 1999, p. 9). Hughes stated that these processes encompass syllabus design, material creation, alterations in teaching methods or content, and learning or examination strategies. Finally, the “product” in Hughes’ framework is defined as “what is learned

(facts, skills, etc.) or the quality of learning such as fluency, etc.” (1993, cited in Bailey, 1996, p. 262). Thus, learning is affected by a test via the test’s stakeholders. All participants in the washback process and the products also exert an influence on the test itself. Hughes (1993) explained how washback works according to the three components in his trichotomy, as follows:

The trichotomy into participants, process and product allows us to construct a basic model of backwash. The nature of a test may first affect the perceptions and attitudes of the participants towards their teaching and learning tasks. These perceptions and attitudes in turn may affect what the participants do in carrying out their work (process), including practising the kind of items that are to be found in the test, which will affect the learning outcomes, the product of that work. (Hughes, 1993, p. 2, as cited in Bailey, 1996, p. 270)

In addition, Hughes (2003) distinguished between negative and positive washback (see 3.3.3). The essence of washback research is to identify ways of ensuring positive washback on learning and conversely reducing, if not entirely avoiding, negative washback. For Hughes (2003), the extent to which the content of achievement tests and techniques are emphasised in the curriculum objectives (rather than in teaching content) indicates the amount of positive or negative washback. He refers to an instance of negative washback in an English course aimed at learning different language skills, but in which writing was assessed in the form of closed questions rather than open-ended responses. This meant that the teachers and students were under pressure to focus on such items rather than acquire the skills. Therefore, if the teaching and curriculum are of high quality but the tests do not accurately reflect this, there is the potential for negative washback on teaching and learning. Hughes (2003) suggested seven ways of encouraging positive washback, as follows:

- 1) Test the abilities whose development you want to encourage,
- 2) Sample widely and unpredictably,
- 3) Use direct testing,
- 4) Make testing criterion-referenced,
- 5) Base achievement on objectives,
- 6) Ensure the test is known and understood by students and teachers,
- 7) Where necessary, provide assistance to teachers.

(Hughes, 2003, pp. 53–56)

Even though Hughes' (1993) washback trichotomy outlines the components of washback (participants, processes and products) that are most likely to be impacted by a test, it does not demonstrate the connections between the three parts in the model and how they could work together to create negative/positive washback. Moreover, the model suggests a

unidirectional process in terms of how washback operates. Similarly, the strategies suggested by Hughes (2003) in relation to washback processes seem to indicate that washback is the sole responsibility of those in charge of the examination, giving teachers and students no obvious roles to play in the process. Even if washback strategies indicate that the teacher might act as an assessor in the classroom, the students, as the assesses, are not included in the washback process.

Furthermore, Hughes' (1993, as cited in Bailey, 1999, p. 9) assumptions about the mechanism of the washback process only provide the description "any actions taken by the participants which may contribute to the process of learning"; why and how participants' actions might occur (or not) is not addressed. For example, the model does not consider differences in individual participants' views of a test and responses to it, despite Hughes' (1993) emphasis on the consequences of tests for learning. Given that the literature has emphasised how complex and elusive washback is, Hughes' (1993) model required further development, as it did not incorporate any factors that could affect the teaching and learning context beyond the test itself.

3.4.3 Bailey's (1996) model

Bailey (1996) conducted a review of the literature on the construct of language testing washback, based on which she proposed a model and strategies for investigating washback (see Figure 3.2), as well as discussing various theoretical perspectives. Her starting point was that language teaching professionals ought to take into account the likelihood of negative washback if tests are not aligned with the principles and practices of CLT. Bailey (1996) highlighted the differences between CLT and traditional testing in terms of their content, design and procedures. The criteria used to construct and develop assessment instruments and their focus on teaching and learning strategies also vary. For Bailey (1996), tests aligned with CLT would evaluate overall communicative competence, while traditional tests measure linguistic competence. For example, testing in CLT is often distinguished by its authenticity and direct assessment of activities and materials. Indeed, Morrow (1991) remarked that in terms of the content and design of tasks in tests aligned with CLT, "reliability is second on construct grounds to authenticity" (p. 115). In this regard, Bailey (1996) suggested that the development of teaching practices that emphasise communicative approaches and principles has created mismatches between the form and design of existing examinations and what it is that test stakeholders (teachers and students) do in the classroom.

To understand the process of washback from testing, Bailey (1996) built on Alderson and Wall's (1993) washback hypotheses, together with aspects of Hughes' (1993) trichotomy, to formulate her own washback model (see Figure 3.2). For instance, Hughes' trichotomy can be seen in Bailey's basic model of washback: a test will direct influence the participants, who are involved in different processes, resulting in a product, related to each category of participants. The dotted lines in Figure 3.2 indicate possible influences of the participants on the test, termed "wash-forward" by van Lier (1989). Thus, according to Shih (2007), the dotted lines in Bailey's model indicate how the participants, influenced by the test, might have a reciprocal impact on the test itself (p. 136).

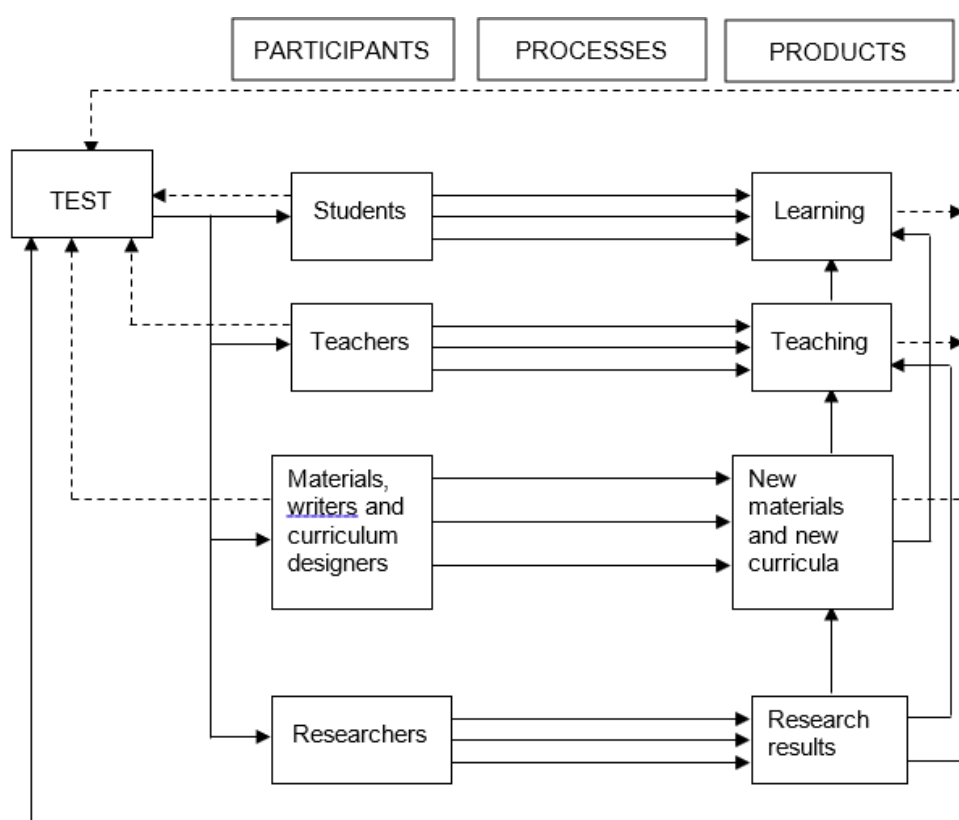


Figure 3.2. Bailey's (1996, p. 264) basic model of washback

In this model, Bailey classified washback into two main types: *washback to the learners* and *washback to the programme*. By *washback to the learners*, Bailey (1996) means the results of the test-derived information that is designed for learners as test-takers and that has important consequences for the learners. She listed 10 examples of the actions and interactions that learners might participate in, a few of which are as follows:

1. Practising items similar in format to those on the test,
2. Applying test-taking strategies,

3. Emphasising test-preparation classes or tutorials,
 4. Studying grammar or vocabulary,
 5. Participating in interactive language practice (e.g., conversations).
- (Bailey, 1996, pp. 264–265)

The outcomes of these processes or actions depend on the extent to which they are used or selected. As per Bailey's (1996) suggestion, selecting specific processes can have either positive or negative effects on students' learning outcomes, depending on whether they improve their language proficiency. Bailey (1996) noted that five of Alderson and Wall's (1993) hypotheses (2, 5, 6, 8 and 10) correspond to her *washback to the learners*. However, while Bailey (1996) mentions that the existence of beneficial or negative washback depends on the participants' use and selection of these processes, there is no particular explanation of the responsibility of the learner (or teacher) in taking decisions about their selection, or even how the learner would use these processes or select from among them. In other words, Bailey (1996) refers to some possible learner actions when faced with an important test, but there is no information about whether these actions are the total responsibility of the learner or if there are other test stakeholders (e.g. teachers, parents, school principals, etc.) who might determine the learners' use and selection of these processes.

The second classification in the model, *washback to the programme*, is more general and encompasses all the different types of participants, including students. Bailey (1996) explains that this mode of washback refers to test-derived information potentially also influencing all the test participants in the educational system, including teachers, administrators, curriculum developers, students, counsellors, and so on. As with *washback to the learner*, six of Alderson and Wall's (1993) hypotheses (1, 3, 4, 7, 9 and 11) correspond to Bailey's *washback to the programme*, such that washback might influence the effectiveness of teaching and attitudes towards the methods and content of teaching. Additionally, Bailey (1996) proposed several highly important factors other than the tests themselves that might promote positive changes in teaching and learning, each of which is addressed briefly in turn below.

Language learning goals

Bailey (1996) posited a direct relationship between tests and learning goals, based on whether a test serves to achieve or impede the achievements of these goals. Hughes (2003) agreed with this perspective, arguing that if achievement tests are based on learning objectives, they will promote positive washback on students' learning. However, even if

Bailey's (1996) argument is valid, a difficulty emerges, as suggested by this thesis, when there are sub-factors that may affect the achievement of language learning goals. Among the most important of these are the different characteristics of the individual stakeholders and the nature of the context.

Building in authenticity

Bailey (1996) cited the views of Carroll (1980), Wesche (1983), Morrow (1991) and Messick (1996) in addressing the direct connection between authenticity and positive washback. Increasing students' motivation and ability to use the target language in real-life situations can be achieved by ensuring that language tests and their objectives are aligned with real-world tasks. Several authors have advocated this approach and proposed that the more authentic a test is, the greater the positive washback effect on learners' motivation. Therefore, authenticity should be a key consideration when designing language tests to encourage positive washback and foster learner motivation.

Learner autonomy and self-assessment

Bailey (1996) maintained that self-assessment in the curriculum promotes washback to the learners as it is a procedure designed to promote learner autonomy and involvement. For Bailey (1996), learner autonomy refers to "the philosophy that learners should have a large amount to say about what, how and how fast they learn" (p. 270). She maintained that self-assessment and learner autonomy are related directly to positive washback, because developing self-assessment mechanisms is one of the key features of autonomous learning.

Score reporting

Another factor that promotes positive washback, as suggested by Bailey (1996), is providing more detailed score reports for large-scale proficiency tests. Bailey (1996) cited Shohamy's (1992) view that examination reports should be "detailed, innovative, relevant and diagnostic" and that they must "address a variety of dimensions rather than being collapsed into one general score" (p. 515).

Bailey (1996) observed that although the above issues (authenticity, autonomy and self-assessment, and detailed score reporting) have been widely acknowledged in the literature as being useful in promoting positive washback, few empirical studies have studied these claims. She therefore posed a series of questions for researchers in relation to the positive washback potential of any "external-to-programme test", as follows (p. 276):

- 1) Do the participants understand the purpose(s) of the test and the intended use(s) of the results?
- 2) Are the results provided in a clear, informative and timely fashion?
- 3) Are the results perceived as believable and fair by the participants?
- 4) Does the test measure what the programme intends to teach?
- 5) Is the test based on clearly articulated goals and objectives?
- 6) Is the test based on sound theoretical principles which have current credibility in the field?
- 7) Does the test utilise authentic texts and authentic tasks?
- 8) Are the participants invested in the assessment processes?

Although Bailey's (1996) model mentions participants and products, it does not provide a clear explanation of what the intermediate processes/or actions are, or how the participant and process dimensions would lead to the corresponding products. Thus, while Bailey's (1996) model was designed to understand how washback from testing works, the "processes" that are supposed to be a vital component of the washback phenomenon (and which are a key focus of this study) are not fully identified or described. Bailey's (1996) strategies and questions emphasise the role of what she calls "standardized, external-to-programme tests" in promoting positive washback, but there is no indication of any other sources of washback, such as curricular innovation or classroom assessment. Similarly, although Bailey (1996) claimed that the potential benefits of washback for both programmes and students stem from the "ideal" characteristics of standardised, external-to-programme tests (p. 267), the test alone is not the exclusive initiator of washback; rather, there are other complex and interrelated factors beyond the test itself, such as teachers' personal characteristics and societal influences (Alderson & Hamp-Lyons, 1996).

3.4.4 Green's (2007) model of washback

Building on the models proposed by Hughes (1993) and Bailey (1996), Green (2007) expanded the model of washback to outline the relationships between i) test design, which is used to determine washback direction, mediated by both ii) participant values, motivations and resources, as key determinants of washback variability, and iii) the perceived importance and difficulty of the test as the major determinants of washback intensity. Due to the lack of focus on test design in previous washback models and a lack of attention to intended washback on the part of test developers, Green (2007) highlighted the importance of evaluating the test instrument and undertaking detailed analysis of its congruence with the planned curriculum.

Based on an extensive empirical study examining washback in relation to writing scores in the IELTS, Green (2007) proposed an alternative washback model, consisting of three dimensions (as shown in Figure 3.3): direction, variability and intensity. Using a variety of data collection methods (questionnaire, observation and interview), Green (2007) explored the relationship between these dimensions and IELTS Academic Writing Module (AWM) scores using two phases of analysis. The first was analysis of covariance, which was used to identify significant differences between the scores obtained by learners in the three different IELTS courses when the students' age was held constant. The second phase of analysis involved the construction of prediction models (a neural method of analysis and traditional linear prediction through multiple regression) to identify which constellation of variables might provide the most accurate prediction of IELTS scores in writing.

Having acknowledged the complex nature of washback on teaching and learning, Green (2007) presented the relationship between test design and the focal construct as one way of determining the direction of washback. When a test accurately reflects the key concepts understood by educators and students, there is a stronger correlation between what is taught and what is assessed. This is known as positive washback and is more likely to occur when the test format, content and difficulty level align with the course's focus. Conversely, as the overlap decreases, there is a higher chance of negative washback. Green's (2007) model acknowledges that test use and test stakes, in addition to a test's characteristics, contribute to individual conceptions of and attention to test demands, as well as perceptions of how the test, the curriculum and the construct relate to one another. When test scores become more important to individuals than their understanding and knowledge of the subject, the test will likely have a negative impact on their learning.

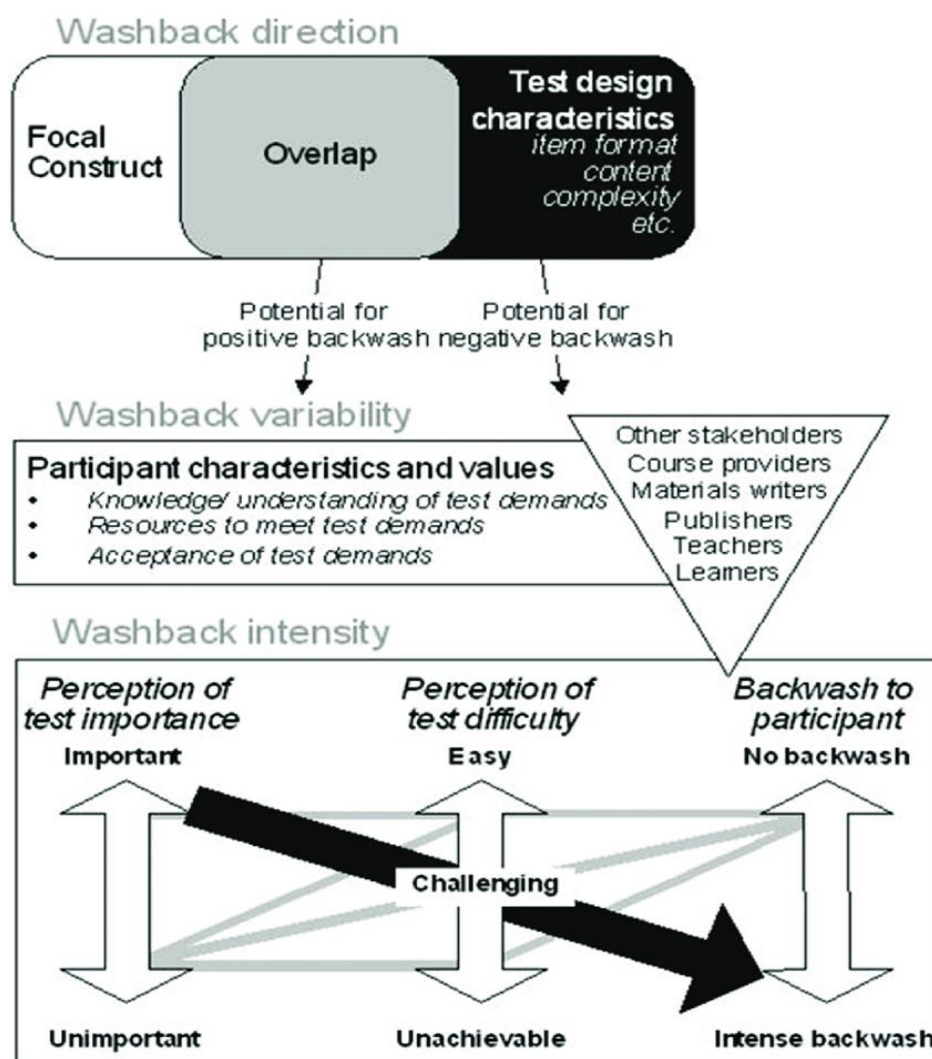


Figure 3.3. Green's model of washback direction, variability and intensity (2007, p. 24)

Green (2007) found that the importance of the test affected students' preparation efforts, with higher stakes leading to more motivation. Design decisions are also influenced by the test's significance, ensuring qualities such as reliability, objectivity, validity and fairness when stakes are high. When it comes to determining washback variability, Green (2007) made the point that because of intervening variables, tests should not, as in Alderson and Wall's (1993) argument, be assumed to be deliberate washback-generating devices; rather, this needs to be established through investigation. In the face of this complexity, Green (2007) highlighted the importance of individual variation in responses to a test, focusing on teachers' behaviour in relation to innovations in testing: individual teachers bring different beliefs and experiences to teaching, which can lead them to react in different ways. Individual participants need to have sufficient understanding of a test's demands and its implications for learning, as well as the resources to undertake test preparation (such as learning materials, teacher expertise, etc.).

In terms of the third dimension in the model, Green (2007) asserted that the influence of a test appears to be different for certain tests and in certain settings. For Green (2007), washback intensity is determined by how important (unimportant) a test is perceived to be and simultaneously whether it is regarded as easy, difficult or unachievable. The relationship between the intensity of washback and the perception of test difficulty, as pointed out by Green (2007), is not simple or direct. If a test is deemed more significant and challenging, it is more likely to have a powerful washback effect (and the opposite also applies). The thick black arrow in Green's (2007) model (Figure 3.3) denotes that washback will be most intense when participants consider success in the test to be attainable but challenging. Green (2007) also related a high level of washback intensity to participants valuing success in the test rather than improving the skills represented in the test (test constructs) and in contexts in which there is a shared perception among stakeholders of the importance of the test. Green (2007) also argued that a test with weak or no consequences may have no washback at all (see Figure 3.3). This means that differences in participants' perceptions of test importance and test difficulty, and their ability to change their behaviours in relation to test demands, will moderate the degree of washback effect, as well as the evaluation of its direction. The model thus highlights the need to investigate the intensity of washback from the perspective of the perceived importance and difficulty of a test as a way of better understanding the complex phenomenon.

In sum, the model foregrounds the importance of test design and characteristics, along with Messick's (1996) idea of construct validity, in determining the direction of washback. The model further considers the importance of participant characteristics and values in directing washback effects. Green's (2003) model takes into account test importance and difficulty to investigate the relationship of these factors and washback intensity. Although the model points to the complex nature and scope of washback, Booth (2012) felt that it did not fully explain the role of other mediating factors that need to be met for washback to be fully realised.

The washback models discussed above show that washback is very complicated – far from a direct and automatic effect – and involves many more factors than just “good” test design. The influence of testing on various aspects of teaching and learning has been described in these models, together with the ways in which washback processes are generated by the interplay between and intertwining of various underlying factors outside

the test itself. Alderson and Wall's (1993) list of washback hypotheses suggests that predictions regarding methods (how), content (what), sequence, rate and the sequence and degree of depth of teaching and learning could potentially be appropriate dependent variables for the washback research agenda. Hughes' (1993) work provides a basis for broader investigation focused on both perceptions and processes, then linked to the products of learning. In Bailey's (1996) model, participants' attitudes will affect their actions in terms of both what they do and how they do it. In turn, these processes will influence the range and sequence of the teaching content. In line with the need for greater explicitness in this area, Watanabe (2004) established various washback attributes to help uncover empirical evidence for the washback phenomenon. These are discussed in the following section.

3.5 Characteristics of Washback

As noted above, while washback is widely associated with tests, the existing literature suggests that positive washback is not directly a function of test design. As Alderson and Wall (1993) pointed out, washback from testing may not be predicted, but must be established through investigation. Watanabe (2004) identified several attributes characterising washback, arranging these under the following headings: (1) dimensions; (2) aspects of teaching and learning that may be influenced by the examination; (3) factors that mediate the process of washback being generated. These various characteristics are discussed below.

3.5.1 Washback dimensions

"Dimensions" here refers to the various factors that characterise the washback concept and the way in which it works. Watanabe (2004) divided washback into the following components, each of which represents the complex nature of washback: specificity, intensity, length, intentionality and value.

Specificity

Washback has been found to be both more complex and more elusive than had been anticipated and it can be characterised as being both general and specific at the same time. Put another way, washback theory has been viewed by researchers (e.g. Alderson & Wall, 1993; Green, 2007; Hughes, 1993) as context-bound, with the nature of washback varying from one specific test type to another, from one task to another, or in relation to the test culture or the test stakeholders in any particular context. From this perspective, the

washback of a particular test is unique and the product of a particular test. This is reflected, for example, in the belief that if grammar as a separate component is measured in a specific test, washback participants will focus especially on this in their teaching and learning. However, the attribute of washback is based on the idea that the effect of testing is not linked to a particular context or cultural beliefs. For example, if the assumption is that a test increases students' motivation to learn more than they ever had before, washback would be generated by any type of test and hence be a general washback effect.

Intensity

Cheng (1998) first coined the term washback intensity, which refers to the level of impact a test has on test-takers' adherence to its requirements. This relates directly to Alderson and Wall's (1993) assertion that tests which associated with important consequences will have a strong impact on teaching and learning, whereas those with low or no consequences will have little to no effect. According to Gipps (2002), test use and its associated consequences determine the major factors in washback intensity, rather than the construct or the skill which the test intends to measure. Madaus (1988) argued that test consequences are linked to the perception of a test's stakes, rather than the fact that it will be used to drive instructional change. If teachers and students perceive a test to be a means of determining progression to the next grade or a potential form of punishment for underperformance, they are more likely to alter their behaviour in reaction to the test, engaging in actions caused by the test, in line with Alderson and Wall (1993). However, other arguments favouring the use of performance assessments to influence educational reform assume that the accountability pressure from high-stakes tests will continue to be needed to improve learning (Crooks, 1988; Gipps, 2002; Shepard et al., 1996; Shohamy, 1992). Yet, as the literature suggests (see 3.3.2), high-stakes tests in themselves may also sometimes generate negative washback resulted from other factors than the tests' qualities and its content.

Length of washback

Washback length concerns how long the test's influence lasts. According to Watanabe (2004), this is determined by two main factors: whether there are any important consequences associated with the test and the nature of the test's influence. For example, if the washback of a placement English test is only present while the students are learning and using certain learning strategies to prepare for the examination and this effect disappears once the test is completed, it is short-term. In contrast, if the impact of a test

with important consequences continues even after the test has ended, the test has long-term washback.

This idea builds on Alderson and Wall's (1993) washback hypotheses, namely that significant consequences attached with a specific test will result in strong washback, and also draws on the attribute of intensity (Cheng, 1998). However, while the potential for washback to have short-term or long-term effects has been studied on a conceptual level, there is a lack of research empirically measuring washback length. This may be because such studies would require a longitudinal design to detect changes in the characteristics of the participants involved over a period of time. Scaramucci and Kobayashi's (2013) washback study suggested that if students decided to take the tests on offer during their elementary and secondary schooling, it would help them get accustomed to the examinations; however, the researchers did not provide any empirical evidence to support this assertion.

Intentionality

References to intended versus unintended washback are frequently found in the literature (e.g. Bailey, 1996; Cheng, 1997; Qi, 2005). For example, Qi (2005) outlined the view that intended washback is regarded as a change – for example, in ELT practices – that facilitates the educational process. Similarly, in Cheng's (1997) study, intended washback was linked to the curriculum changes that Hong Kong secondary schools were seeking to bring about. Thus, intended washback has been seen as reflecting changes (for example, in the curriculum, teaching quality, students' learning, or individuals' perceptions) which the educational system is aiming to achieve. However, there has been insufficient investigation examining washback processes from the perspectives of test designers or policymakers (Cheng, 1998; Qi, 2007; Shohamy, 2001), as discussed later in (3.6.2). Nor has there been much empirical research on unintended washback, although it is expected there will soon be more, as seen in this research and other recent washback studies.

Value

The value of washback refers to the reason(s) for researching washback. Although test writers may intend to promote beneficial washback on teaching practices, explicit references to the negative consequences of tests and any change in teaching and learning tending not to be significant or less than expected were first made more than 50 years ago (see 3.3.2). The literature on washback has consistently sought ways of maximising the

positive washback effects of testing on teaching and learning, and conversely reducing or avoiding the negative washback effects. In this regard, Watanabe (2004) drew researchers' attention to a further level of washback complexity, in that the value of judgments about washback can vary according to the different evaluators/stakeholders involved in the process. One type of learning objective may be evaluated as positive by examination writers, for instance, whereas the same objective might not necessarily be judged in the same way by teachers, and so on. The different views that stakeholders can hold leads to the need – in relation to any testing system – to identify whether or not there is any interaction involving different stakeholders in which they might have a different agenda or different objectives.

3.5.2 Aspects of learning and teaching potentially influenced by the examination

Based on Alderson and Wall's (1993) 15 washback hypotheses and Bailey's (1996) washback model, Watanabe (2004) proposed two types of variables that are caused by the examination: "washback to the learner" and "washback to the programme". The former covers the how, what, rate and sequence, and degree and depth of learning, while the latter involves the what, how, rate and sequence, and degree and depth of teaching. Watanabe (2004) argued that comparatively more emphasis has been placed on washback to the programme in research than on learner washback due to the difficulty of gaining access to learners.

3.5.3 Factors mediating the washback process

In view of the complexity of the washback in terms of its the occurrence, strength and type, Watanabe (2004) listed a set of factors driving washback other than the test, identified by previous washback literature as influential in effecting washback processes (Alderson & Hamp-Lyons, 1996; Alderson & Wall, 1993; Brown, 1997; Spart, 2005). Alderson (2004) argued that washback studies should take into consideration not just the context of the investigated test, but also all the myriad factors that can enhance or challenge any changes assumed to arise from a test. In a review of second language acquisition research, Spolsky (1989) noted 79 variables that may influence the degree and rate of teaching and learning and suggested that these interact in a complicated way. Some of the factors identified were related to the learners' background, such as their age, social class and L1, and psychological characteristics, such as intelligence, motivation, language learning strategies and language aptitude. Shih's (2007) study further listed factors that represent the full domain of washback based on a review of empirical studies examining

aspects that could affect the degree and rate of washback on the teaching context. These include contextual factors, such as school management (Wall & Alderson, 1993), course objectives (Hayes & Read, 2004), class size (Alderson & Hamp-Lyon, 1996), the duration of the course, teaching time (Shohamy et al., 1996; Watanabe, 1996), the professional development of teachers (Hawkey, 2006), and resistance from class teachers owing to class schedules and individual differences among students (Green, 2007; Hawkey, 2006; Skehan, 1989).

Other washback studies have included a range of test factors (e.g. Alderson & Hamp-Lyons, 1996; Green, 2007; Hughes, 2003; Hyland, 2011; Shohamy et al., 1996), such as test methods, content, skills, purpose and decisions made as a consequence of test scores, the stakes of a test and the status of a test in the educational context. Teacher factors comprise teachers' abilities to teach and their English language abilities (Qi, 2007), teaching experience (Watanabe, 1996), teacher training and professional opportunities (Green, 2007; Shohamy, 1993; Wall & Alderson, 1993; Watanabe, 2004), learning experiences (Watanabe, 2004), teachers' concerns for students' levels of achievement (Watanabe, 1996), teachers' familiarity with effective teaching methods (Watanabe, 2004), teachers' perceptions of the importance of the tests (Shohamy et al., 1996), perceptions of the test's qualities (Shohamy et al., 1996), other teaching responsibilities in schools (Wall & Alderson, 1993), the degree of teachers' commitment to the teaching profession (Wall & Alderson, 1993), and teachers' capabilities to innovate (Wall & Alderson, 1993).

Watanabe (2004) also suggested that research into washback should consider variables that affect the learning context to explore the washback phenomenon in a more tangible way. This is in line with Wall's (2012) statement concerning "the difficulty of separating out the influence of tests from the effects of other variables at work in the educational contexts" (pp. 83–84). Empirical washback studies to date describing the interaction between these factors, or even between teaching and the context in which the test is used, are scarce. The interactions between various factors and their varying degrees of depth suggest that testing washback may not always occur. In addition, when it does occur, it may take on different intensities and forms depending on the context (Spart, 2005). Addressing any variables other than the test itself, therefore, would make the complexity of washback more tangible (Watanabe, 2004). Further considerations of such variables are discussed later in this thesis.

Examining the literature assists in understanding the nature of washback and how it can function within a given educational system. Although the studies reviewed outline a general framework for the washback phenomenon, involving participants, processes and products, there is still a need to delve deeper into the intricate nature of the phenomenon and how or why these components might shape washback. This complexity can be partially accounted for by examining not just the components of washback (participants, process and product) that seem to influence and be influenced by a new examination, but also the interactive relationships among them and other factors that may mediate the teaching and learning process beyond the examination itself. In particular, this study explores washback complexity at the process level, where the interactions between different participants contribute to the nature of the washback on teaching (positive or negative). All the participants in this study are in some way responsible for the nature of the washback product resulting from the introduction of the new examination. The primary focus of the data to be considered later in this thesis will be the washback on teachers in particular high-stakes examination situations. The following sections thus set out the role of washback participants and processes work to influence the nature of the washback product.

3.6 Washback to Participants

3.6.1 Washback on teachers' perceptions

In Hughes' (1993) view, as mentioned earlier, the perceptions and attitudes of participants are important in determining the washback effect in teaching and learning because they exert a crucial influence on behaviour (see also Cheng, 2005; Green, 2007; Shohamy, 2000; Wall, 2005). This behaviour can involve teachers and students making special preparations for tests by covering test-specific content, principals putting pressure on teachers to raise student achievement to improve the standing of the school, or an educational system allocating special resources on the basis of the level of national or international test results (Shohamy, 2007). Harlen and Crick (2003) argued that teachers are perhaps the most important stakeholders in testing, since "they are the only ones whose actions directly affect students" (p. 203). Other washback studies (Alderson & Hamp-Lyons, 1996; Bailey, 2005; Kim & Isaacs, 2018; Shohamy, 2007; Wall, 2005; Watanabe, 2004) have highlighted the critical role of teachers in facilitating or impeding the effectiveness of language testing. Shohamy (1992) asserted that one requirement for the successful use of test results is the involvement of "agents of change" – principals and teachers – in the different stages of the testing process, because "they are the ones

expected to carry out change” (p. 515). Similarly, Burrows (2004) concluded that it is crucial for test designers seeking beneficial effects on teaching and learning to consider the “teacher variable” when planning strategies for implementing new tests (p. 127).

Pajares (1992) argued that teachers’ beliefs are a key element influencing the way they make decisions about their lesson preparation. These beliefs are determined by various factors, such as teacher training, teachers’ prior experience and societal requirements and expectations. Brown (2004) conducted a study of assessment practices with 525 New Zealand primary school teachers and determined that to ensure the successful implementation of an assessment system, it is necessary to comprehend and consider the complex nature of teachers’ beliefs and their understanding of assessment. In general terms, beliefs are formed through interactions in daily life and depend on personal experiences and how individuals interpret events. According to Bauch (1984, p.3), these beliefs underlie attitude formation, which in turn influence intentions, thus becoming the basis for decisions that govern teaching behaviours. The implications for this in an educational context have been addressed by Borg (1999).

Extensive research has been conducted on teachers’ cognitive psychology, encompassing their beliefs, perceptions, knowledge and thought processes (Yin, 2005). This work has illustrated “how teachers interpret what goes on in their classrooms and how they will react and respond to it” (Johnson, 1994, p. 440). Woods (1996) suggested that teachers’ beliefs, assumptions and knowledge (BAK) are essential factors in understanding their behaviour and preparation for classroom teaching (p. 195).

Burrows (2004) proposed that washback will differ from one teacher to another and that teacher beliefs need to be regarded as individual and variable influences. She offered three models representing different views of the washback process on teachers (see Figure 3.4). The “traditional” view shows the effect of a new test as necessarily influencing the responses of all teachers in the same way. The 1990s view acknowledges the existence of individual responses to a new test, resulting from teachers’ differing beliefs. The third model, “curriculum innovation”, incorporates “models of responses”, acknowledging that different teachers can share a similar response and thus teacher responses need not be viewed simply in individual terms, but can be grouped in different types or models (Burrows, 2004, p. 126).

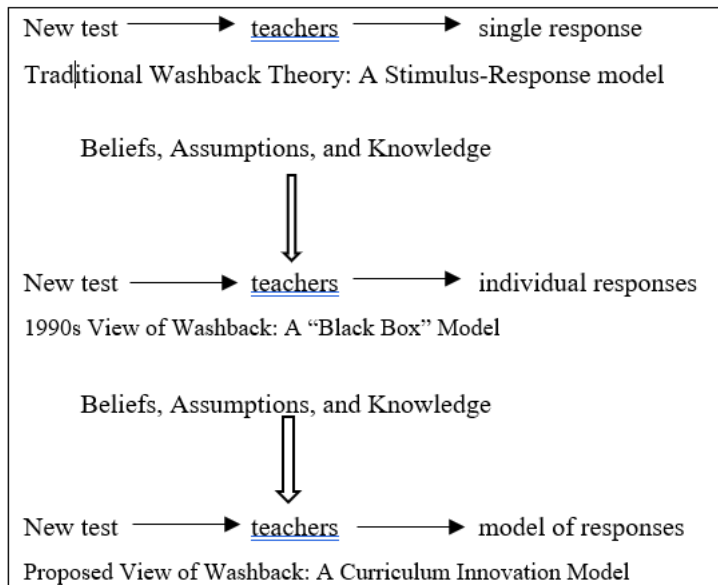


Figure 3.4. Burrows' (2004) set of models of washback responses

Burrows' (2004) curriculum innovation model embodies the view that when it comes to teacher behaviours, there is a link between washback and the introduction of new curriculum. In other words, the change in testing and curriculum affects teachers' BAK, revealing patterns in their responses. When a new test or curriculum is introduced, teachers' teaching behaviours or responses vary according to their BAK regarding the new test and the new curriculum.

Attempts at educational reform through testing may be hampered by a general lack of resources, a lack of teacher involvement in the development process, poor teacher understanding of how to improve test scores, insufficient knowledge regarding teaching content or teaching strategies, and the failure of the proposed changes actually to improve students' performance levels (Agrawal, 2004; Azadi & Gholami, 2013; Barnes, 2017; Chio, 2008; Kilickaya, 2016; Orafi & Borg, 2009; Rahman et al., 2021; Rao & Haque, 2019; Read & Hayes, 2003; Salehi & Yunus, 2012; Shohamy, 2001; Yildirim, 2010). Wall and Alderson (1993) explored teachers' attitudes towards changes in the O Level English examination in Sri Lanka. They concluded that the new test had no impact on the way teachers were teaching. Their study attributed this result to a lack of knowledge on the part of teachers' about how to prepare their students for the examination. Similarly, Shohamy's (1993) study of the introduction of three new national language tests in Israel showed that they were imposed upon teachers, who were given no input into the tests in advance. In addition, the teachers were not given any instructional guidance and received no specific training on the new topics introduced in the new examination. Consequently,

teachers felt a great deal of fear and pressure to ensure their students attained high scores in this examination.

Rahman et al. (2021) studied the washback effect of the Secondary School Certificate (SSC) English examination in Bangladesh found that teachers' lack of knowledge of curriculum goals and their lack of training in language assessment augmented the negative washback on their perceptions. Their failure to understand assessment objectives and principles indirectly encouraged their students to ignore those skills not tested in the examination and their negative attitudes again affected other areas, such as test quality, administration practices, lesson preparation, and the use of test scores and information. Despite providing several significant insights, the sample size in this study was limited and thus the results may well not be representative of all schools in Bangladesh. Also, while the study used document analysis to examine the test design and format and their compatibility with the objectives of the English language curriculum, it did not take into account the perspectives of the test developers or examine the teachers' and students' views and compare/contrast them with the test developers' intentions.

A study by Dammak et al. (2022) in Tunisia investigated the impact of the National English Baccalaureate Exam (EBE) on teaching practices. Through a mixed-methods approach, the study found that teachers' attitudes and beliefs affected various aspects of classroom instruction, such as content, teaching strategies, materials and class time devoted to different elements of teaching and learning. Teachers' attitudes were mixed, but on the whole negative. They claimed that although the curriculum encouraged the CLT approach and integration of language-learning skills, they still used traditional methods in their teaching. Moreover, there was narrowing of the curriculum and they concentrated on practising test-related skills. Similar results were also found in the Saudi Arabian context in a recent study by Rao and Haque (2019), which showed particular variation in the attitudes of experienced and less experienced teachers.

In a study of TOEFL, Alderson and Hamp-Lyons (1996) concluded that most teachers held negative attitudes towards the test because "they considered it inauthentic and noncommunicative" (p. 285), as well as being confused about the appropriate material to teach. They also felt that teaching for TOEFL was boring and that they did not have enough time for proper in-class preparation. Despite the challenges, the two instructors had a positive outlook when teaching the TOEFL course. They found fulfilment in helping

students tackle important subject matter. Alderson and Hamp-Lyons (1996) also noted that the teachers experienced frustration and guilt when they could not make the content engaging or guarantee better student scores, despite following the book's guidance. The researchers noted the need for more work on this issue and suggested that such attitudes are typical among teachers who teach TOEFL preparation classes. In addition to analysing the teachers' statements about class requirements and expectations, the study also interviewed students to gain insights into their perspectives and test-taking strategies. However, the interviews revealed nothing about the students' progress in learning during the TOEFL classes.

When high-stakes examinations are used for accountability purposes, and certainly when sanctions or awards are linked to test scores and teachers are held responsible for students' poor results, teachers are liable to find ways of increasing students' test scores without necessarily developing their English language abilities (Abu-Alhija, 2007; Stecher, 2002; SuKyadi & Mardian, 2011). Smith et al. (1991) conducted interviews with teachers and other stakeholders (testing professionals and administrators, students, members of the public) and recorded statements they made about testing during school observations. The main findings of the study showed that the teachers were the only ones to have access to an "interpretative context"; that is, they alone were able to judge the meaning of the test scores in relation to students' daily performance. Although the teachers valued the test results for the information they conveyed about student achievement, testing professionals and administrators seemed to use the test scores as an organisational tool to punish or reward, or to control the information obtained from the participants' real achievements. The test results were a source of great anxiety for teachers and led to a shift in teaching away from a focus on particular skills and attainments. In their long-term school observations, the researchers saw evidence of a "narrowing of curriculum happening before our eyes" (p. 10). Similar conclusions also emerged from Gunn et al.'s (2016) study, in which the teachers stressed that tests should not be the only tool used to assess students' achievement and felt that the drawbacks of testing tended to outweigh its benefits.

Waltman (2008) examined the perceptions of teachers concerning test preparation practices to determine the factors affecting their perceptions of a given teaching activity. The study found that most teachers viewed practising test-like questions as unethical and teaching test-taking skills as ethical. There were actually no significant differences

between perceptions of teachers towards the ethicality and appropriateness of particular practices and the different levels of student achievement, but use of practices and perceptions did vary across grade-level configurations. The negative feelings and pressures felt by teachers in these studies can be seen to resonate with Alderson and Wall's (1993) concern that "for teachers, the fear of poor results, and the associated guilt, shame, or embarrassment, might lead to the desire for their pupils to achieve high scores in whatever way seems possible" (p. 118).

Furthermore, other studies have reported individual variations in responses to test (Haladyna et al., 1991; Imsa-ard, 2020; Mahmoudi & Baker, 2013). Cheng's (2005) study explored the perceptions of teachers and students concerning the Hong Kong Certificate of Education Examination in English language (HKCEE) introduced in 1993. The investigation focused on how the change in the public examination would affect classroom teaching in the Hong Kong system and which of the different variables in the classroom teaching context would be affected. The findings over the two-year study period revealed that although the importance of the new examination did result in a change in teachers' attitudes, this was not reflected in any changes to their role in teaching and learning or the methodology and approach they adopted. The researcher also found clear divergences between the teachers' attitudes and those of their students in relation to such areas as teacher talk and teaching and learning activities. The study further concluded that "conceptual changes are seldom achieved without attending to the beliefs of those who are the targets of change – teachers – and the conditions of the environments in which they function – schools and student levels of proficiency" (p. 269).

In Burrows's (2004) study, teachers tended to have divergent views regarding the new Certificate in Spoken and Written English (CSWE) and its assessment scheme. Teachers were asked whether the new assessment had impacted their teaching and whether they had observed any modifications in the way they teach due to its implementation. Some immediately adopted the new curriculum in its entirety; others took a more gradual approach, maintaining useful elements from the old curriculum; others only adopted the new assessment to some extent; yet others resisted it completely and said that their teaching was unchanged. Attributing these diverse responses to the differences between the philosophy underpinning the new assessment and the teachers' own beliefs about teaching and learning, Burrows (2004) suggested that evidence of resistance on the part of some teachers implied that washback could only occur to the extent that the participants

allow. However, given the apparently low stakes of the new assessment, it is not clear if the teachers in this study were subject to strong pressures from other stakeholders in the educational system, meaning that they may have been better placed to resist the proposed changes than teachers working in a context of high-stakes testing.

Ferman (2004) found that a grade 12 Oral Matriculation test in Israel for EFL had both positive and negative impacts. Teachers were able to focus on the skills tested, encourage oral skills development and develop test familiarity, which were all beneficial practices. However, adverse effects were observed, such as a reduced scope of teaching, increased anxiety among teachers and students regarding the test results and intense pressure to cover all the test materials. This indicates that what is taught is tested and acquired (Spolsky, 1995). Moreover, the negative washback effect resulted in a greater emphasis on more straightforward teaching tasks and less on skills requiring more complex thinking. According to Messick (1996), washback's positive and negative impacts can be traced back to the implementation and utilisation of a test, resulting in either good or bad practices. Another main finding in Ferman's (2004) study was that there were divergent perspectives among different groups of participants. While the inspectors and students believed that the test was having a considerable impact in improving both oral and reading skills, the teachers felt it was having little effect in relation to reading skills. Ferman (2004) proposed that these conflicting perspectives may be due to teachers' reluctance to endorse the satisfaction expressed by the inspectors about the test's outcomes.

Shohamy et al.'s (1996) study painted a more complex picture, examining the effects of the Arabic test (ASL) and English test (EFL), implemented nationally in Israel. The researchers explored the views of teachers and language inspectors through interviews while questionnaires were administered to students. The results showed that the teachers viewed the ASL test as low quality and although many students considered it necessary, they were concerned about taking it. This was in contrast to the EFL test, which engendered considerable anxiety in both teachers and students. The teachers were anxious that their students' results might reflect poorly on them, while the students were worried that the test could negatively affect their future studies. Thus, both the language being tested and the importance of the test influence washback among teachers and students.

McCarthy (2008) conducted a study which identified differences in responses among teachers in low-income schools and those in high-income schools: whereas the former

experienced great pressure to increase their students' scores, the latter generally felt the school environment provided good grounding and were confident that their students would succeed without much intervention on their part. Here, the different school settings resulted in the same test stakes generating different effects among the two groups of teachers, leading eventually to a contrasting impact on the teachers concerned.

It could be argued from this that a similar variability in the effect of testing was found in both these studies, albeit it presented in different ways. The former showed that the differing stakes of tests exerted a direct influence on the participants, whereas the latter showed that the stakes of the test were moderated by the school setting. These results echo Yamashita's (2011) finding that teachers in a low-stakes testing environment felt that tests affected the content of their teaching but rarely their teaching practices. Teachers are more likely to change their instruction methods in line with the introduction of a new curriculum or the degree of support they receive in understanding the meaning of test scores. Given the role that high-stakes tests play in determining future of students' lives, it may be said that the level of teacher anxiety will keep increasing as long as the status of tests continues to rise.

However, some studies (Lane & Stone, 2006; Stecher et al., 2004; Tzagari, 2011) reported either a positive impact or a totally no impact on teachers' attitudes and feelings. Stecher et al. (2004) focused on the impact of the new Washington Assessment of Student Learning (WASL) test on the teaching of writing. The study conducted two state-wide surveys among principals and teachers. The study found that the new test reform had positively influenced the curriculum and the methods used to teach writing. Teachers reported that they could allocate additional time to specific areas of writing instruction. Although there was no definitive proof that teachers prioritised test content and format over state standards, it was apparent that they did not explicitly prepare students for the test. Additionally, Cholis and Rizqi (2018) used a questionnaire to examine the effect of a high-stakes test (the SBMPTN in Indonesia) on teachers' attitudes and teaching methods. Although the findings showed that the test caused teachers to undertake more lesson preparation tasks and increase their planning to help students meet the requirements of the test, the tendency to teach for the test was viewed as a positive practice, since the teachers were still teaching according to the wider curriculum rather than to the specific focus of the SBMPTN. This meant that the test led teachers to pay

more attention to the integration of the four main skills, despite the fact that only reading was actually assessed in the exam itself.

Research has found that teachers' positive beliefs and perceptions concerning the demands of innovations in testing are determined by the local context, time needs to be allowed for the introduction and adoption of the change and there needs to be recognition that the new ideas will be interpreted in different ways from one teacher to another (Burrows, 2004; Turner, 2009; Wall, 2005; Watanabe, 1996). Turner's (2009) study focused on teachers' responses to a new high-stakes English as a second language (ESL) provincial exit examination in high schools. The main focus was on exploring formative assessment and summative assessment in the classroom teaching context in light of the washback dimensions of value (positive or negative) and intentionality (intended/unintended). The data revealed a "blurring of formative and summative assessment" (p. 118) in the alignment of classroom assessment practices with the new exam requirements. It appeared that teachers believed that adapting their teaching practices according to the format of the examination construct (e.g. speaking with peers) was essential, yet they interpreted this alignment in different ways. Some teachers were happy to follow the exact format and requirements of the examination task; some teachers felt restricted by the new format; some adopted different, broader strategies; others felt it appropriate to prepare students for the provincial speaking examination by conducting regular practice throughout the year. The study concluded that while the teachers' behaviour and perceptions may have had common characteristics that indicated positive washback, the variations in the teachers' underlying beliefs and their teaching and assessment strategies refer to the complexity of the washback phenomenon in such a context. The research suggested that one possible reason for this variation could be the teachers' perspectives and stances towards the new provincial examination. This is reflected in Fullan and Stiegelbauer's (1991) comment regarding innovation:

If we know one thing about innovation and reform, it is that cannot be done successfully to others. It is not as if we have a choice whether to change or not. Demands for change will always be with us in complex societies; the only fruitful way ahead is to carve out our own niche of renewal and build on it. (p. xiv)

Studies have shown that tests designed to enhance teaching and assessments that align with classroom instruction can improve education (Moss, 2003; Tan & Turner, 2015; Turner, 2009). Messick (1996) suggested that the move away from learning to test-taking should be seamless, with little differentiation between language learning activities and

test preparation. Tan and Turner (2015) conducted a study on the high-stakes Secondary Five ESL exit writing examination developed by the Ministry of Education (MELS) by examining the behaviours of 11 trained teachers and 2 MELS assessment officials in rating a sample of students' writing. The study analysed the raters' behaviour and undertook content analysis of the interviews with both groups of raters, information from teacher and rater surveys and notes from the marking centres. The main findings showed that the teachers' experience of practising the rating processes with the MELS officers was useful in employing the marking rubric. The teachers gained a better understanding of the MELS officers' expectations concerning each criterion in the rubric and the grade level during this experience, which helped to improve the validity of their marking judgments. The study revealed that the teachers became confident that they could enhance their students' language skills and therefore they felt empowered to modify their lesson planning and teaching methods to match the needs of students' writing ability. The researchers emphasised that because the examinations closely reflected the teaching content and standards, this reinforced the intended washback effect.

Zhang (2021) conducted a study of teachers and students' perceptions in relation to a revised test in China, the Test for English Majors Grade Four (TEM4). The study employed questionnaires and a follow-up interview and found that teachers and students expressed positive views towards the test's design, administration and content. However, participants also raised concerns regarding its validity, citing issues such as the need for more relevance to university coursework and its apparent excessive difficulty. The researcher attributed this negative washback to a lack of resources devoted to ensuring participants understood the intentions of the TEM4 developers. Insufficient training in assessment and testing may also have accounted for the stakeholders' differing opinions. The study revealed that students prepared for the test primarily by practising previous past papers or sample published test items. Teachers also seemed to rely on test-oriented activities in teaching. While test preparation practice is considered detrimental in the literature, this study showed that the strategy can be effective. This was mainly because the students perceived item practice to be helpful in increasing their test scores and teachers of low-performing students considered it a good way of increasing their students' motivation for learning.

Kennedy and Lui (2013) conducted a similar study that corroborated these results. This study explored the views of students and teachers concerning a high-stakes English test

in China. Questionnaire and interview data indicated that the teachers and students believed that English classes should focus on preparing students for the final-year test rather than on improving their language levels. For the teachers, teaching to the test was useful in this respect, but practising materials and activities not related to the test was unnecessary. The researchers emphasised that the teachers' perceptions were influenced by changes made to the content and design of the test, as these changes prompted them to alter their teaching activities.

3.6.2 Perceptions of test writers

Several studies support the idea of recruiting different groups of participants when investigating washback. Although some of these participants figure in the main washback models (e.g. Bailey, 1999; Hughes, 2003), many washback studies narrow their scope to treating just students and teachers as the principal participants in the assessment process (e.g. Alderson & Wall, 1993). Regarding assessment reform, those who develop assessments must carefully consider whether their proposed assessment is suitable for its intended purpose. Sometimes, the interpretations of teachers, students and other stakeholders may differ from those of the developers. These developers aim to create assessments that will yield positive outcomes and results. However, despite their best intentions, their efforts may not always be successful. Hence, the assessment may not have the desired impact, or worse, it could have unintended consequences that may cause harm to those involved (Bachman & Palmer, 2010). This is why Shohamy (2001) suggested that the intentions of policymakers and test designers should be one focus when investigating the washback effect from testing. However, there has been little discussion in the literature of policymakers' intentions towards examination reform and what they do to facilitate intended washback.

Some contributions have been made, though, towards addressing the intentions of other personnel in the education process (e.g. decision-makers, school administrators, supervisors, textbook writers, inspection bodies, etc.). Alderson and Wall's (1993) empirical study in Sri Lanka, for instance, was the first to investigate top-down efforts in introducing changes in national English examinations. Their study listed the expected washback effects, based on their analysis of the official statements concerning the goals for the test and the textbook series it was intended to support. However, the study did not describe the design of the test and how it would fulfil the rationale and goals of the revised approach. Therefore, the researchers stated that it would be useful for their study to

complement the data obtained from classroom observations with interviews or questionnaires. Andrews (1995) investigated the impact of a new exam-related textbook on the pedagogical strategies used by teachers in preparing students for the examination. The researcher used parallel questionnaires to examine the perspectives of the Use of English (UE) oral examining body members and teachers in Hong Kong. The findings revealed a mismatch between the views of the UE oral examination members and those of the teachers. These contrasting views were related to the relative importance and goals of the different pedagogical strategies. Interesting as this finding is, there was no explanation as to why this was the case. Another principal finding was that the views of test designers and textbook writers regarding the expected forms of washback were diverse. Cheng's (1998) study did address the fact that the Hong Kong Examination Authorities (HKEA) had failed to specify their intentions for the new test and did not seek to investigate the tests constructors' intentions in any way.

Shohamy (2021) argued that testing can be used a tool to influence the behaviours of test takers and others concerned to meet tests' requirements. Tests may be used to dictate what teachers and students need to know, what they will learn and what they will be taught, as a so-called "disciplinary tool" (p. 17). Underlying this coercive, top-down approach, in which policymakers and test constructors try to shape classroom teaching, is the assumption that testing possesses the power to exert the desired changes in teaching and learning because of its consequences. However, Qi (2007) found that the assumption that high-stakes tests can bring about desirable pedagogical change is not supported. Her study compared policymakers' intentions to the actual classroom practices of teachers to determine any discrepancies between them. She found that the teaching of writing in Senior III did not occur communicatively in a simulated real-world situation as intended by the test creators. Instead, teachers focused on grammatical accuracy and continued using the traditional instructional approach. Therefore, to gain a more comprehensive understanding of the role of testing and its objectives, it is necessary to identify the expectations of washback.

A study conducted by Fan et al. (2020) examined how recent changes to the Test for English Majors Grade 4 (TEM4) impacted English language instruction and learning in China. The researchers surveyed tertiary-level teachers to gauge their opinions on the changes made to the test and whether they aligned with the objectives of the TEM4 committee members. The results showed that teachers had a positive view of the new

TEM4. They found it a well-designed and valuable tool for assessing students' language abilities. Teachers praised the integrated read-to-write task, which accurately reflected real-world language use. They also noted that incorporating integrated writing activities helped motivate advanced learners. The interview findings also revealed that teachers' perceptions were well matched with the views of the TEM4 examination members. However, the researchers reported that there was some divergence between both the participants in terms of their views of the consequences of the test on teaching and learning. While the teachers felt that the new format of the listening tasks was less authentic due to the removal of news broadcasts, the test developers explained that the news broadcasts were limited in terms of their expected value for measuring listening comprehension. Because the test developers' intentions concerning the new listening test were unknown to the teachers, they interpreted this change as discouraging the use of news clips to teach listening skills. Significant as this study was, it did not provide information about the specific areas of the TEM4 examination members' intentions or their expectations of teaching and learning. Another point is that this study was undertaken during the post-implementation stage of the reform process and it would perhaps have been more useful had it been conducted throughout the assessment reform, as advocated by So (2014).

Winks' (2011) study of the English Language Proficiency Assessment (ELPA) used an online survey to examine teachers' perceptions. The responses were both positive and negative. Teachers were critical about the impact the test had on the ESL curriculum and its psychological effect on students. They expressed the view that the recently introduced test was excessively lengthy for kindergarten children. Additionally, they claimed that the style and requirements were not fitting for the developmental stage of young children. However, they also revealed some positive attitudes due to the enhanced focus on English language learners and the increased funding and educational opportunities provided by the test. The study showed that the opinions and feedback of teachers can provide valuable insights into the impact of tests. While this research serves as an excellent example, it did not consider the perspectives of the test designers, nor did it compare the viewpoints of the two groups. Fan et al. (2020) showed that teachers may not understand the principles underpinning testing reform and thus respond in ways that are not intended or anticipated. However, Winke (2011, p.652) argued that the different perceptions of teachers need not necessarily be regarded as due to misunderstandings but may rather be legitimate doubts about the broad concept of test validity and its property based on their own knowledge

and experiences. These two studies thus suggest that the process of tests' reform is based on the extent to which key stakeholders understand and accept the specific expectations and/or preferences for test formats as intended by the test designers (Green & Andrad, 2010, p.331).

East (2015) conducted a study to examine how teachers felt about implementing a high-stakes English language speaking test, called "Interact", for foreign language students in high schools in New Zealand. The new test was designed to replace the previous test, "Converse", and was significantly different in its approach. Interact included a series of spontaneous peer-to-peer interactions throughout the year, while the old test consisted solely of a single conversation between the teacher and the student. Survey results revealed that the new test was perceived as much more helpful in accurately assessing students' spoken language skills. However, it is worth noting that this study only explored teachers' perspectives on the matter, despite the researcher's suggestion that it is crucial to consider the opinions of both teachers and students.

Gaining insights into teachers' perspectives on tests during the development process can enhance the validity argument in interpreting and using newly created tests according to So (2014), who used the design of TOEFL test, as an example to demonstrate that teacher feedback was essential in making design choices throughout the test's development and implementation. The study focused on teachers as the primary participants since their input was crucial to the assessment developer's teams. The researcher's findings demonstrate how including teachers' input in developing an international language test can significantly impact the test design. However, the study only involved 10 teachers, limiting its generalisability.

Zou and Xu (2017) investigated washback from the Test for English Majors for Grade Eight (TEM8) and its implications for fairness and validity based on questionnaire data from 724 university programme administrators, including programme heads, English department heads and a Vice-Dean of English language teaching. The study sought to examine washback in light of the participants' perceptions of the former teaching syllabus contrasted with the revised TEM8 test syllabus. These participants were selected as they would likely be knowledgeable about the test and syllabus and about instructors' teaching methods in their respective universities. Hence, their perceptions could "provide illuminating insights into the reform of the test and teaching and learning as well" (Zou

& Xu, 2017, p. 152). The results showed that most participants were familiar with TEM8 and overall held positive views of the course design. They also reported that the test affected the teaching of skills in training courses in certain universities, even though this was not mandatory for the advanced stages of the undergraduate programme. The researchers considered that the participants provided valuable insights into the washback effect of the test. However, it is important to note that the study only involved programme administrators; the perspectives of other stakeholders, such as test developers and users from different groups, could have provided even more insights into the washback effect. Additionally, the study relied solely on a questionnaire survey, meaning that the data collected were based on self-reporting and might not have accurately reflected the actual effects of the test on teaching and learning.

Several studies have compared language testers' and teachers' perspectives in language assessment literacy (LAL) studies. In one such study, Malone (2013) investigated both testers' and teachers' perspectives concerning the content of an online LAL course. The findings revealed that the two groups had differing orientations and viewpoints when reacting to a tutorial on basic language assessment. The testing experts focused on the technical precision of definitions and aspects of proper test usage, while the language teaching experts highlighted the "ease of use with regard to presentation and delivery, as well as clarity of definition" (Malone, 2013, p. 342).

In light of the studies reviewed above, it can be concluded that research in general has identified an overall negative washback from testing where teachers trying to cope with the requirements and demands of examinations. Some examples of the negative washback on teachers' attitudes and beliefs include the following: loss of instructional time on extensive practice of the examination format; narrowing of the curriculum with an increased emphasis on skills that appear in the examination; an increase in test scores without any corresponding rise in students' skills. Positive washback on classroom teaching can be found in fewer studies, but is evidenced in some, as follows: when teachers integrate task characteristics in their teaching; when they are invited to share their opinions and experiences in testing reform process; when teachers' awareness and understanding of tests' validity and its requirements are increased; when teachers are provided with the necessary training prior to a new examination. Hence, positive attitudes to testing demonstrate contexts in which teachers are participated and guided in testing development process. Solomon (2002) labelled this "engagement and nurturing".

Furthermore, the literature suggests that there are variations in washback effects among teachers. As Monfils et al. (2004) stated, “Test preparation is not a uniform thing. There are different forms to prepare children (students) for a test...” (p. 52). In their study, the teacher variables that apparently contributed most to the variation in test preparation practices in the classroom were their beliefs and attitudes about teaching and learning, their knowledge about a test’s status and its stakes and their knowledge about test design and how test scores would be used. Shih (2009) stated that washback varies from one teacher to another and that it is a phenomenon which is inextricably linked to the context under investigation. When teachers prioritise achieving high scores on a test rather than promoting English language proficiency, or when external decision-makers impose such beliefs, the negative impact of washback is most pronounced. As a result, anticipating and managing washback is a complex undertaking. Alderson and Wall (1993) argued that the effects of tests may not be predicted but must rather be established through investigation. This raises the issue of how washback can be measured and observed. It further shows that because of intervening variables, washback is more complex than it might have initially appeared (Bachman & Palmer, 1996).

Although many washback studies have relied exclusively on questionnaires to investigate washback on participants’ attitudes and perceptions, with the advantage that it is possible to make comparisons across a large sample, the results obtained may not provide sufficient data about the research context. This is because the response options in the questionnaire may not reflect participants’ understandings of the study context. Wall and Alderson (1993) point to inadequacies in studies relying on questionnaire data with no other forms of evidence, but they acknowledge that questionnaire data are useful in explaining teachers’ behaviour in classroom and probing their beliefs and perceptions. Therefore, there is a need for a more flexible method to explore teachers’ understanding and beliefs and reveal the complexity in their world views. In this regard, Watanabe (2004) pointed to the advantages of qualitative research methods, such as interviews, in understanding washback in a particular context.

Stakeholder involvement plays a crucial role in successfully implementing testing policies in educational systems. The effectiveness of a new examination in a local context depends mainly on the extent of various stakeholders’ involvement in the implementation process. There is an interplay not only between the groups of washback participants, such as teachers and test writers or teachers and students, but also between the participants

within each group. In the following section, this relationship between washback participants is explained to understand how washback operates and to tease out the complexity and multi-directionality of the phenomenon.

3.7 Washback Processes

3.7.1 Washback on teaching content

Most washback studies have addressed the effect of examinations on teaching processes and in particular on the content of teaching (Chen, 2002; Cheng, 1997; Hawkey, 2006; Saglam, 2018; Saif, 2006; Tsagari, 2011; Wall & Alderson, 1993). The examinations investigated seem to have had an effect on the official materials, textbooks and curricula that teachers use in the classroom – both in relation to teaching and the type of skills that they focus on – and also in terms of examination preparation (Chen, 2002; Cheng, 1997; Saif, 2006). In Cheng’s study (1997), for example, the introduction of the revised HKCEE resulted in the publication and adoption of new textbooks and teaching materials. As Cheng (1997) remarked, “by the time the examination syllabus affected teaching in Hong Kong secondary schools in the 1994–95 academic year, nearly every school had changed their textbooks for the students. Almost all textbooks are labelled specifically ‘For The New Certificate Syllabus’” (p. 50).

Many commercial test preparation materials and past examination papers are used in classroom teaching to prepare students to sit examinations (Hawkey, 2006; Hoa, 2020; Shohamy et al., 1996; Tsagari, 2011). As many studies have shown, teachers rely heavily on new textbooks as a main source for teaching, making extensive use of materials based on examination content and format (Andrews, 1994; Azadi & Gholami, 2013; Belkbir, 2019; Birjandi & Taqizadeh, 2015; Hwang, 2003; Li, 1990; Read & Hayes, 2003; Tsagari, 2011). Andrews (1994) investigated the impact of the Hong Kong Use of English (UE) test assessing the oral component and found that “two thirds of class-time is spent working with exam-focused published materials” (p. 78). Similarly, Read and Hayes’s (2003) washback study in New Zealand found that 90% of the 60 schools used commercial test publications for IELTS preparation courses, whereas only 30% of 19 schools used in-house materials for their own teaching programmes.

As has already been emphasised, reviewing the literature reveals the complex nature of washback on teaching content. Both types of washback effect can occur; negative and positive washback may even exist at the same time. Additionally, the degree of washback

intensity on what teachers teach may vary depending on other circumstances. Various examples of teachers' behaviour in the literature have revealed the way in which examinations compel them to improve students' scores, leading to a neglect of teaching materials that do not seem important for the test (Smith et al., 1990), adjustments to lesson plans to fit in with test formats (Herman & Golan, 1991) and changes in the teaching syllabus to reflect the content of mandated tests (Mathison, 1987). Shohamy (1997) suggested that new tests can be used to create a "de facto curriculum" (p. 344), meaning that teachers align their classroom instruction to the test and thus create a narrower curriculum that ends up over-riding the existing one.

This is illustrated by Hwang's (2003) study of the influence of an entrance examination on teaching and learning in South Korean secondary schools. Both general and vocational schoolteachers in this study agreed that they could not teach what they felt was appropriate in terms of their students' needs and their own philosophy of teaching but had to limit the curriculum and use of materials to focus on the text. Moreover, whereas the general teachers covered all the listening and reading elements in the official curriculum, the vocational teachers were not able to teach the whole of the syllabus. They felt this was because they had not been allocated the necessary time to teach English in their schools. Other teachers believed there was a misalignment between the official curriculum, the textbooks and the test.

Similarly, in a recent study, Belkbir (2019) investigated the possible influence of the Moroccan baccalaureate examination on teachers' behaviour in second-year baccalaureate classes. The research used various observation techniques to record teachers' language-teaching methodologies and other aspects of their teaching, as well as the supplementary materials used in class. Classroom observation was conducted for a period of two months, focused on six teachers. The analysis showed a negative influence on teachers' behaviours. The researcher attributed this to an apparent emphasis on grammar, reading comprehension, and writing – the targets of the examination – at the expense of other language skills.

A study conducted by Mutereko (2017) used questionnaire and interview methods to understand the washback of the National Senior Certificate (NSC) examinations on teaching practices in South Africa. The results indicated that the use of the NSC examinations as a source of accountability could be leading to negative practices in the

classroom, such as the manipulation of test scores and a narrowed-down curriculum. It also seemed to be forcing teachers to teach at an inconvenient pace to complete the syllabus on time – what Mutereko (2017) refers to as the “finish syllabus syndrome” (p. 138). In similar terms, although the curriculum in Costa Rica promotes the use of communicative techniques, the national standardised English test, the Bachillerato, only assesses reading skills, causing teachers and students to focus on these over the other communicative skills mandated in the English curriculum (Quesada Inces, 2001). Once again, teachers and students were found to practise the topics most relevant to the examination. The teaching objectives were aimed at developing reading skills by focusing explicitly on grammar and vocabulary (Quesada Inces, 2001).

A study by Saif (2006) of a single teacher used observations of class activities and an interview to examine how this particular teacher taught for a test. The findings offered different results to the other studies discussed so far. Here, the researcher found a positive relationship between the test under investigation and how the teacher taught, with the changes in classroom practices largely based on the contents and objectives of the test. Similarly, a study conducted by Xerri and Briffa (2016) reported that the new speaking component of a test resulted in positive washback because the teachers themselves were responsible for the design. The researchers reported that there was explicit attention to the development of speaking skills in the classroom, even though the teachers were operating in an educational system that prioritised writing and reading skills over oral skills. Teachers changed their old teaching practices to accommodate the new speaking component, incorporating new opportunities for students to speak in class.

Some studies have found that curriculum narrowing can still result from examinations which are based on the communicative language approach and incorporate all four language skills. One example of such work is Tzagari’s (2011) study of the washback effect of the First Certificate in English (FCE) test in the context of Greece. The teachers here reported that the test impacted the content of their teaching and they concentrated on grammar and vocabulary, believing these were crucial for students to perform well in the test. Thus, they placed considerable emphasis in class on reading and writing tasks, as tested, while spending less time on speaking and listening. Tzagari (2011) considered this type of curriculum narrowing could be due to aspects of the test design, such as the relative weighting of the four skills and the nature of the marking criteria specified for the speaking skills. It could also be the result of the absence of training opportunities,

teachers' misunderstanding of the underlying principles of the test, or the local practices in mainstream Greek schools. These findings again indicate that the impact of test design on teaching content is not always linear since it was moderated here by external factors beyond the test itself.

Other research has found that test changes can also have little or no impact on teaching content. Although the reform of test design aimed at developing teaching has been discussed in much research, there is no guarantee that changes will actually happen. This can be seen in Shohamy et al.'s (1996) study in Israel on the long-term effects of two language examinations (the Arabic ASL and the English EFL). After several years of testing, the researchers reported that although both tests had some initial impact on teaching when they were first implemented, the situation changed over time. The ASL had almost no impact a few years later: teachers were no longer teaching to the test, there was little preparation for it, they had little knowledge of the test design and content, those who knew about it believed it to be of poor quality, and the materials had not changed for several years. In contrast, the EFL test appeared over the years to have an impact on what and how the teachers taught. Teaching materials had been reformed, the key test stakeholders were highly aware of its design and use and it had a high value among them; it was regarded as important and the test was still creating anxiety among its users. Shohamy et al. (1996) concluded that the washback effect of testing could take different forms over time and that its form would be based on different external factors: test stakes (high-stakes vs low-stakes), the purpose of the test, the status of the assessed language within society, the test design and its content.

Some studies have provided more detailed data on the variability of test impact on the teaching content and the factors that contribute to this. Cheng's (1997) study of the revised HKCEE examination in Hong Kong found that the new examination exerted a considerable influence on the type of materials employed in the lessons and on the activities teachers were using. The researcher suggested, however, that these changes only related to the form of teaching, not the substance, and that teachers' perceptions of change were determined more by the textbook publishers' interpretation of the examination than by their own experience. Similarly, a study by Chen (2002) found that one of the significant changes in the content of teaching in Taiwan caused by the reformed entrance test, the Basic Competence Test (BCT), was that teachers integrated oral skills in their classroom teaching. However, the researcher reported that this change was not simply due

to the test itself as it did not actually assess these two skills. Rather, these skills were included in the new textbook and other curriculum materials issued by the MoE. Given the key role the textbooks were seen to play in the study context, the changes in the textbook content for nationwide junior high schools explained why teachers emphasised listening and speaking in their classes.

Alderson and Hamp-Lyons (1996), in their study of TOEFL tests, conducted observations of two different types of classes taught by the same teachers: TOEFL preparation classes and “normal” language proficiency classes. The teaching behaviours of each teacher differed between the two types of classes, “but ... the differences between the teachers were as great as any difference between the two classes” (p. 290). This led them to conclude that a test is not the exclusive initiator of washback, but that a complex network of external factors is actually involved, including:

...the status of the test, the extent to which the test is counter to current practice, the extent to which teachers and textbook writers think about appropriate methods for test preparation, and the extent to which teachers and textbook writers are willing and able to innovate. (Alderson & Hamp-Lyons, 1996, p. 296)

Along the same lines, Watanabe (1996) studied the influence of university entrance examinations on the use of the grammar-translation teaching method in Japan. The researcher compared the lessons taught by two university teachers, one who emphasised grammar translation and one who did not. The researcher observed the types of activities in various private extracurricular institutions at Yobiko (examination preparatory schools) in Tokyo to gather data. Based on the findings, Watanabe (1996) concluded that the entrance examinations focused on translation affected some teachers but not all. Like other researchers, he suggested that teachers’ personal characteristics (in this case, educational background and/or experiences of and attitudes to teaching) and the examination timing outweighed the possible effects of the entrance examinations on how teachers conducted their lessons.

Finally, Qi (2005) reported significant findings from her investigation of the impact of the National Matriculation Test (NMET) in China. The NMET had two main functions: to select students for higher education and to exert positive washback on English teaching and learning as proposed by the test developers. The main findings were that the examination format and content exerted a significant impact on what teachers taught in the classroom, but not the kind of influence that was intended by the test constructors.

She concluded that the main reason for this was that the aims of the “selection function” and the “function of promoting change” conflicted with each other: high-stakes testing puts pressure on teachers to strive for good results rather than to help their students develop their language proficiency. Qi (2005) argued that neither high-stakes testing nor low-stakes examinations could produce the intended washback proposed by the test’s constructors and the policymakers.

3.7.2 Washback on teaching methods

Some washback studies have found substantial positive washback on how teachers teach. Amengual-Pizarro (2009), for example, investigated the washback effects of a high-stakes test on various aspects of teaching. The survey data showed that most of the teachers contended that their in-class methods seemed to be test-related. The teachers reported that they would change their teaching methods and pay more attention to students’ oral skills if they were allowed to teach other than for the test. They argued that it would be beneficial if an oral component were added to the test to improve teaching strategies in language lessons. The survey study concluded that these findings confirmed Alderson and Wall’s (1993) hypothesis that tests influence “how” teachers teach (hypothesis 4). In similar terms, Stecher et al.’s (2004) examination of the reform of the writing skills section of the Washington Assessment of Student Learning (WASL) test indicated that changes introduced in the test induced a strong positive influence on teachers’ behaviours in writing lessons. The effects of these changes on classroom practices improved the standard of student learning and raised scores on the test.

Yu (2020) used both questionnaire and interviews to explore the influence of a new English writing test on teaching practices in high schools in China and found that the teachers devoted more effort to teaching English classes as a result of the new test. The emphasis on writing capabilities (such as coherence, cohesion, creative thinking, summarising, etc.) tended to cause teachers to focus more on these topics as a way of achieving higher marks. Thus, changes in the teaching content, method and teaching time, and targeted training resulted in developing students’ language ability.

Other studies, however, have found variability in relation to how teachers teach, namely that washback can vary from one teacher to another and that individual teachers employ different teaching strategies in response to an examination (Hamp-Lyons, 1996; Hayes & Read, 2004; Watanabe, 1996). For example, Hayes and Read (2004) compared two IELTS

preparation courses in two different schools in New Zealand using classroom observation. The study found differences among the teachers and schools in the implementation of the two courses. For example, the washback in school (A) was negative as the teacher's methodology was entirely based on tasks geared towards the examination. In contrast, the teaching methodology for the classes in school (B) appeared to address students' academic language development rather than test familiarisation. The researchers pointed out that these differences were due to the nature of the course.

In a study investigating the English National Examination (ENE) in Indonesia, Furaidah et al. (2015) examined how teachers at senior high schools prepared their students for the ENE, focusing on two different groups of schools, one for high achievers (SHIGH) and one for low achievers (SLOW). The study found evidence of washback in the two types of schools, but there were differences in the degree of washback intensity, which was noticeably stronger at the SLOW schools. The SLOW schools allocated more time to preparing students for the ENE than the SHIGH schools. The researchers concluded that:

...the washback manifested in the increased amount of time allocation for UN [Ujian Nasional or National Examination] subjects is likely to be beneficial as it can potentially support the development of communicative competence as stated in the standards of competences of the English subject. (Furaidah et al., 2015, p. 51)

They also reported some tentative links between the time spent on practising for the test and enhancing students' communicative competencies. However, a key finding in this study was that although there was variability in the amount of teaching time spent on test-oriented activities in the different schools, the teachers used the same methods of teaching in both groups of schools.

Washback studies that have investigated the role of classroom assessment practices in influencing teaching and learning have also resulted in mixed findings. One example here is a longitudinal study by Wall and Horak (2006), which investigated the impact of a new TOEFL test on teaching, learning and classroom assessment in parts of Europe. The researchers explored the validity of these new tests, as they tended to focus on writing skills rather than giving balanced weight to the other tested skills. In addition, there was evidence that the marking of writing skills was based on the teachers' own feedback, rather than on the TOEFL writing scale. The researchers noted that the teachers needed more guidance on this scale, as they did not feel confident in using it. Wall and Horak

(2006) also stressed their concerns regarding the reliability of these tests as they were administered under different conditions than the TOEFL itself. Ultimately, the study concluded that these tests were helpful in familiarising students with the TOEFL test format and structure and would help students to prepare adequately for the TOEFL test, but that the tests would not improve students' language proficiency. In light of studies such as this, the influence of high-stakes examinations on classroom assessment practices appears to vary, ranging from no reported washback to considerable washback. Some researchers have suggested the need for further research into the reliability and validity of classroom-implemented tests (Turner, 2012; Winke, 2011). Certainly, more research is needed to address how high-stakes examinations influence classroom assessment and this is an issue which this study aims to scrutinise.

Other research findings in this regard have claimed that tests have no or only limited washback on teaching methods. Wall and Alderson (1993), for example, in their Sri Lankan washback study, raised the point that while there was a change in the content of teaching, the teachers continued to teach in the same way in all their classes. Similarly, in Cheng's (2005) work on the HKCEE, the data collected indicated that there was no washback on how the teachers taught in the classroom. The washback of the examination influenced what the teachers taught rather than how they taught it. For instance, read-aloud activities were exchanged for role-playing scenarios and group discussions, but both activities were taught through drilling. While the examination was underpinned by a communicative and task-based approach to teaching, the teachers' behaviours remained test-oriented. Teaching was highly teacher-centred and content-based, thus exhibiting a negative influence of the examination on pedagogy.

Examining the impact of the Basic Skills Test (BCT) in Taiwan, Chen (2006) reported that even though it was designed to encourage CLT and foster the communicative abilities of students, the teaching methods focused on de-contextualised language rather than communicative abilities under the influence of the test. Glover (2006) sought to explore the influence of the Hungarian school exit examination on teaching methods by comparing teachers in two different contexts: examination and non-examination lessons. The findings indicated that a few aspects of how the teachers taught showed some slight empirical evidence of a washback effect, but other elements were not affected by the examination. The researcher attributed this variation in teachers' classroom practices in

examination lessons to the fact that “the teachers have a different pedagogic purpose: teaching the examination, not the language” (p. 324).

Looking at the variability in the results of the above studies, the complexity again indicates that the existence of a certain test in itself will not be the only source of washback. The findings from washback studies in different contexts, such as Sri Lanka (Wall, 2005; Wall & Alderson, 1993), Israel (Shohamy et al., 1996), Hong Kong (Cheng, 1999), and China (Qi, 2007), all support such a claim. Shohamy (1992) suggested that although the use of tests for the purpose of bringing about change has become a well-known phenomenon in various educational contexts, researchers have been sceptical about the extent to which a reformed test in itself can improve teaching and the use of tests to affect and drive teaching practices may not guarantee that any meaningful change will occur. The words of Alderson and Wall (1993) are also relevant here:

...a naive deterministic view would assume that the fact of a test having a set of qualities is sufficient in itself, by virtue of the importance of tests in most societies, to bring about change. However, this takes little account of other factors in the environment which may also influence teaching: the teachers’ basic competence, their understanding of the principles underlying the test, levels of resourcing within the school system, etc. (p. 118)

Alderson and Hamp-Lyons (1996) made a similar point, as did Bailey (1996), who commented that “the processes involved in washback will vary widely, depending on which constituency of participants we consider” (p. 264). Hence Watanabe (2004) asserted that washback is “a highly psychological phenomenon” (p. 31), meaning that it is much too simplistic to suggest that it is inherent to any test. Cheng (2000) is also worth quoting in this context:

...perhaps the single most important theoretical development in language testing since 1980’s was the realisation that a language test score represents a complexity of multiple influences. Language test scores cannot be interpreted simplistically as an indicator of the particular language ability we want to measure... What makes the interpretation of test scores particularly difficult is that these factors undoubtedly interact with each other. (p. 3)

Differences among these studies are also noticeable in terms of the research methodology, as some used classroom observation and teacher interviews, while others relied only on questionnaires. Alderson and Wall (1993), Shohamy (2001), Turner (2001), Watanabe (2004), Cheng (2005), Wall (2005) and Green (2007) have all stressed the need for a triangulated approach, employing a combination of quantitative and qualitative methods

using different sources. They also recommend the triangulation of perspectives, examining the views of key test stakeholders, such as teachers, students and test developers, and stakeholders more broadly, such curriculum writers and policymakers.

Furthermore, notwithstanding the importance of questionnaires and interviews in exploring participants' beliefs and attitudes about testing and providing evidence of how they are affected by the test, it is argued that classroom observation can provide a deeper understanding of the responses given in interviews and questionnaires and contextualise otherwise irrelevant or ambiguous responses (Wall & Alderson, 1993). However, Wall and Alderson (1993) did not find evidence of observable variability in how teachers teach. Cheng (2005, citing Bailey, 1999), agreed that observation can provide richer data on test impact than questionnaires or interviews alone and argued for the triangulation of methods. Moreover, Wall and Alderson (1993) pointed out that observation need not be constrained to classroom events but can also include analysis of teachers' classroom materials and classroom assessment work. All in all, it is important to employ various research methods, as "we need to look closely at classroom events in particular, in order to see whether what teachers and learners say they do is reflected in their behaviour" (Alderson & Wall, 1993, p. 127).

Herman and Golan's (1991) washback study used only a survey to compare teaching practices in two distinct categories of schools: those in which test scores had risen and those in which test scores had remained the same or decreased. One of their findings was that teachers reported they felt pressure from district administrators, school principals and administrators, other teachers, parents, the community and the media to improve students' test scores, but they did not establish the extent to which these pressures influenced the way the teachers taught. This example indicates that without the use of direct observation or other sources of information beyond questionnaires, it is not possible for researchers to find answers not yielded by direct questions, or to provide evidence that might enable clarification or verification of questionnaire/interview responses.

We can tentatively say from the findings of empirical research into washback on teaching methods that many studies have revealed the presence of washback on teaching methods, but the nature and consistency of the washback effect is not yet clear, and indeed has not been extensively studied. As already noted, previous studies have found that the presence of washback in relation to how teachers teach ranges from being quite considerable in

some situations (e.g. Amengual-Pizarro, 2009) to no washback at all in others (e.g. Wall & Alderson, 1993). It is also noticeable that studies have found several variations in terms of the washback of tests on teaching behaviours, with some teachers using approaches in the classroom that tend to rely on exam-related materials, while others create their own new strategies and processes in teaching. Thus, one question is whether the examination is the only source of washback causing these contradictions and variations, or whether there are other mediating factors, as suggested by previous researchers. The following section tackles these issues.

3.8 Factors Influencing Washback

Although potential intervening variables have been not widely investigated empirically in washback studies, there seems to be agreement in the literature, referenced earlier in this thesis, that several factors might contribute to the nature and intensity of washback (Alderson & Wall, 1993; Dunkely, 2010; Fulcher & Davidson, 2012; Lam, 1993; Taylor & Weir, 2009; Wall, 2005).

Lam (1993) developed a series of 10 hypotheses to examine the influence of the Revised Use Examination (RUE) in Hong Kong. The researcher surveyed around 60 teachers through a questionnaire and analysed documents and textbooks that were in use before and after the implementation of the examination. The author attempted to trace the impact of factors such as: the duration of time spent on teaching English; teachers' own attitudes and their perceptions of their students' abilities and attitudes; whether a school set aside extra teaching time for test preparation classes; the content of teaching; students' levels of achievement. The responses from the teachers indicated mixed views concerning washback with regard to most aspects and Lam (1993) gave some interesting explanations for the interaction between the different factors, developing a more complex picture of washback than the examination developers might have expected. In particular, Lam (1993) referred to "teacher culture", namely differences in the teachers' reactions to the introduction of the new test depending on their own language competence, their teaching experience, their perceptions regarding the new test design and its aims, their own commitment and motivation to the teaching profession, and the challenges and anxieties they faced in implementing changes in their classroom teaching.

These findings echo Markee's (1997) view that teachers are the key participants in dealing with such a change in the educational system, making curriculum innovation work in their

classroom. As discussed earlier in this thesis, teachers' beliefs and attitudes concerning an examination determine whether or not it has washback, as well as its effects on how they teach and plan their lessons. If it does not affect teachers' perceptions, any proposed change in testing, even one purporting to follow CLT principles and strategies for EFL teaching, will not achieve any significant impact on teachers' actual practice in the classroom (Young & Lee, 1984). Supporting this, Chapman and Snyder (2000) argued that any changes in teachers' practices are brought about by the teachers' own beliefs about the actions they need to take to prepare themselves and their students for those changes. In similar terms, Mark (2011) undertook a study to investigate the interaction between teachers' beliefs and their teaching practices. The main finding was that there is a need to trigger teachers' awareness of the effects of their beliefs on their instructional decisions.

In conclusion, teachers' perceptions concerning examination change are often used as the primary source for evaluating the implementation of a new test once it has been adopted in schools (Morris, 1988). Teachers' feelings about and reactions to the introduction of new test could be due to its normative or intrinsic features. However, teachers' perceptions regarding a new test design and use, including associated teaching activities, textbooks, content, methods of teaching and classroom assessment tools, will also be determined by various other and possibly more influential factors. The most common of these are discussed below.

3.8.1 Teacher characteristics

It is widely accepted in the literature that teachers' personal traits are essential elements that shape and determine the extent of washback. The features involved here include the teachers' educational background and their experience and training. Also important in terms of teachers' instructional practices are gender-related issues. The reason this factor is relevant in a washback study is the consistent and substantial divergence in achievement between male and female students, not only in English language, but in all subjects and national and international assessments conducted by the MoE in Oman (Al Bulushi et al., 2018; Al Kharusi, 2011; Rassekh, 2004). However, teachers' gender has mainly been discussed in relation to teacher–student interaction patterns (Hopf & Hatzichristou, 1999), and this factor has very rarely been considered in determining the occurrence of washback. Thus, the influence of teachers' gender on the washback from a

new examination is of interest in this study, alongside the other previously identified personal variables.

Watanabe's (1996) study of the washback from an entrance examination on use of the grammar-translation method in Japan found that the test had limited influence on the teaching approach in the classroom. Rather, "teacher factors, including personal beliefs, past education, and academic background, seemed to be more important in determining the teaching methodology a teacher employs" (p. 130). Similarly, Spratt (2005) concluded in his review of washback studies that teachers are essential in determining the extent and type of washback and deciding what to teach and how to teach it most effectively. In this regard, several variables can play a role in the teacher's response to the test, such as the teacher's educational qualifications, experience and gender. In Alderson and Hamp-Lyons' (1996) study, all the teachers exhibited a negative outlook towards their teaching practices related to the TOEFL test. While the test induced effects on both the content and teachers' classroom behaviours, this influence varied in degree and kind between the different teachers. In a similar vein, Watanabe (1996) considered teacher factors an important element in his study of the Japanese national university entrance examination, arguing that "teacher factors, such as educational background, personal beliefs and teaching experience may outweigh the possible effect of the entrance examinations" (p. 318).

Green's (2007) study sought to discover the influence of IELTS writing course by comparing the differences between the IELTS preparation courses and English for academic purposes (EAP) courses. The findings showed very few differences between the courses in terms of classroom organisation, student modality and teaching content. Green's (2007) determined that teacher-related factors, such as their level of professional training and beliefs regarding effective learning, played a significant role in determining specific practices that could not be foreseen based on the format and content of a test. As a result, the following paragraphs explore the various factors that potentially influence teachers' reactions and responses.

Teaching experience

Many studies examining the relationship between test washback and classroom teaching in the ESL context have provided evidence that teaching experience is a key factor that helps to explain why washback varies from one teacher to another (Cheng, 2005;

Shohamy et al., 1996; Watanabe, 1996). In Shohamy et al.'s (1996) study, the more experienced teachers showed greater positivity towards the new test and were thus happy to use it to support their classroom teaching. Another significant finding from Lam's (1993) study was that more experienced teachers were most likely to adopt a traditional exam-oriented approach, while those with less experience were more likely to use authentic materials and adopt activities requiring classroom interaction. In Lam's (1993) view, the more experienced teachers are, the more confidence they will have in their own ability to know what best suits their students' needs and the more faith they will have in the value of pragmatic teaching. Lam (1993) suggested that the disparities identified were because teachers sometimes have different attitudes, philosophies and teaching cultures regarding teaching and learning. Similarly, Cheng (2005) argued that it is difficult for less experienced teachers to change their practices in response to the introduction of a test because they lack the required resilience. This contested issue will be one of the focuses of this study, especially since English language teachers in Oman are assigned to teach grade levels based on their teaching experience: the more experienced the teacher, the higher the grade he/she is assigned and vice versa.

Teacher training and awareness of examination reform

While studies have shown that teaching experience is important for teachers to adapt and improve their teaching methods, research on the relationship between tests and instruction has indicated that proper methodological training is also vital for promoting creativity in instructional practice (Cheng, 1997; Davison, 2008; Turner, 2009; Urmston & Fang, 2008). This is evident in Cheng's (1997) study, which noted that even when teachers held positive attitudes towards the new HKCEE, they still found it difficult to adopt and practice changes in their teaching. The teachers revealed their concerns about challenges in classroom management and the availability of teaching resources, factors they felt needed addressing for them to be able to implement the new curriculum requirements in the classroom. A similar point was made by Falvey (1996), who noted that "the majority of English teachers in Hong Kong are unprepared either for recent changes to the curriculum or for pedagogical changes" (cited in Cheng, 2005, p. 16).

According to Khaniya (1990), "a large number of teachers help students cope with the examinations in order to preserve reputation as good teachers. This situation is unavoidable because of the extrinsic values of examinations" (p. 51). Alderson and Wall (1993) reported that teachers' feelings of anxiety, guilt and embarrassment regarding

students' poor results in examinations led them to seek to raise the students' examination scores and thus they taught to the test. Even so, they concluded that there was no specific sign of a change in the teachers' methodology and that teaching methods for the rest of the year were the same as before the new English examination was introduced. Thus, the findings of this study suggest that there is no evidence to support the idea that examinations affect how teachers teach, or that teachers cannot understand how to implement a new textbook if they lack training in the teaching methodology required by the new examination syllabus.

Chapman and Snyder's (2000) review also supports this finding, as they concluded that reforming a test to change teaching practices can be an effective tool, but its success is not necessarily assured. This is because policymakers fail to understand the changes required in test design, content, or use to achieve their intended washback on teachers' instructional practices. It is important to reiterate here that the degree and intensity of washback induced by the new English elective diploma examination in the Omani context may be partially related to teachers' understandings of the examination's content, format, and use. This is because teachers who are aware of the examination's requirements and purpose are more likely to adapt the proposed changes in their classroom teaching (Alderson & Wall, 1993).

Muñoz and Álvarez (2010) studied the influence of an oral examination in the EFL classroom. They concluded that it is imperative to provide teachers with the required facilities and guidance to help them understand their role in teaching and use appropriate strategies to meet students' needs. In Hughes' (1988) view, teaching to the test concerns the content of the course. Coaching specifically for the examination becomes less of an issue when teachers have more awareness of the design, format and requirements of a new test or examination.

3.8.2 Educational background

In addition to the above personal factors, teachers' academic qualifications and educational background play an important role in determining the occurrence or absence of washback and why and how it occurs or not. For example, Lodhi et al. (2018) found that teachers' academic qualifications were among the factors that affected the performance of ESL students in examinations and specifically that teachers who held a Master's qualification in linguistics made a conscious effort to develop their students'

communication skills. Similarly, Watanabe (1996) argued that one of the influences on washback is a teacher's background. His study found that teachers who held a postgraduate degree in theoretical linguistics were better at implementing effective teaching strategies than teachers who had obtained only a Bachelor's degree. According to Richards and Nunan (1990):

In second language teaching, teacher education programmes typically consist of a knowledge base drawn from linguistics and language learning theory, and a practical component based on language teaching methodology and the opportunity for practise teaching. (pp. 49–50)

Thus, variations in teachers' academic backgrounds may affect their reactions in terms of washback, an issue that warrants further investigation in this study.

Examining the literature on washback suggests that the factors beyond the test itself play a crucial role in shaping instructional practices in the classroom (Andrews et al., 2002; Chapman & Snyder, 2000; Cheng, 2005; Qi, 2007). As such, it is essential to consider these additional elements when studying washback. These studies have shown that such factors are likely to include teachers' personal characteristics, which appear to be inextricably linked to an examination achieving the intended washback on teaching practices. Lam (1995) concluded that the RUE in Hong Kong made some positive contributions to the way teachers teach, but that these contributions were limited "because of the exam-oriented English teacher culture which is influenced by many complex issues, such as teachers' language competence, personalities, the changing student learning culture, etc." (p. 96).

Researchers have argued the importance of familiarising students with the requirements of the test for which they are preparing and ensuring teachers find support and help if they face challenges in implementing examination change (Hughes, 1989). Schools need feedback and information about the test and teachers should be part of the different phases of examination development since they have to implement the change in their teaching practices (Shohamy, 1992).

To date, the main question left unanswered is why the washback from testing shows the form and the degree of intensity that it does. As a complex phenomenon, research has demonstrated that the factors underlying its unpredictable nature should not be overlooked and indeed that considerable attention should be paid to the complexities. Hence, more research on the washback phenomenon is required in different contexts

(such as Oman), especially when a new examination has become familiar to test stakeholders (Spratt, 2005). It is possible that there are still unexplored differences in the impact of the factors influencing washback in different educational systems, since every educational context has specific variables that reflect its own identity. This relates closely to Cheng's (2004) view that while substantial research has already been conducted into washback in the EFL/ESL context, further empirical research is needed to investigate washback "in specific educational settings [to] fully understand the complex issues that shape the relationship between testing, teaching and learning" (p. 148). This is why these factors were worthy of investigation in this study. Moreover, since these external factors are often associated with the variability of the washback effect on teaching, they were included in this study as independent variables, so that a deeper understanding could be achieved about the way teachers teach.

Given that this study sought to examine the educational context and testing culture of Oman, to set the stage for this discussion, the following section discusses the work already conducted in the Middle East region, not least because there has been only one study conducted to date on washback specifically in relation to Oman. Remarkably, there is no empirical research in the EFL field in Oman that could be said to have investigated and established the occurrence of washback or its process.

3.9 Washback Studies in the Middle East

Although Oman has been the focus of language teaching and testing research, there is a lack of empirical studies examining the washback effect in the country. The washback studies available in this context involve only tentative comments and reviews of the phenomenon; no concrete conclusions have been reached to support assumptions about what or how washback might appear in this context (see 1.2). Hence, it is useful to turn to work carried out in other Middle Eastern countries, which may offer insights of special relevance to Oman.

Ibrahim's (2019) investigation of the impact of the Thanaweya Amma English language test in Egypt is one of the few studies conducted in a Middle Eastern country to address the washback effect on students and teachers. Given the highly examination-oriented education system in Egypt, as in almost all Arab countries, the stakes of public examinations are of great importance, since they serve to determine which students will be allowed to continue to higher education. The Thanaweya Amma test is administered

by the MoE across the whole of Egypt, where it is taken by students in grade 12 at the same time and under the same circumstances (600,000 students took the examination in 2017). It has not been changed for more than 30 years. The study employed structured interviews with teachers and students. The overall finding was that the test design significantly and negatively impacted students' learning habits and teaching methods. The grammar-translation approach continued to be widely used in classrooms, with both teachers and students emphasising the use of Arabic in class and students memorising the language functions assessed in the examination beforehand. The research participants also indicated that they focused explicitly on speaking and listening skills because these were included in the test. However, this study relied only on interview data. There was little evidence showing which aspects of washback affected the teachers and learners, or in what ways, and the study did not provide sufficient evidence or data to support its findings. This is mainly because the interview data were not triangulated with other methods, such as direct observation, as recommended in the washback literature and elsewhere in this thesis. More flexible methods needed to be adopted to understand the participants' perspectives and to reveal some of the complexity of participant behaviour in the classroom.

Tayeb et al. (2014) investigated the washback of the General Secondary English Language Examination (GSEE) used in Yemen as an exit test for the grade 12 secondary school stage. As is the case in Egypt, the educational system in Yemen is test-driven and the stakes of examinations, especially public, are of great importance from a societal and educational perspective. As Mathews (1985) pointed out, "it is in the developing countries where the pressure of selection by examination may be most severe" (p. 23). The researchers conducted semi-structured interviews with three teachers. Based on the interview responses, a questionnaire was constructed and then administered to 30 Yemeni grade 12 English language teachers. All the teachers reported that the GSEE exerted a strong influence on how they taught because the ultimate goal was to increase students' scores on the test. This meant familiarising the students with the test requirements and design rather than focusing on language development. The teachers indicated that a third of their teaching time was usually spent on explaining the examination format and content. They also reported that their students were under constant pressure to pass the test and get a high score. According to Tayeb et al. (2014), the questionnaire data provided clarification of the interview responses and all the data were confirmed through triangulation. Although the study aimed to investigate the students' learning styles and

learning activities, as well as their attitudes and motivation, it did not account for the students' own perspectives and thus failed to confirm the teachers' concerns about their students' learning. Furthermore, the findings were based on interviews with only three teachers. More varied and comprehensive data might have been obtained had the researchers undertaken more interviews with teachers, rather than asking them to respond to a questionnaire. Even though the questionnaire data were triangulated using data from the interviews, it was still hard to assess whether the teachers' perspectives concerning their teaching practices and the students' learning process reflected their actual practices in the classroom. Similarly, the absence of any classroom observation in this study made it hard to determine whether any aspects of the teachers' classroom practice could be due to the design and format of the GSEE.

Al Sheraiqi (2010) conducted a study of washback from the introduction of the Common Educational Proficiency Assessment (CEPA) English test in the United Arab Emirates (UAE), including its possible effects on curriculum content, teaching methods and classroom assessment practices. Students in grade 12 in the UAE take this test to graduate, making the CEPA a high-stakes test with a strong impact on students. The researcher used a questionnaire distributed to 12 teachers to gather their views on their teaching content and methodology. In addition, two interviews were conducted to obtain more data about the washback effect of this test, along with analysis of the mock CEPA English tests to see whether their content was aligned with the grade 12 English syllabus. The document analysis in this study indicated that the content of the mock CEPA English test was not completely related to the content of the grade 12 curriculum. Moreover, the teachers indicated that they made changes in what and how they taught due to the test. Some were negative, such as neglecting parts of the textbook not included in the CEPA test, teaching and drilling test items and emphasising certain supplementary materials to help students become familiar with the test. Other changes were reported to be positive, including the use of new teaching methods (such as process writing) and the inclusion of student-centred activities. This study serves as a good example of washback on classroom teaching practices; however, it only involved a small overall sample of 14 teachers. Although the purpose was to discover the alignment between the content of the test and the grade 12 syllabus, the researcher used mock CEPA English tests instead of the actual CEPA test. This could have affected the validity and reliability of the data obtained from the document analysis. Another significant limitation in this study was that even though the questionnaire data were supplemented by data from interviews, the absence of

baseline data meant that it was difficult to judge whether there were any changes in the way teachers taught, or which aspects of the teaching process could be attributed to the effects of the CEPA test itself.

Haddadin et al. (2008) examined the washback of a public examination on English language instruction in the secondary school stage in Jordan. Initially, surveys were conducted with teachers and students, followed by interviews and classroom observation to examine the teachers' responses in greater depth. The study revealed strong negative washback on teaching at the secondary school level. This was evident from the data, which showed that the examination was steering teachers towards covering what was required in the examination. The test was also encouraging students to study only the skills covered in the examination, namely reading and writing. The teachers were neglecting the two other main skills, listening and speaking. In other words, teaching to the examination was leading to curriculum narrowing. Although this study investigated washback on English instruction at the secondary level, it only addressed the effect of the examination on what the teachers taught, rather than which aspects of teaching the test affected. There was no indication of whether the test caused teachers to change the way they were teaching.

In conclusion, an overview of washback research in the countries of the Middle East suggests that although high school examinations in these countries serve important functions and purposes in both higher education institutions and the local culture, empirical studies are scant. There is remarkably little research in the Arabic ELT area that can be said to have investigated why or how washback operates. From the above analysis of empirical studies, three major limitations can be seen in current washback studies in the Middle East. The first relates to the research methods employed by ELT researchers in these countries. Classroom observation has been employed rather less frequently than questionnaires and interviews, despite being highly recommended as a primary research method for researching washback (see Bailey, 1999; Cheng, 2005; Wall & Alderson, 1993; Watanabe, 2004). The second limitation of these studies is that the results they have yielded seem to be far from adequate, since in some studies the researchers seem not to be able to provide enough data to back up their claims. The third drawback is that the research evidence on how teachers teach seems to conflict with other data collected in the studies. One source of conflicting findings lies in what is taken as evidence of teaching methodology. The studies that claim to have found evidence of strong negative washback

have tended to use self-reported data from teachers rather than observation. Studies that rely on self-reports of classroom behaviour without observing teaching may indicate washback on attitudes, rather than washback on how teachers teach.

In light of the above empirical studies and the literature reviewed, there seem to be some gaps in existing research into washback on teaching. These are addressed below.

3.10 Gaps in the Existing Washback Literature

From the above review of the literature, it is clear that several washback studies have tried, in one form or another, to address a key research area – how washback from testing influences the process of teaching. Most of the research findings cited above have arrived at conclusions regarding the occurrence and nature of washback. However, despite the studies having good qualities, they are all limited to some extent. Some of these limitations are discussed in this section.

One clear limitation of existing studies is that they generally only shed light on those factors that are associated with testing itself, which indicates that insufficient research has examined the factors mediating washback. Because of this narrow focus, many assertions and claims made about the nature of washback overlap in meaning although they differ in wording (see Turner, 2009). Furthermore, although such variables, and specifically teacher factors, have been widely investigated empirically by various researchers (Alderson & Wall, 1993; Fulcher & Davidson, 2012; Lam, 1993; Shohamy et al., 1996; Taylor & Weir, 2009; Wall, 2005), additional data are still needed to examine and address in more detail the interactions and functions that connect these factors and above all, to illustrate whether factors associated with particular educational contexts also apply in other research contexts. Cheng (2004), Watanabe, (2004) and Spratt (2005) have all argued that filling this gap would be a valuable addition to the washback field.

From the review of studies on the washback on teachers' perceptions, it is apparent that most have concluded that teachers exhibit negative attitudes, such as anxiety and fear, when an examination is newly introduced. In only a few studies did the teachers show a positive attitude, while in others they expressed mixed feelings. Furthermore, there has been little discussion about what influences teachers' perceptions and/or what causes their negative reactions towards the introduction of a new examination. Similarly, there has been little focus on how well teachers react towards the implementation of a new examination, or how well-informed they are about such a change before it occurs. This

information is crucial in that it is difficult to expect teachers to be involved and interact in examination change if messages from the policymakers are not clearly communicated. These flaws are addressed in this thesis.

As clearly observed in this chapter, washback cannot be treated as a simple area of study as its effects are not direct or automatic. While studies have shown that the influence of tests is clear with regard to various aspects of teaching, the same studies have also found that washback can be superficial and there is little evidence of substantial change in how teachers teach, particular as the influence can differ in terms of both “form” and “intensity” (Cheng, 1997, 2005). Studies have also revealed that there can be inconsistencies between teachers’ perceptions about their classroom teaching and their actual practices (Cheng, 2005; Wall, 2005; Wall & Alderson, 1993; Watanabe, 2004). For example, Wall and Alderson (1993) found that while teachers claimed that they had changed the way they were teaching by emphasising the communicative approach, they were actually adopting a teacher-centred approach that the curriculum developers had not intended to encourage.

Further, the literature review has shown that there have only been a few studies to date (e.g. Alderson & Wall, 1993, Wall & Horak, 2006) that have examined the effect of examinations on classroom assessment practices. These studies have looked at the impact of classroom testing questions and arrived at various findings, from no washback in some cases to considerable washback in others. Studies conducted into how teachers teach have revealed that washback is broad and multi-faceted. Some studies have indicated a strong enough impact from washback to constrain teachers’ methods to become more test-related (e.g. Amengual-Pizarro, 2009), while others have shown that both negative and positive washback can occur together (Shohamy et al., 1996). It seems that there is a gap in the literature on the washback on how teachers teach. To attain a more in-depth understanding of the washback process, this examined the washback from the new examination by considering the central role of teachers in terms of their teaching methods in the classroom and their perceptions of what they did as a result of examination change. This approach has been emphasised in previous studies that have suggested that what the teachers’ say about a test may shape how they teach in the classroom, most notably teachers’ practices in preparing students for an examination (Norris, 2008; Spratt, 2005; Wink, 2011). This study also aimed to examine the way in which factors other than the test itself (e.g. teachers’ characteristics, the challenges they face in teaching, their teaching workload,

etc.) can facilitate or hinder the various forms and intensity of washback. This approach was designed to provide thick descriptions of washback in the research context.

The literature review indicated that there has been very little discussion of policymakers' intentions concerning examination reform and what various educational stakeholders do when introducing a new examination. For example, to understand the impact an examination has on teaching, textbook writers, test creators and inspection teams need to investigate the washback effect. Previous studies by Watanabe (2004), Qi (2007) and Shohamy (2001) have explored this topic.

A more comprehensive range of research methods has been employed in more recent washback studies than in those undertaken earlier (for further reflections on methodology in washback research, see Cheng & Curtis, 2004). Those studies that have used different research instruments have found these methods to be useful not only in describing washback effects, but also providing reasons why washback occurs or does not, or why it takes a specific form. The literature review has also suggested that studies which adopt a range of instruments to measure the influences of tests that may be considered positive or negative are likely to offer a more comprehensive picture of washback from examinations, bring to light new issues in the field of washback and increase the overall validity of the research findings. There is still a need in the literature, however, for more research on test washback using a range of research instruments "to investigate some of the apparent contradictions in the findings to date" (Spratt, 2005, p. 27).

Drawing on the above gaps identified in the literature on washback, the following section reiterates the research questions and the research method(s) used to address each research question, before introducing the conceptual framework.

3.10.1 Research questions

From the literature review above, it can be seen that most research in the field has focused on what happens in the classroom as a result of test reform (teaching behaviour, materials, classroom activities, etc.), rather than on how any changes occur (the process of examination change). It is thus necessary to study the factors that facilitate or hinder the perceived washback effects, as well as the ways in which these factors interact together (in particular, teacher characteristics). Moreover, as suggested by the previous literature, it is important to understand what determines how teachers teach. Teachers' beliefs and perceptions are critical factors which mediate the washback of testing. In addition to these

factors, investigating the intended washback as proposed by the MoE decision-makers (test developers and curriculum writers) in this context could provide insights into any misalignment between expectations and objectives among test developers and teachers concerning test reform, as well as ensuring that the introduction of a test and reform measures are taken into consideration. Formally stated, the research questions are as follows:

- RQ1. What are teachers' perceptions of the new English elective diploma examination?
- RQ2. What is the intended washback from the new English elective diploma examination according to MoE decision-makers?
- RQ3. What are the nature and scope of the apparent washback effects resulting from the new English elective diploma examination on teachers' classroom practices?
- RQ4. How does the intensity of the apparent washback effects from the new English elective diploma examination differ according to teachers' personal characteristics?

The first question explores teachers' perceptions of the washback from the new examination and how these beliefs influenced their behaviour. According to Bachman and Palmer (1996), a critical aspect in designing and developing a test is to consider its intended use. Thus, the second question aimed to determine the intended washback as perceived by MoE decision-makers. Intended washback helps understand the role of the new examination as an agent of change in teaching and learning. The third question concerns the apparent washback effects from the new examination on teaching practices and the relationship between actual teaching and the intended washback. The fourth question investigates why the perceived washback effects from the examination took the shape they did, the degree of this influence and the reasons for the presence or absence of the intended washback on teaching outlined by the MoE decision-makers. Specifically, it measured personal factors (educational background, beliefs about best practices in teaching and learning, experience, professional development and gender) that might potentially influence the intensity of the perceived washback effects on teaching.

3.10.2 Conceptual framework

Having discussed the previous washback literature, the research questions and their purposes, this sub-section describes and elaborates on how the washback effect works in relation to the conceptual framework of this study. Although the models discussed in

section 3.4 contribute to the overall picture of washback, the model in this study is adapted from Hughes' trichotomy (1993) and focuses on the participants, processes and products as these concepts are directly interrelated and represent the mechanisms by which test impact occurs in classroom teaching (see Figure 3.5). The process model outlined by Hughes (1993) made a tripartite distinction between the effects on participants, processes and products. The trichotomy is here adapted to focus on teachers and their teaching practice, to which end the study examines washback effects on teachers' attitudes and their beliefs, what they teach and how they teach. It is also important to understand how washback in this context operates with a view to enhancing the occurrence of positive washback. This study seeks to explore and explain why washback has taken the form it has among local Omani teachers, using a framework adapted from Hughes (1993).

As discussed in this chapter, the "participants" should include the educational system administrators, teachers and students, curriculum designers and examination writers (Hughes, 1993, cited in Bailey, 1996). They also need to include all those whose work relates to teaching and learning, such as users of the examination (teachers and students), parents, school principals and other examination stakeholders. Due to the scope and time available to conduct this study, however, there were only two groups of participants: those who initiated and developed the new English elective examination (MoE decision-makers) and those who implemented it (teachers). The roles and responsibilities of these participants within the Omani educational system determined the degree of intensity and nature of the washback of the examination on teaching.

I chose teachers as the key participants in this investigation because research on washback considers them to be among the most important stakeholders in the examination process and classroom assessment practices and thus it is important to take into consideration the variety of effects that tests can have upon them (Cheng & Curtis, 2004; Wall, 2005). Teachers' attitudes towards their responsibilities within the school could be affected by the examination (Bailey, 1996, p. 262). Teachers are commonly influenced by their knowledge and perceptions of what their responsibilities are towards their students and what parents expect of them in terms of how they teach (Khaniya, 1990; see 1.4 and 4.6). The model illustrates the direction in which the new elective diploma syllabus was communicated to each party at each level – macro and micro – within the Omani educational context.

According to Hughes (1993), “processes” refer to “any action taken by the participants which may contribute to the process of learning and teaching” (cited in Bailey, 1996, p. 262). Hughes suggested that processes may include “materials development, examination syllabus design, changes in learning strategies, changes in attitudes, changes in the use of test-taking strategies, changes to how teachers teach, etc.” (cited in Bailey, 1996, p. 262). This study thus explored what occurred drawing on the perceptions of the two groups of participants involved in the new examination: the MoE decision-makers (examination writers, curriculum writers, supervisors) and the teachers.

The proposed model indicates that the washback effect of the new elective diploma takes place when the teachers perceive that examination preparation is important (Alderson & Hamp-Lyons, 1996; Green, 2007; Qi, 2005). As discussed in this chapter, teachers’ teaching behaviours or responses vary according to their beliefs and perceptions regarding the new testing design and content. The way that teachers think determines why they would teach to the test rather than teach according to the textbook design. The model also proposes an interaction between teachers’ beliefs and their practices such that the implementation of the new test and its demands and requirements may affect teachers’ beliefs and their perceptions about what constitutes good teaching practices. In addition, the study explored the possibility that the way in which washback operates and its complexity could stem from factors other than the test, such as teacher-related factors. Furthermore, the study considered processes in terms of elements that changed in the educational system, such as the elective diploma syllabus, teaching strategies and content. This study focused on the washback effect in relation to actual classroom teaching, looking both at behaviour in terms of teachers’ practices in the classroom (through the use of classroom observation) and attitudinal aspects (MoE decision-makers’ views of intended washback and changes in teachers’ attitudes).

Thus, the washback phenomenon in this study was addressed in terms of the teachers’ beliefs, teacher-related factors and classroom practices in relation to the new examination. In a sense, the influence of the new examination is represented in this study as a product of the teaching processes involved. The product (the third perspective in Hughes’ trichotomy) has a different meaning in this study. In Hughes’ (1993) model, the product was defined as “what is learned (facts, skills, etc.) or the quality of learning such as fluency, etc.” (cited in Bailey, 1996, p. 262). However, this study did not focus on the product as defined by Hughes, largely due to time constraints. Instead, it sought to

examine the overall process of the washback effect to understand the various interactions between the participants and processes within the Omani educational context. This study thus examined the influence of testing on the participants and the processes. Consequently, the changes in the objectives of the new examination, classroom teaching practices and teaching activities and materials were all treated as both process and product within the research framework. Moreover, teacher-related factors that might facilitate or impede the perceived impact on classroom teaching were also highlighted in the proposed model. These factors could affect teachers' perceptions of their capacity to implement change and could relate to the teachers themselves, for example age, gender, teaching experience and so on, or the wider teaching context, such as teacher training, workload, or the nature of the teaching situation.

To sum up, the nature and intensity of washback from the new examination was conceptualised as affecting first the teachers' perceptions of their teaching and their personal characteristics (studied through a questionnaire). These perceptions, in turn, affected their classroom teaching practices, i.e. processes, which eventually led to the final product, the teaching (studied through observation, interviews and questionnaire).

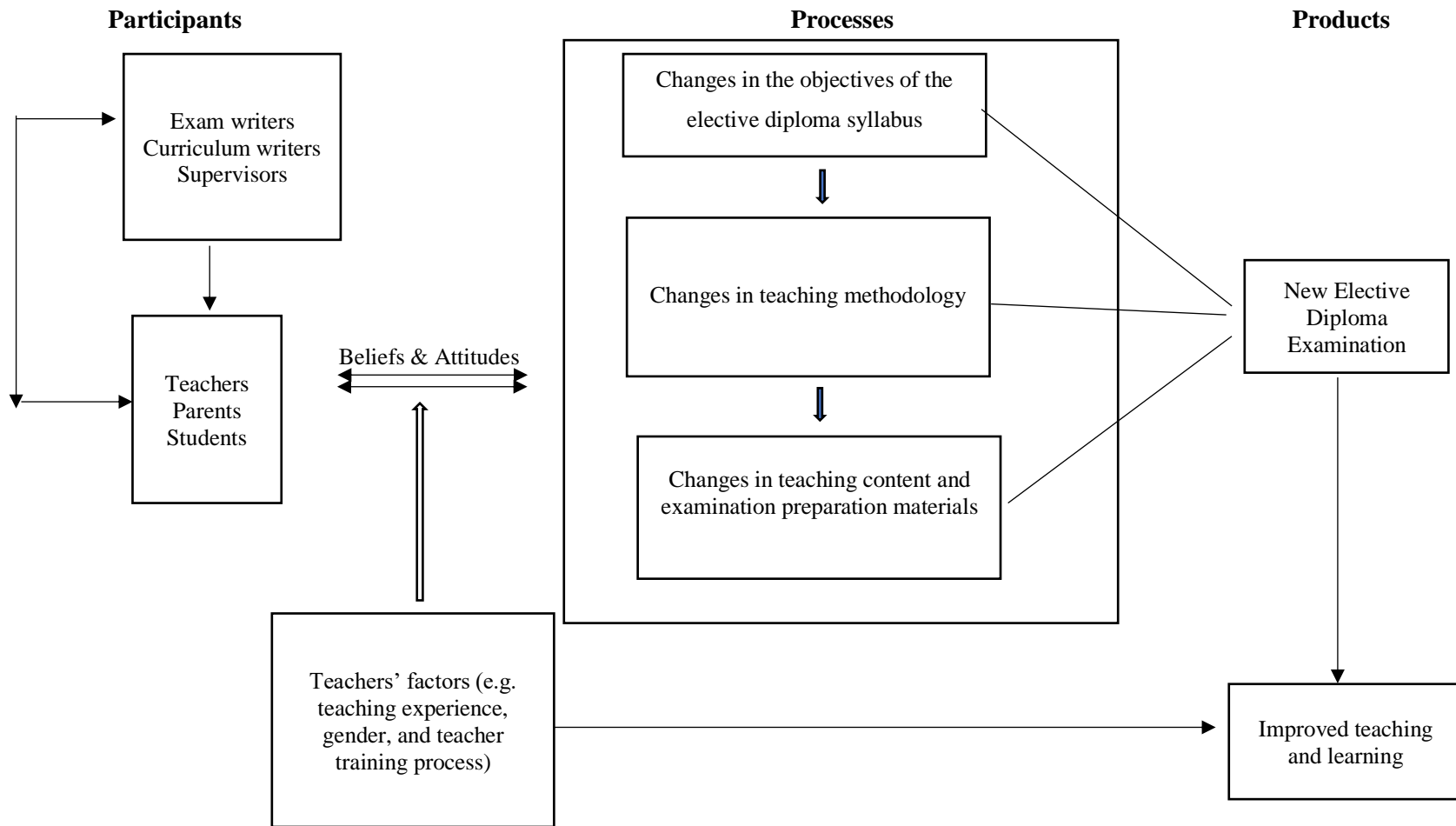


Figure 3.5. Proposed model of washback on teaching (adapted from Hughes, 1993, cited in Bailey, 1996)

3.11 Summary

This chapter has discussed definitions, models and the characteristics and nature of washback. Various relevant empirical studies have been reviewed to identify the gaps in this research area and position the focus of this research in relation to earlier studies. This research investigates test influences on teachers' beliefs and teaching behaviour, using insights from previous washback studies. It is clear that the implementation of policy employing tests in any educational system involves a variety of stakeholders, such as test constructors and curriculum writers. Such stakeholders play a key role in the development processes of a new examination in the local context. Furthermore, the review shows that numerous factors can affect teachers' perceptions of the washback from public examinations on their instructional practices. These factors were investigated to understand the particular impact they had on teachers' instructional practices in relation to the new English elective diploma examination in Oman.

The conceptual framework, which provides the theoretical framework for the investigation undertaken in this research, suggests that there will likely be no washback without engagement on the part of the participants and the nature of washback from the new examination or its products will be a consequence of those processes. It also assumes that washback, as discussed in the review of washback hypotheses and models, may not always operate in a direct way, but can instead be mediated through a variety of external factors beyond the test itself, most importantly contextual factors and teacher-related factors. Thus, the washback phenomenon may be considered in this model as uncontrollable, as its nature and intensity can be facilitated or hindered not only by the format and design of the examination, but also by other factors, which are all issues worthy of investigation for this research context.

The studies reviewed in this chapter show the use of different research methods adopted in accordance with the nature of the inquiry, the requirements of the researcher and the context under investigation. The literature review has thus also informed the empirical stage of this study in terms of the adoption of an appropriate design and methodology, as discussed in the following chapter.

Chapter 4. Research Methodology

4.1 Introduction

This study aimed to explore the washback effect of a high-stakes English language examination – the grade 12 English elective diploma examination – on teaching in Omani EFL public schools. The purpose was to explore how changes in the examination influenced classroom teaching, including aspects of teachers’ perceptions and teaching practices in the classroom. It focused in particular on whether the intended washback appeared to have been achieved in classroom practices.

This chapter discusses the theoretical paradigm and methodology drawn upon to design this study. It begins (4.2) by restating the position adopted in terms of the research paradigm. It then describes the methodology employed in previous washback studies (4.3), and the research design and methodology adopted for this study, including the specific research techniques used (4.4). The section on research design and methodology (4.5) presents the research instruments and addresses major factors considered in their design, as well as the research participants and sampling, the piloting stage and the procedures used for data collection. The chapter describes the methods of data analysis (4.6–4.9) and the researcher role in this study (4.10), before concluding with a review of ethical considerations (4.11).

4.2 Research Paradigm

A “paradigm”, according to Kuhn (1970), defines the set of views and beliefs of a particular group about what constitutes an appropriate way, or “an accepted model or pattern” (p. 23), of determining the most important questions in researchers’ field and the best ways of addressing those questions. That is, a paradigm defines the kinds of questions asked by researchers, how these questions are to be understood, what data are significant, how to investigate a phenomenon, what predictions and recommendations can be made and how results are to be interpreted (Bergman, 2010; Clark & Creswell, 2008; Morgan, 2007).

Researchers believe that inquiries and research methods should follow a systematic plan wherein the researcher articulates the paradigmatic underpinning the research, as well as a specific research design including methods for collecting, analysing and interpreting research data. There are various paradigms (Lincoln & Guba, 2003), the most widely acknowledged within social sciences research being positivism, post-positivism, constructivism, the participatory paradigm and pragmatism (Shannon-Baker, 2016; Teddlie & Tashakkori, 2009).

The nature and purpose of research can be clarified and organised by paradigmatic thinking (Cohen et al., 2011) and this research was guided by pragmatism, as were previous washback studies (see 3.4). Pragmatism is concerned less with the ultimate nature of reality and more with whether a meaningful solution can be found to a problem. In this vein, the truth is “what works” best for the specific task at hand (Johnson & Onwuegbuzie, 2004), whether the methods are quantitative, qualitative, or some combination of the two (Cohen et al., 2011). In pragmatism, what something “means” is manifested in empirical solutions to a particular problem in the real world, rather than, for example, an idealistic approach that adheres to a particular value system or epistemology; it is “practice-driven” (Denscombe, 2008, p. 280). This approach is consistent with the argument made by Maxcy (2003) and Creswell (2009) that the research questions should have primacy over the methods or the paradigmatic thinking underlying each method.

Furthermore, pragmatists believe that research into any given problem involves a process that falls somewhere along the deductive–inductive spectrum. At any given point during research, researchers may start to adopt an inductive approach (i.e. generating specific meanings from the data set about the phenomenon being studied), while others may choose a deductive approach (i.e. testing theories, hypotheses, conceptual frameworks, or causal relationships between variables, and seeking to investigate whether these concepts apply to specific instances). However, pragmatics may employ abductive reasoning, where the deductive outcomes of a quantitative method can be utilised as inputs for the inductive objectives of a qualitative approach in a sequential fashion (Teddlie & Tashakkori, 2009). This suggests alternating between qualitative and quantitative methods within a single study and making meaningful connections between them (Clark & Creswell, 2008). Therefore, pragmatism is a suitable framework for understanding a specific phenomenon.

There are four fundamental, interrelated assumptions with respect to pragmatism: ontology, epistemology and methodology (Guba & Lincoln, 1994); axiology is an additional consideration (Reason, 1998). Ontology is “the study of being, that is, the nature of existence and what constitutes reality” (Gray, 2017, p. 21). Essentially, ontological assumptions raise queries about “how things really are” and “how things really work” (Guba & Lincoln, 1994, p. 108). Pragmatism embraces features associated with two opposing ontological views: first, that reality exists independently of individual consciousness and knowledge and hence reality exists “out there”, waiting to be discovered, regardless of a researcher’s perspective or belief (Cherryholmes, 1992. P. 14); second, with regard to the real world, it holds that truth cannot objectively be determined, that the criterion for choosing one reality over

another is how well that truth produces anticipated or desired outcomes (Cherryholmes, 1992) and therefore different individuals access reality in different ways (Guba & Lincoln, 1994).

This study adopted a mixed view of reality: realism and relativism. One stance concerns evidence from previous studies that washback does exist (e.g. Cheng, 2005; Green, 2007; Shohamy et al., 1996; Wall, 1996) and that the study needs to uncover that reality as it is perceived by the research participants. The other stance is the belief that the reality of washback exists inside the human mind and may be viewed and comprehended differently depending on individuals' personal perceptions, practices and experiences. The factors that provide support for positing that teachers deal in subjective reality include the following. First, teachers translate and reinterpret messages about policy according to their own understanding of the situation. Moreover, according to Fullan (2001), teachers often work independently, so there is little chance for them to discuss matters with colleagues; instead, they are pressured to accomplish a great deal of the curriculum within a short timeframe and they are also expected to carry out changes devised by others. Unintended washback might occur in teaching if such changes are imposed upon teachers without consultation, leading naturally to resistance (Curtis, 2000). Additionally, there tend to be discrepancies between proponents of examination change, that is, between the decision-makers and teachers, who are directly affected by the decisions (Andrews, 1995). The main aim of this study, which investigated the nature and scope of the washback phenomena, thus derived from the existing reality in the social world with an objectivist orientation. However, I also recognise the important contribution of individual beliefs and thoughts to one's understanding of the realities of the outside world, namely teachers' personal interpretations and views of washback phenomena.

While ontology attempts to understand "what is", pragmatist epistemology embodies an understanding of "what constitutes knowledge". Epistemology refers to the relationship between the "knower" (the nature of knowledge that is being gathered) and the "known" (the ways in which knowledge is interpreted by the researcher and the participant) (Teddlie & Tashakkori, 2009). In terms of attaining a comprehensive understanding of the relationship between the knower and the known, within pragmatism there are two opposing views. The first is the positivist view, which posits that this relationship is objective and tangible and that researchers detach themselves from what is being observed and thus operate as observers to ensure that data are value-free. The second is the constructivist view, which holds that knowledge is subjective and personal and thus must be grounded in individuals' interactions

and experiences to co-construct social realities (Grey, 2013). Researchers holding this view interact with other participants when searching for information on a specific case or situation (Cohen et al., 2011) and put little distance between themselves and what is researched (Collis & Hussey, 2003). In this case, pragmatism once again rejects the need to choose between objectivity and subjectivity. Those with a pragmatic perspective understand that epistemological assumptions fall on a spectrum rather than being confined one of two contrasting viewpoints of subjectivity and objectivity. During certain stages of the research, the researcher and participants may need to engage in extensive collaboration to address intricate inquiries, while at other times, the researcher may not require interaction with the participants, mainly when collecting quantitative data (Teddlie & Tashakkori, 2009).

In terms of the epistemological assumptions of this research, therefore, two views were adopted: one objective and independent and the other subjective and interpretative. Both views were necessary because the objective aspects of teaching provide the means of discovering the subjective reasons why changes are necessary. Without careful grounding based on scrutinising decision-makers' intentions for and reactions to examination change, one cannot explore the subjective processes constructing participants' beliefs and attitudes, nor understand what actually happens in the classroom setting: various factors influencing the phenomenon determine the way in which research data are collected, analysed and interpreted, as well as the exploration of relationships among the reported data.

It is evident from what has been discussed thus far that these two assumptions of pragmatism – ontological and epistemological – are concerned with conceptions of “truth”: the essence of reality being investigated and the nature of truthful knowledge of reality. The third, axiological assumption concerns the nature of values and value judgments that are attributed to the inquiry process, especially ethical and moral issues (good versus bad, right versus wrong). Axiology refers to “values of being, about what human states are to be valued simply because of what they are” (Heron & Reason, 1997, p. 287). Axiology is important for this study in that it focuses on what is valuable and ethical in all stages of the research. This is especially relevant, as axiology has a direct bearing on the ethical issues that must be considered when planning qualitative research and it provides the foundation for understanding what the important additions to knowledge in any field of inquiry might be. Moreover, values are important both in conducting this research and in drawing conclusions (Cherryholmes, 1992). Based on axiological assumptions, the research topic of this study was shaped by what I considered to be essential based on my personal values and interests,

including the choice of variables I believed would most likely yield interesting findings (Teddlie & Tashakkori, 2009).

The fourth type of methodological assumption involves the “choices we make about cases to study, methods of data gathering, forms of data analysis, etc., in planning and executing a research study” (Silverman, 2005, p. 99). To the general concept of research methodology, Henn et al. (2006) added ethical considerations and consequences, the accessibility of the research field and the role of values – both the values of the researcher and of others who have influence over the field in which research is conducted. Methodological assumptions are constrained by ontological and epistemological stances. This means that, for example, researchers who view reality as being reliably external to the individual will tend to use quantitative methods, such as structured questionnaires, surveys and experiments, whereas those who believe in subjective realities will tend to favour qualitative, less generalisable methods, such as observations and interviews. Some researchers have argued that the various paradigms are incommensurable (Biesta, 2010; Cohen et al., 2011; Gray, 2013; Johnson & Onwuegbuzie, 2004). Even if it is possible to mix quantitative and qualitative methods, there is the question of whether it can ever be sensible to integrate approaches that imply seemingly incompatible worldviews and following on from this, what researchers should do to overcome such concerns.

To resolve any commensurability problems of mixed methods, inquirers have taken several stances when adopting paradigms to underpin their research, including dialectical, a-pragmatist and single paradigm (see Mertens, 2012; Teddlie & Tashakkori, 2009). Rather than assigning constructivism and positivism two different epistemological and ontological assumptions, this research focuses on a pragmatic stance that represents an intermediate point of view on the paradigmatic continuum in terms of mode of inquiry. The distinct views of the pragmatic paradigm are “logically independent and therefore can be mixed and matched, in conjunction with choices about methods, to achieve the combination most appropriate for a given inquiry problem” (Greene & Caracelli, 1997, p. 8).

By adopting a pragmatic stance, then, this study was free to utilise both types of data, quantitative and qualitative, and offered a flexible research design, so long as the methods applied to the research question; hence, one can choose to use abductive reasoning to address the multiple significances of research questions. Moreover, data sources were triangulated to gain a fuller understanding of practices and specific actions in classroom teaching. A more

in-depth justification for adopting the mixed-method approach following the washback literature is given in the following sections.

4.3 Methodology in Washback Studies

Given the complexity of the washback phenomena in the context of teaching and learning, a wide array of methods has been employed in prior research, including surveys, interviews, classroom observation, diaries, testing and document analysis (see 3.6 and 3.7). The findings from previous washback studies that have relied exclusively on quantitative data with no other forms of evidence may be limited since they are based on self-report (see Amengual-Pizarro, 2009; Andrews, 1994; Bailey, 1996; Cheng, 1998; Dammak et al., 2022; Green, 2007; Hawkey, 2006; Li, 1990; Rao & Haque, 2019; Shohamy, 1992; Stecher et al., 2004). Bailey (1996) considered purely experimental designs to be “futile and unhelpful” when used to measure washback, because it is difficult to separate washback effects from many other features of teaching and learning. Hawkey (2006) had a similar view of experimental approaches and asserted that an experimental design is rarely used in washback and impact studies because the nature of washback requires a research tool that measures interrelationships between variables and processes.

Despite the criticism of experimental approaches, other researchers have argued that experimental baseline data are useful to identify changes in classroom practice before and after the introduction of a new examination (Andrews et al., 2002; Cheng, 2005; Wall, 2005; Wall & Alderson, 1993; Wall & Horak, 2007). Weir and Roberts (1994, p. 46) noted the need to consider the conditions before a “treatment” takes place, so that there will be a point of comparison for changes that take place several years afterwards. Bray and Luxon (1999) pointed out that in research with no baseline design, “it is extremely difficult to provide convincing qualitative and quantitative evidence of change” (p. 34). Moreover, other researchers have stressed the importance of understanding the characteristics of the teaching situation before the introduction of innovation, i.e. the “antecedent conditions” (see Henrichsen, 1989; Wall, 2005). However, although baseline data may be deemed necessary to help decision-makers at the macro level shape new tests and introduce them successfully (Luxon, 2004; Wall, 2005), they cannot be used as a reference point for measuring a process of change until “the later follow-up research has been undertaken” (Wall & Horak, 2007, p. 101).

Many washback studies have mainly drawn on findings from surveys of the effects of testing (see 3.6 and 3.7). In contrast, Watanabe (2004) argued that qualitative methods were more

suitable than quantitative methods, although he further explained that this does not necessarily disqualify quantitative research methods for studying washback. Other researchers (e.g. Alderson & Hamp-Lyons, 1996; Cheng, 2005; Hayes & Read, 2004; Wall, 2005; Watanabe, 2004) have similarly suggested that observation should be used to gather data to understand the washback phenomenon. Cheng (2005) argues that further research is necessary to explore the impact of tests on teaching and learning based on both surveying teachers' perceptions and observing teaching and learning in action. As Watanabe (2004) maintains, due to the complex nature of washback, "the methodology that attempts to disentangle the complexity has inevitably to be multifarious" (p. 20). Furthermore, Wall and Alderson (1993) emphasised the benefits of mixed-method research designs to provide insights such as "why the teachers do what they do, what they understand about the underlying principles of the textbook and examination, and what they believe to be effective means of teaching and learning" (p. 62).

For example, Herman and Golan (1991) employed only a survey to investigate teachers' instructional practices in two distinct categories of schools. The findings indicated that survey data alone can be useful but are insufficient to understand what is happening in classrooms (see 3.7.2). In this regard, Thorne (2000) suggested that researchers need to access different methodological strategies because "a distinction between explaining how something operates (explanation) and why it operates in the manner that it does (interpretation) may be a more effective way to distinguish quantitative from qualitative analytic processes involved in any particular study" (p. 68). Moreover, Zhan's (2009) investigation of the influence of a new test on the learning of English revealed that the survey method restricted the participants from expressing their perceptions due to its rigid nature and structured format, whereas qualitative methods, such as semi-structured interviews and journals, were useful for collecting detailed data and gaining a comprehensive understanding of participants' "insider" perspectives. Thus, recent washback studies have used both types of methods, qualitative and quantitative, concurrently to explore in depth the complexity of the washback process in relation to instructional practices in the classroom (Allen, 2016; Cinkara & Tosun, 2017; Khodabakhshzadeh et al., 2017; Phan & Effeney, 2015; Yıldırım, 2010).

Following on from the above discussion, this thesis argues that in investigating the influence of a new test on teaching practices as perceived by teachers, attention needs to be paid to the many factors that may contribute to changes in teaching practices, as well as teacher-related and examination-related factors that seem to influence the intensity and nature of the

perceived washback effects beyond the examination itself. This thesis focuses on understanding how changes appear from the point of view of decision-makers (examination writers and curriculum developers), as well as teachers' perceptions. Therefore, this research collected both quantitative and qualitative data to explore teachers' perceptions and practices and determine whether the perceived washback effect appeared to have been achieved as intended by the MoE decision-makers. Furthermore, this thesis argues that the examination cannot in and of itself dictate the "how" and "what" of teaching practices in the classroom. Rather, the intensity of washback occurs through other intervening factors related to teachers (see 3.8). Hence, a mixed-methods approach was considered necessary for this study. Based on the above philosophical and methodological justifications, the design of this study research and the methods are discussed in the following section.

4.4 Research Design

4.4.1 Approaches to studying washback

Considering the intricate, multivariate nature of a washback study, it was necessary to identify and address a variety of factors related to examination change that may influence classroom teaching, as well as to consider factors other than the examination itself as possible influences on the intensity of the apparent washback effects in classroom teaching. As discussed above, this washback study investigated teachers' beliefs and thoughts and what they actually did in the classroom following the introduction of the new English elective examination. Accordingly, data triangulation was used in this study to shed light on the complex aspects of the research phenomenon and potentially increase the truth value of the findings in the case that the research methods and instruments yielded similar results (Denscombe, 2014).

Quantitative and qualitative methods each have their own advantages and disadvantages. The quantitative method is rooted in the positivist perspective, often seen as the classic scientific method, as it involves recognising a problem or occurrence, formulating an initial hypothesis and verifying the hypothesis by collecting and examining empirical data with exact techniques to generate reproducible results. Thus, this method offers the possibility of exploring the research questions objectively, avoiding biases or prejudices when there is more than one interpretation of the findings and controlling alternative explanations, with the possibility of generating findings that are generalisable to other contexts (Creswell, 2009). This results in what some researchers believe to be a comprehensive and reliable description of the world (Ary et al., 2010). One of the distinctive features of the quantitative approach is that it is centred on numerical data, typically analysed (and thus "explained")

using inferential statistics. These data need to be precisely defined by the content and the boundaries of the variables under study and by the different values within the variables (Dörnyei, 2007). However, while the quantitative researcher has the power to generalise findings, this generalisation may refer only to a restricted population (Bryman, 1988). For example, Freeman (1986) said of researchers in the field of organisation studies that “they rarely work with samples that are representative of even the restricted types of organizations they choose to study” (p. 300). Moreover, quantitative research tends to treat the overall tendency of responses from individuals as the centre of attention for researchers (Bryant, 1985; Creswell, 2009), a focus deriving from the assumption that the survey is administered to individuals as a discrete object of inquiry (Bryman, 1988).

Qualitative methods, on the other hand, can be defined as investigating the meanings that individuals or groups bring to a certain research problem or phenomenon (Creswell, 2009). According to Denzin and Lincoln (1994), qualitative research can be associated with multiple methods, involving both interpretive and naturalistic methods to interpret a social or human problem. This means that “qualitative researchers study things in their natural settings, attempting to make sense of, or to interpret, phenomena in terms of the meanings people bring to them” (p. 2). According to Flick (2014), qualitative approaches allow researchers to analyse the subjective meanings that guide research participants’ interpretations, to grasp latent meanings in specific settings, and to describe the social practices and experiences of participants. For Corbin and Strauss (2008), committed qualitative researchers tend to discover participants’ inner thinking and experiences, and as a result understand a world that they do not have access to. This indicates that qualitative methods are employed to capture a clearer picture of reality and to provide in-depth information about complexities in the situation under study (Chalhoub-Deville & Deville, 2008).

Despite these strengths, qualitative research has its limitations. One is due to the typically small sample size, which leads to a lack of generalisability since respondents often have idiosyncratic characteristics (Thompson, 2011). For example, in Lam’s (2015) study of pre-service teachers’ preparation for using tests to promote effective learning in Hong Kong, the case study findings should not be generalised because of the small number of participants. Thus, Duff (2006) warns that qualitative data might be helpful in describing a phenomenon, but the specific descriptions or conditions may not be applicable to other contexts. Another deficiency of qualitative research is that, as mentioned above, researchers make interpretations based on the perspectives of individuals within the generalised population of

research participants; it is unclear how those researchers can assess the validity of research interpretations based on those perspectives (Bryman, 1988). Furthermore, methods of analysis for qualitative data often take up a considerable amount of time, relative to quantitative methods (Flick, 2011; Miles, 1979; Richards & Richards, 1994); Sallee and Flood (2012) suggested that, for this reason, policymakers tend to favour quantitative methods. Berg and Lune (2012) similarly asserted that “qualitative research is a long hard road, with elusive data on one side and stringent requirements for analysis on the other” (p. 4).

Although there may seem to be clear differences between the types of methods as outlined above, there is in fact more overlap than difference. Moreover, as Pring (2004) pointed out, “The distinctions within the so-called paradigms are often as significant as the distinctions between them” (p. 48). Brannen (1992) argued that many claim that qualitative data are only concerned with the meaning individuals ascribe to a certain phenomenon, concept, or situation, whereas quantitative research is centred around the relationships amongst variables rather than individual cases – but that these claims are not fully supported, since both approaches may be concerned with individuals’ perceptions and their actions. Moreover, Ritchie and Lewis (2003) asserted that the claim that qualitative research lacks generalisability is only true if generalisability is based on statistical inference – that is, when the study sample are generalised to the whole population. However, there are other ways of understanding generalisability in qualitative research: these may involve “transferability” (Lincoln & Guba, 1985), which is achieved when a reader feels that the story of the research overlaps with their own situation and they intuitively apply or transfer the research ideas to their own situation, or “naturalistic” generalisation (Stake & Trumbull, 1982), which occurs when the research resonates with the reader’s own intuitive understanding and experience. Therefore, Brannen (2005) concludes that data produced by both quantitative and qualitative research methodologies can be mutually interdependent due to existing strengths and weaknesses shared by both types of research methodologies. In a similar vein, Sandelowski (2003) has argued that the distinction between qualitative and quantitative research is not clear, because the former means many things to many people, so there is no consistent way in which a comparison can be made.

Based on this, it can be inferred that using only quantitative or qualitative methods would not be sufficient to achieve a complete understanding in addressing each research question in this study. The purpose was to provide a thorough overview that required a combination of methods.

4.4.2 Mixed-methods design

To overcome the weaknesses and biases of a sole method, and based on the research problem and its objectives, a mixed-methods design involving quantitative and qualitative methods was adopted. Johnson et al. (2007) presented several definitions of mixed methods which differ in terms of what is mixed, when or where in the design the mixing occurs, the scope of mixed research, the motivations for mixing and the orientation of the mixed-methods research. Other researchers have suggested that mixed methods involve data collection, analysis and interpretation of the quantitative and qualitative methods that, singly or together, offer a more balanced and complete understanding of a particular research phenomenon (Leech & Onwuegbuzie, 2009, p. 265). Mixed-methods research can be conducted in all research phases and aspects: paradigmatic foundations, research design and questions, data collection, analysis and interpretation, as well as recommendations and uses of research results (Tashakkori & Creswell, 2007). As Yin (2006) notes, “the stronger the mixing between the methods at all stages, the stronger the results of the mixed-methods research” (p. 46).

Naturally, a mixed-methods study design combines different aspects of qualitative and quantitative methods. The key challenge in employing this design appears when researchers attempt to articulate how the two distinct sets of characteristics of each method relate to each other (Tashakkori & Creswell, 2007). Moreover, doubts have been raised about what mixed-methods research comprises, how it can be organised and conceptualised, which elements of the research need to be explained, and how and when they can be explained (Bryman, 2007; Tashakkori & Creswell, 2007). According to Teddlie and Tashakkori (2009), the choice of design depends on four criteria: “the number of methodological approaches used, the number of strands or phases, the type of implementation process, and the stage of integration of approaches” (p. 140).

This study employed a sequential mixed-methods design so that the first phase, drawing on quantitative data, determined the subsequent phase, drawing on qualitative data and synthesising the main findings from both data sets (Teddlie & Tashakkori, 2009). In line with this design, this study explored aspects of classroom teaching related to teachers’ perceptions of the new examination, first with a large sample and then in more depth with a small sample during the qualitative phase. The study thus gave priority to the qualitative phase despite the quantitative phase in this study taking place first. This decision was made with a view to the data from the qualitative phase building and expanding on the data obtained from the first quantitative phase and thus enhancing understanding of the washback

effects of the new test. The qualitative data could help explore relationships between what the teachers perceived about the examination change, what they actually did in the classroom and whether the new examination affected teaching in the ways intended by the MoE decision-makers.

Specifically, the research instruments developed for the study were a teacher questionnaire, a classroom observation scheme and an interview guide. The main purpose of the quantitative questionnaire was twofold: first, to explore aspects of classroom teaching related to teachers' attitudes and reactions as influenced by the introduction of the examination (RQ1); second, to investigate teacher-related factors that might influence the perceived washback of the new elective diploma exam on teachers' practices in the classroom (RQ4). In the qualitative phase, classroom observation was used, applying the responses given in the teachers' questionnaire to explore further aspects of the apparent test washback on teachers' classroom teaching, investigating the relationships between the perceived examination change and teachers' actual practices in the classroom (RQ3). While classroom observation is best suited to richer exploration of a phenomenon, which can inform the research design (Creswell & Plano Clark, 2011), Wall and Alderson (1993) point out that observations alone can only describe one aspect of what is happening within classroom teaching and do not provide justifications for teachers' practices in the classroom or give a clear picture of what teachers assume effective teaching and learning to be. Therefore, interviews were also employed to elicit participants' views concerning whether the intended washback of the new examination appeared to have been achieved (RQ1 and RQ2).

Hence, the four instruments were designed to complement each other. This type of mixed-methods approach applied to a single phenomenon offers cross-checking mechanisms in examining washback, often referred to as "triangulation" (Denzin, 1978). Hammersley and Atkinson (2007) noted that one form of data collection may lead to undetected errors in the interpretation of research findings, whereas when "different types of data lead to the same conclusion, we can be a little more confident in that conclusion" (p. 183).

In this study, two types of triangulation were employed (Allwright & Bailey, 1991; Denzin, 2009; Dörnyei, 2007). The first was *data* triangulation, which means bringing to bear information from a variety of sources as a way of answering the research questions (e.g. data from teachers, examination writers, curriculum developers and supervisors). The second was *methodological* triangulation, which refers to using more than one method for eliciting data (e.g. surveys, classroom observations, and interviews). As a check on validity, this study

employed “between-method” triangulation by combining classroom observation and the survey as methods (Denzin, 1978, pp. 301–302).

4.5 Research Phases

The data for this study were collected in three phases: a pilot phase and two phases of data collection. These phases are described below; details regarding the design of each research technique can be found in 4.7.

4.5.1 Phase one

Phase one involved the piloting of the questionnaire, observation and interviews. Van Teijlingen and Hundley (2000) defined pilot studies as small-scale studies designed to try out or pre-test a particular research instrument in preparation for a major research study. For Marshall and Rossman (2010), piloting studies provide researchers with opportunities to eliminate barriers and allow for adjustments and revisions in the actual study. In this vein, the piloting of data collection in this study was developed with the overall purpose of identifying any issues that might arise in the main study (Kim, 2011; Oppenheim, 1992) and in the specific research context (Williams et al., 2008). To ensure the validity and reliability of the research (Cohen et al., 2011), the pilot study evaluated the teacher questionnaire, the classroom observation scheme and interview guides for the decision-makers and teachers that had been developed for use in the main study (phases two and three). This pilot phase lasted from mid-December 2020 to the end of March 2021. A detailed description of the piloting stage for the three instruments is given in the discussion of each research technique.

4.5.2 Phase two

Phase two was primarily concerned with teachers’ perceptions, using the teacher questionnaire as the research instrument. The aim of this phase was to examine teachers’ perceptions of the new English elective diploma examination and how they perceived themselves to be affected by the introduction of the examination. Hence, the findings of this phase addressed the first research question: “What are teachers’ perceptions of the new English elective diploma examination?” The data in this phase related to teachers’ attitudes concerning their teaching practices in the context of examination change. Moreover, this phase focused on the extent to which the teachers’ personal characteristics might have influenced the intensity of the apparent washback effects from the new examination (RQ4). In phase two, data were collected throughout February and March 2021. A detailed description of the implementation of the questionnaire during this phase is given in 4.7.1.

4.5.3 Phase three

Phase three was concerned with the impact of the new examination on teaching in the classroom, focusing on what actually happened in the classroom following its introduction. As well as classroom observation, in-depth interviews with teachers and MoE decision-makers were also conducted. Hence, this phase was particularly focused on the second and third research questions: “What is the intended washback effect of the new English elective diploma examination according to MoE decision-makers?” and “What is the nature and scope of the apparent washback effect of the new English elective diploma examination on teachers’ classroom practices?” This phase took place throughout April and July 2021. A detailed description of the implementation of the instruments is provided in the discussion of each research technique (4.7). Figure 4.1 illustrates the design of the three phases in this study.

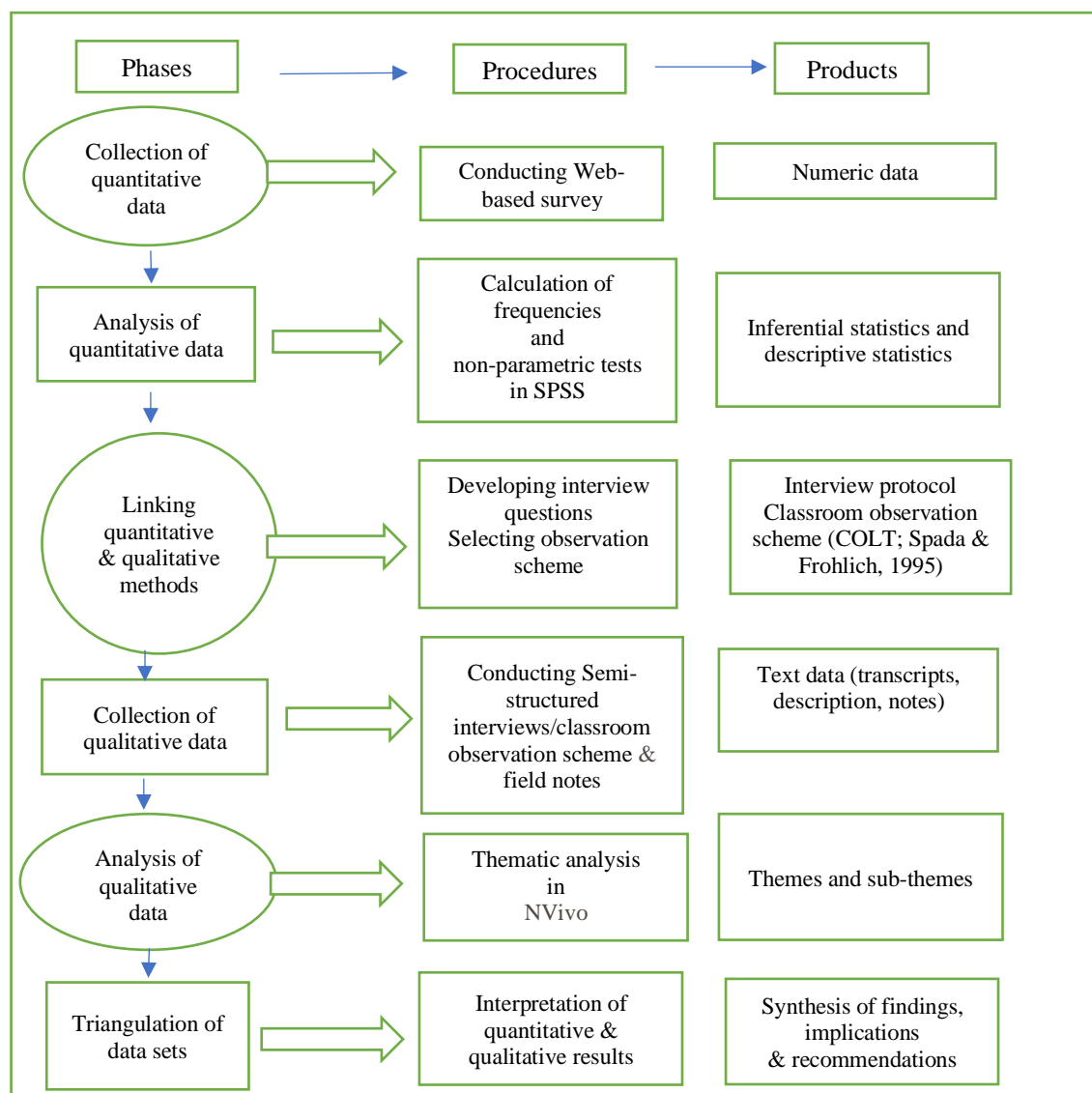


Figure 4.1. Proposed sequential explanatory research design

4.6 Research Participants and Sampling Techniques

Although the target population comprised mainly teachers, as they would play an essential role in shaping whether and how washback operates and to what extent the examination ran counter to existing teaching practices (see Burrows, 2004; Kim & Isaacs, 2018; Spratt, 2005; Woods, 1996; Yin, 2005), the perceptions of examination writers and curriculum developers were also examined to understand the different dimensions of both intended and unintended washback. Decision-makers were targeted as participants for this study because they ultimately set goals and initiate action in English language teaching. As emphasised by Watanabe (2004) and by Cheng and Curtis (2004), the importance of a test depends on the perceptions of the key test stakeholders involved in the development process of the test and the different distinctive approaches used by tests' participants within a particular educational context. In addition, previous studies have found that there tend to be discrepancies between the purported intention of such a change in the educational system and the perceptions of teachers who are expected to implement this change (Andrews, 1995; Smith et al., 1994). Therefore, exploring teachers' and decision-makers perceptions in this study was crucial to provide insights into how and in what areas a change in examination design and requirements might influence teaching practices and why the intended consequences were manifest or not.

The distinction between *sample* and *population* has frequently been discussed in the literature, with the former referring to “the segment of the population that is selected for the investigation” and the latter representing “the universe of units from which the sample is to be selected” (Bryman, 2012, p. 187). A primary aim in determining the sample size in quantitative research concerns the extent to which findings can be generalised to the whole population to ensure validity and avoid sampling errors or biases (Cohen et al., 2011). For this study, the decision was made to apply the questionnaire nationwide to understand teachers' perceptions broadly and potentially improve response rates (see Appendix C).

A frequently stated sampling issue is whether to opt for a probability (random) sample or a non-probability (purposive) sample. In a probability sample, each participant in the research context has the same chance of being selected, which is useful for making generalisations as it provides random representatives drawn from the wider population. In a non-probability sample, the members do not need to be a random representative of the wider population; this approach seeks to increase the depth of data at the cost of less breadth to the findings of the study (Cohen et al., 2011). As this study adopted a mixed-methods approach, the sampling techniques used were a random sampling technique in phase two (survey) and a purposive sampling technique in phase three (classroom observations and interviews). I deployed a

purposive sampling technique to select “individuals who can provide rich and varied insights into the phenomenon under investigation so as to maximize what we can learn” (Dörnyei, 2007, p. 126). Additionally, the teachers and the context were purposefully selected for interviews and classroom observations, as opposed to random selection, to achieve the key principle of qualitative sampling “[working] with small samples of people, nested in their context, and studied in-depth” (Miles & Huberman, 1994, p. 27). I chose this from among other sampling techniques as the most effective way of overcoming certain challenges, such as time and expense, that could have prevented access to the research sample, specific locations and those people who were willing and available to take part in the study.

Therefore, the research population and sample varied in terms of purpose and size depending on the research instruments being used (questionnaire, classroom observation, interviews) and the research questions being addressed in phases one, two and three. Following is a brief description of the participants in the questionnaire, interviews and classroom observations.

4.6.1 Description of the teachers

For the questionnaire in this study, the research population comprised teachers who had taught the existing elective diploma subject since its implementation in 2018/2019. Those not teaching at the time of data collection (not available to be observed) or teaching the subject for the first time in 2020/2021 (with less than a full year’s experience teaching for the examination) were excluded. This was because some of the items in the questionnaire concerned teachers’ opinions of the examination quality and format, which these teachers would not be able to comment on due to the change in the examination system caused by COVID-19 school restrictions, which meant that examinations were only to be implemented once at the end of the school year rather than in each semester. Table 4.1 provides a summary of the demographic characteristics of the teachers who responded to the online questionnaire.

Table 4.1. Demographic characteristics of teachers in the survey sample

Variable	Category	Frequency	Percentage
Gender	Female	117	60.6%
	Male	76	39.4%
Age	20–30	14	7.3%
	31–40	117	60.6%
	41–50	51	26.4%
	> 50	11	5.7%
	Bachelor’s	163	84.5%
Academic qualification	Master’s	29	15.0%
	PhD or equivalent	1	0.5%
Years of teaching experience	0–2	4	2.1%
	3–6	12	6.2%
	7–10	37	19.2%
	11–15	68	35.2%
	> 15	72	37.3%
	0–2	123	63.7%
Years of teaching grade 12 English elective	3–6	39	20.2%
	7–10	21	10.9%
	11–14	10	5.2%
	Yes	70	36.3%
Currently teaching grade 12 English elective	No	123	63.7%
	4	61	89.7%
Current teaching hours	8	6	8.8%
	12	1	1.5%
	Yes	59	30.6%
English elective diploma examination training	No	134	69.4%

Most of the teacher participants were female (60% vs 40% male). They were mostly aged between 31 and 40 years (60.6%). The second largest age group in this research (41–50 years) represented almost a quarter of the entire population at around 26.4%, and the other age groups constituted the lowest number of participants: 7.3% were in the 20–30 range and 5.7% were over 50 years old. The majority of teachers held a Bachelor’s degree (84.5%); very few had postgraduate qualifications: 15% held a Master’s degree and only one person (0.5%) held a PhD. The teachers with over 15 years of teaching experience (37.3%) and 11–15 years’ experience (35.2%) together constituted over half the population. A small group of around 19.2% of teachers had taught English for between seven and ten years, 6.2% had taught it for between three and six years, and only 2.1% had taught it for two years or less.

Most participants had taught the elective for two years or less (63.7%) and about 20% of teachers had taught the elective for three to six years; very few teachers had taught the elective for more than six years. Most of the teachers were not teaching the English elective during the time of data collection, except for around (70%) who taught the grade 12 elective

in the year of data collection (2021/2022). In terms of teaching load, most teachers (around 89.7%) only taught 4 hours per week, while six teachers were teaching 8 hours per week and one teacher 12 hours per week. It should also be noted that only about one third (30.6%) were subject-trained.

As stated above, non-probability purposive sampling was used in this study for the interviews and observations. Five teachers were selected purposefully from the questionnaire respondents to participate in the interviews and another three teachers participated in classroom observations. All teachers in this study were female. Aiming to have balanced representation, I contacted male teachers, but none of those approached at the time of the study were willing to participate. They gave various reasons for not participating in either the interviews or classroom observations, such as “suspecting Covid cases in their schools”, “heavy teaching workload” and “difficulty in arranging the observation schedule due to Covid restrictions”. To protect the participants’ anonymity and confidentiality, they were referred to as T1 through T5 in the interviews (Table 4.2), and Teacher A through Teacher C in the classroom observations. Similarly, the MoE decision-makers were referred to as MoE-1 through MoE-4 (see 4.6.2).

With regard to the interviews, the five teachers provided in-depth information concerning their beliefs and attitudes towards their practices inside the classroom. They were all Omanis with varying experience in teaching English. All the teachers interviewed had studied English language-related subjects at university: three teachers (T1, T4, T5) held a Master’s degree and two (T2 & T3) had a Bachelor’s degree. Four teachers (T1, T2, T4, T5) had taught English for period ranging from 12 years (T1, T2, T5) up to 19 years (T4). T3 had the least teaching experience at five years. For this study, it was required that participants at least two to three years of experience teaching the new elective diploma subject. Two teachers (T2, T4) had taught the new elective diploma for three years and the other three (T1, T3, T5) for two years. The participating teachers were selected for this study based on their responses in the questionnaire. They were also selected because they had taught both the old and the new elective diploma subject, except for T3, who had only taught the new elective (Table 4.2).

Table 4.2. Descriptive information for interviewed teachers

	Academic qualification	Years teaching English	Years teaching the new elective diploma	Years teaching the old elective diploma
T1	MA in English curriculum	12	2	10
T2	BA in teaching English language	13	3	9
T3	BA in English literature	5	2	5
T4	MA in TESOL	19	3	19
T5	MA in Philosophy	17	2	15

Three teachers participated in the classroom observations. Teacher A had been teaching English at school A for approximately 10 years and had taught the old English elective for 4 years and the new elective for 3. She was the only teacher who taught the new elective in her school. Teacher B had been teaching the new elective for two years. She had taught grades 11 and 12 for approximately 15 years. Teachers A and B were both doing their BA studies in language teaching at the time of the study. Teacher C had a BA in Education. She was a senior teacher with long years of experience in teaching English at post-basic school level, grades 11 and 12. She had approximately 17 years of teaching experience and had taught the new elective for around 2 years.

4.6.2 MoE decision-makers

The other group of participants in this study comprised the MoE decision-makers. They participated in the interviews (phase three) and were chosen for their experience as agents of the new policy change through a self-selection sampling method. The four participants were the only ones who had participated in the development and implementation of the new elective diploma curriculum at the Ministry level. MoE-1 was working at the Centre of Measurement and Evaluation as an exam writer. She had a total 23 years of work experience and held a Master's degree in English language curriculum and teaching methods. Some of her responsibilities at work were preparing English language assessment documentation for all grade levels, writing diploma examinations and participating in committees relating to curriculum, supervision and English language training. MoE-2 worked in the Supervision Department as a supervisor. He held a Master's degree in teaching methods and had completed 16 years of work experience. His main responsibilities in supervision were collecting and analysing feedback from all governates about any changes in curriculum, training and assessment of English language. He had joint tasks and responsibilities with curriculum, assessment and training in English language. MoE-3 represented the Curriculum Directorate as a curriculum developer/writer. He held a Master's degree in curriculum and teaching methods. He had completed 12 years in his position. MoE-4 also represented the

Curriculum Directorate as a curriculum developer/writer. He had a Master's degree and was working on his PhD research. He had completed a total of 25 years of work experience. A detailed description of the research population and research sampling techniques are given in the discussions of the design of each research technique.

4.7 Research Instruments

This section introduces the quantitative and qualitative data-gathering techniques used in this study and articulates how these techniques align with the research paradigm and the tools used in previous washback research (see 4.2 and 4.3). It also provides details of the development of the research techniques and how each functioned in relation to the mixed-methods approach.

The empirical work sought to investigate the impact of the examination on teaching by asking teachers about their views and perceptions of the test and by observing classroom teaching. The investigation was concerned with teachers as they are deemed to be the principal agents of change within this particular educational context (Fullen & Stiegelbauer, 1991). Furthermore, the intentions and perspectives of decision-makers were also examined to explain how the assessment change operated within the Omani context and whether the intended washback of the test was well understood by teachers. The study also investigated factors other than the examination itself that could affect the intensity of washback on teaching. The designs of the research techniques used here are based on the theoretical and methodological derivations from previous research (as discussed in chapter three), which can be used to probe in some depth the general characteristics of a system of testing and thus determine the consequences of a change made to a high-stakes examination in the educational system in Oman. What follows are detailed accounts of the questionnaire, interview protocols and classroom observation instrument used in this study.

4.7.1 Questionnaire

Questionnaires are one of the main data-gathering instruments used in the social sciences. They are variously referred to as “inventories”, “forms”, “opinionnaires”, “tests”, “batteries”, “checklists”, “surveys”, and “indexes/indicators” (Aiken, 1997). Moreover, multiple terms may be used to describe similar characteristics, such as “self-administered” vs “self-completion” (Bryman, 2012).

According to Brown (1997), “Questionnaires are any written instrument that presents participants with a series of questions or statements to which they should react either by selecting from existing possibilities or writing out their answers” (p. 111). Dörnyei (2007)

further notes that questionnaires measure data about the participants which can be factual, behavioural and attitudinal. Factual questions are likely to be used to collect data about who the respondents are, such as demographic characteristics (level of education, material and socio-economic status, residential location, religion, occupation and so on). Questions of a behavioural nature are typically employed to determine the actions of individuals presently or in the past, concentrating on habits, undertakings, ways of life and individual histories. Attitudinal questions are typically concerned with what people think or believe, and thus investigate respondents' interests, opinions, norms and values.

Proponents of questionnaires (e.g. Bryman, 2012; Robson, 2002; Wilson & Sapsford, 2006) point to several advantages associated with this instrument. First, questionnaires can be easily employed in field settings such as classrooms and they gather numerical data, rather than discursive data, such as transcripts, field notes, or documents (Nunan, 1992). According to Dörnyei (2007), questionnaires are more convenient than other methods in terms of duration of time needed to collect data and the efforts and financial resources spent on their design. As noted by Mackey and Gass (2005), conducting research with questionnaires is particularly efficient at providing large amounts of data compared to interviews and questionnaires can gather standardised information from almost any human population. Other reported benefits are that questionnaires are relatively straightforward to code and analyse and they reduce certain forms of bias because the same questions are used for all participants (Cohen et al., 2018). Finally, questionnaires have the potential advantage of being administered without any interference from the researcher on the respondent's own view or even the need for the researcher's presence (Gillham, 2007).

However, there are also some key disadvantages that must be considered before employing such a method. One disadvantage is that respondents may not treat the questions seriously, especially if completing the questionnaire without the researcher present and researchers may not be able to probe responses. Additionally, as questionnaires often consist primarily of close-ended questions, many do not allow respondents to justify or elaborate on their responses, which can result in superficial data (Mackey & Gass, 2005). Furthermore, prestige bias has implications for research using questionnaires: respondents may not necessarily report their true beliefs and attitudes, wishing to please or impress the researcher by providing favourable responses; thus, there may be a weak relation between the information provided by the respondents and their actual behaviours and opinions (Robson, 2002). Another type of bias prevalent in questionnaire data collection is acquiescence: people tend to agree with statements about which they feel unsure (Robson, 2002).

Among the several advantages mentioned above, the literature shows that questionnaires have increasingly and widely been used in washback research (see 3.6 and 3.7). In most washback studies, there has been a tendency to explore participants' reported beliefs and attitudes through questionnaire responses (Cheng, 2005; Green, 2007; Pan, 2014; Xie & Andrews, 2012). This tendency was also noted by Alderson and Wall (1993), who suggested that the questionnaire, possibly used in conjunction with interviews, is a useful method for learning about perceptions of events. Hence, for this study, a teachers' questionnaire was used to elicit perceptions, attitudes and behavioural changes in relation to the new examination. The questions in the questionnaire specifically aimed to investigate the nature and scope of the washback effect on teachers' perceptions regarding their classroom teaching, as stated and operationalised in the research questions. According to Denscombe (2003), questionnaires require respondents to "reveal information about feelings, to express values, to weigh up alternatives etc., in a way that calls for a judgment about things rather than the mere reporting of facts" (p. 146). In addition to this, the questionnaire was used to learn how teachers' characteristics (their qualifications, teaching experience, training opportunities and gender) might contribute to the intensity and nature of potential washback.

In line with the pragmatist approach, the questionnaire was deemed a practical and effective way of learning the perceptions of the research participants regarding the impact of the new examination on teaching practices using deductive reasoning, numerical quantification and statistical procedures. This tool allowed me to test the washback hypotheses regarding aspects of teaching in a series of logically related steps and answered the research questions in an "objective" manner, while trying to reduce the influence of my own bias, thereby resulting in an accurate and reliable description of the washback phenomena.

Questionnaire design

After reviewing the literature on research methodology related to item writing for questionnaires, I considered several technical guidelines when devising the items in an attempt to exploit the advantages of the method. In addition, it was noted that many survey researchers call for a substantial investment of time and effort in the development of questions, making the items as reliable as possible (meaning that questions mean the same thing to different respondents), and making sure that items motivate respondents and make them interested in the topic (Cohen et al., 2018; Dörnyei, 2007; Robson, 2002). Other issues were considered during the piloting process and are discussed in the following section.

Before embarking on the design of the research questionnaire items, other general issues taken into consideration were as follows:

- the type of information to be collected from the respondent to address the questionnaire objectives and research questions;
- the desired demographic characteristics of the target respondents;
- the method(s) of reaching the target respondents;
- the content of each question;
- the research context;
- the quality of questions (wording, order and format, length, clarity, etc.).

Having identified the above elements, this research adapted some items from a questionnaire used by Cheng (2005) with teachers in her Hong Kong washback study (see Chapter 3). There were several reasons for adapting this questionnaire: first, it has successfully been used in empirical washback research to examine the attitudinal and behavioural responses of teachers in relation to new examinations (e.g. Hsu, 2009; Mahmud, 2018; Onaiba, 2013); secondly, the reformed examination in Cheng's (2005) research was relevant in relation to the purpose, type and importance of that investigated in this study (both are high-stakes examinations used to provide grade 12 leaving certificates at the national level); third, the validation procedures (namely qualitative input and piloting procedures) used in Cheng's questionnaire design saved time and resources in formulating the items for this thesis, helping to conceptualise the aim of this study and making items as clear as possible for each questionnaire respondent (Cohen et al., 2018; Dörnyei, 2007).

Although Cheng (2005) extensively analysed her questionnaire items for validity and reliability, it was not appropriate to use the questionnaire in its entirety in this study for the following reasons. First, there were contextual differences that could have confused the teachers who participated in this research. For example, the term "target-oriented curriculum principles" (Part 2, Q2) would not have meaning in the Omani context and such differences may significantly affect the validity and reliability of a questionnaire (Creswell, 2009, p. 150). The second reason was that there were some questions that were not relevant to the purpose of this research. For example, the items in Part 3 focused on aspects of learning and teaching materials that were not part of this study. Furthermore, some questions in the questionnaire included long lists of alternative answers, which was impractical in the research context and could have unduly influenced participants' response rates (Burchell & Marsh, 1992). More importantly, there were some redundant questions in the questionnaire,

most of which could be answered during the interviews and classroom observations (Part 2, Q11 and Part 3, Q2, Q7, Q8, and Q9).

After reviewing related empirical washback studies (see Chapter 3) and selecting the most salient items from Cheng (2005), the adapted questionnaire included both close- and open-ended items tailored to English language teachers in the research context (see Appendix D). One of the main advantages of writing close-ended questionnaire items is that they can be numerically coded and analysed statistically more easily, whereas open-ended items can lead to superficial information and may also discourage respondents (Dörnyei, 2007; Robson, 2002). However, other survey researchers hold the view that open-ended questions are useful if close-ended questions would require long lists of possible answers (Cohen et al., 2018), if the possible answers for the questions are unknown, or if the questionnaire is used for an exploratory function (Bailey, 1994).

In this study, the questionnaire included four parts, each comprising close-ended questions with the addition of some open-ended questions. Part 1 consisted of seven closed questions concerning teachers' demographics (gender, academic qualifications, age and teaching experience), current teaching situation (grades currently taught, class size and number of lessons per week), and training courses they had taken in relation to the newly introduced examination in 2018–2019. Part 2 used items from Cheng's (2005) questionnaire adapted and modified for this study. The minor modifications included reducing the number of possible responses, changing the terms used in certain questions and removing irrelevant questions. This part of the questionnaire used a five-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). This scale was suitable for this study as it is commonly used in the field of education to gather data on attitudes and opinions (Cohen et al., 2018; Dörnyei, 2007). The items in the Part 3 of the questionnaire focused on the teachers' reactions to the new examination; these consisted of four multiple-choice questions written specifically for this study. Another multiple-choice question was included in Part 4, containing a series of statements that were designed to be straightforward and easy to complete.

The open-ended items aimed to obtain qualitative input to provide greater understanding of teachers' feelings about the impact of the new English elective examination. The first open-ended question (Part 4, Q1) was written to expand on the responses to the close-ended questions related to teachers' practices in the classroom, while Q2 concerned the teachers' perceptions of the impact of the new examination, not covered by the close-ended questions.

In addition to these two open-ended questions, other option statements were added to the close-ended questions to allow participants the opportunity to add relevant information that may not have been considered in the original development.

Piloting the questionnaire

The value of piloting a questionnaire represents a paradox for researchers. Some researchers have argued that not a great deal is known about the extent to which piloting serves its intended purposes or what researchers think it should do (e.g. Presser et al., 2004). However, many researchers (Bryman, 2012; Cohen et al., 2018; Dörnyei, 2007; Jaeger, 1988; Robson, 2002) have pointed out that piloting (also referred to as “pre-testing” or “field testing”) performs the important role of ensuring that the questionnaire items are understood and are consistently interpreted by the targeted respondents as much as possible (Cohen et al., 2018; Jaeger, 1988), hence increasing the validity and practicality (Dillman et al., 2014). Piloting is of particular value given Boynton’s (2004) point that questionnaires fail “because participants don’t understand them, can’t complete them, get bored or offended by them, or dislike how they look” (p. 1372).

Piloting of the teacher questionnaire was conducted in mid-December, before the end of semester one of the academic year. The questionnaire was distributed online to gain an understanding of the technical aspect of the online survey software (Qualtrics) and mitigate problems, as well as to ensure that the format of the questionnaire was easy to follow and it was complete. The pilot study also checked the clarity and coherence of the language used in instructions and items and established the construct and content validity of the items, enabling elimination of items that presented as inappropriate, unclear, ambiguous, or irrelevant. In this process, teachers were asked to answer the open-ended questions either in Arabic or English as they wished, as these types of questionnaire items are demanding in terms of respondents’ time (Cohen et al., 2018). Thus, piloting aimed to test whether the overall length and composition of the questionnaire would affect response rates and/or the quality of the data collected.

The following procedures were followed in piloting:

- I started by completing the questionnaire from the imagined point of view of a teacher.
- Two supervisors from the research context volunteered to comment on the questionnaire.

- A draft was tested with a random sample of 10 teachers, 2 exam writers and 2 curriculum developers to check if the questions were interpreted as expected.
- Two language education PhD researchers were asked to participate in the piloting and contribute to the final draft for the main study.
- I conducted a final review of the questionnaire, including corrections, additions and omissions.

Most participants confirmed that all the items were clear, appropriate and relevant. The time estimated to complete the questionnaire was 20–30 minutes. Similarly, the Qualtrics survey showed that the questionnaire took an average of 20–25 minutes to complete. However, three respondents expressed doubts about the pilot version for the following reasons. First, some terms and statements were slightly ambiguous and caused confusion due to the wording. For example, in Part 2, Item A, participants noted that “tertiary education” is not commonly used by teachers, rather “higher education” was clearer and more appropriate. Accordingly, all ambiguous terms and items identified were revised and rephrased. Second, the teachers were confused about how to interpret Q4 in Part 3, as they were not sure which decision-makers were responsible for teaching arrangements and the medium of instruction changes to the English elective diploma examination in 2018–2019. Another concern expressed was that some items contained a long list of alternatives, which might distract respondents from selecting the option that was most appropriate (e.g. Part 4, Q3). The participants also reported that the alternatives for Q3 in Part 3 were irrelevant and unclear and so they were replaced with other options. Other minor comments were related to the questionnaire structure within the Qualtrics survey. For example, they reported that the structure of the questionnaire did not allow them to go forward or backward throughout the survey, so this function was added for the respondents in the main study. A modified version of the teacher questionnaire was used for the main study (see Appendix D).

Questionnaire administration

It is often argued that questionnaire administration is merely a technical procedure that can be relegated to research assistants, whereas in fact administration should be regarded as an essential stage in determining the quality of questionnaire responses (Dörnyei, 2007). Based on the data from the pilot study, the questionnaire was administered electronically using the Qualtrics survey software. The online format was chosen to help distribute the questionnaire to teachers in all districts in Oman within a short period of time, in particular due to its ease of use on different mobile devices such as tablets and smart phones. Moreover, this mode offered a secure and simple way to organise, save and analyse the data, as well as to export

and import data to the Statistical Package for Social Sciences (SPSS) version 25 for further analysis (see Chapter 5). Another reason why I opted for this mode of data collection was that the process took place at a time of unprecedented change and disruption as a result of COVID-19; thus, it was advisable to use a “socially distanced” method to eliminate ethical concerns, both for the researcher and participants, which was especially important given that the target participants of this questionnaire comprised an entire research population.

Following written correspondence, an anonymous Qualtrics link was distributed through the MoE portal, which is accessible to all teachers working in the sector. The portal made it feasible to approach a large number of teachers, including those working in distant areas. What contributed to the success of questionnaire data collection was that the teachers’ experiences of online teaching during the time of COVID-19 increased their overall efficiency and confidence in dealing with online tools; indeed, most teachers now feel more comfortable about actively participating in data collection.

Moreover, the questionnaire was distributed in February 2021, when teachers in Oman were back in schools for the second semester of the school year. This was an appropriate time to ask teachers to participate as they were less busy since their marks for students’ performance in the classroom assessments, short tests and quizzes were not yet due. Thus, the timing of the questionnaire distribution enhanced teachers’ participation and ensured a good response rate, as highlighted by Cohen et al. (2018, p. 502). The research participants’ willingness to take part in data collection was reflected in the number of complete responses to the questionnaire. Nevertheless, the process of distributing the questionnaire to the entire research population was very demanding as many teachers needed to be reminded three times to complete it, once when distributing the questionnaire link, once before collecting it and once during the collection period. Initial emails, follow-up emails and follow-up phone calls were required, in a process that amounted to approximately six weeks in total.

Specifically, the questionnaire launched on 7 February and was taken down at the end of March 2021. A total of 209 questionnaires were distributed to 209 teachers covering the entire population of teachers who were teaching the English elective during the period 2018–2021. Of the returned questionnaires, 16 were not counted in the analysis as the preliminary data in Qualtrics software showed that they were incomplete or blank and others did not consent to participate in the study. Altogether, this meant 193 questionnaires were received with all types of questions completed in full. The questionnaire data represented the completion of phase two of this study, which provided data on the perceptions of teachers.

4.7.2 Classroom observation

Classroom observation is a systematic approach that allows researchers to make common-sense judgments about people, behaviours, events and settings (Simpson & Tuson, 2003). Similarly, Angrosino (2007) states, ‘Observation is the act of noting a phenomenon, often with instruments, and recording it for scientific purposes’ (p. 2). Observation is also regarded as a useful research tool in providing “thick descriptions of the target culture” (Dörnyei, 2007, p. 130). According to Cohen et al. (2000), it is one of the research methods used to “understand the context programmes, to be open-ended and inductive, to see things that might otherwise unconsciously be missed, [and] to discover things that participants might not freely talk about in interview situations” (p. 305). Observation has become a more prevalent method used in researching washback on language teaching and learning (see Alderson & Hamp-Lyons, 1996; Burrows, 2004; Ren, 2011; Wall, 2005; Wall & Alderson, 1993). Such washback studies have used classroom observation to measure what happens in the classroom and to understand interactions between teachers and their students, as well as to identify the teaching and learning processes in a research context. Other washback studies have found that classroom observation helps to understand the washback effect of new tests, which have varying impacts on teaching practices depending on differences among teachers (Blewchamp, 1994; Burrows, 2004; Watanabe, 1996). Importantly, observation may go some way to addressing inconsistencies and gaps in data gathered using interviews or self-report instruments. This was also suggested by Ren (2011), who stated that classroom observation helps to provide a clearer picture of the multi-directionality of the washback process than simply relying on self-report data. Another important use of observation in the washback literature has been to identify variables that may contribute to the impact on teaching and learning as a result of a test, such as teachers’ various personalities and teaching styles, making it possible to characterise and compare and contrast lessons and teachers (Alderson & Hamp-Lyons, 1996).

However, the literature on washback has shown that classroom observation tends to be employed slightly less frequently than other research techniques, despite being highly recommended as a primary research method when researching washback (see Wall & Alderson, 1993). According to Wall’s (2005) findings on the impact of the O-Level English examination, although observations were useful in revealing a great deal about the relationship between teaching and the examination, it was difficult to reach a large number of respondents through this time-intensive technique. The study also revealed that observations could not provide much detail about teachers’ reasons for engaging in certain

practices in the classroom, either because the observer did not have the confidence or time to go into detail or because the observer saw no reason to ask for clarification regarding what he/she had observed in the classroom. Other problems with classroom observations (also as suggested by previous washback studies) include that they make teachers feel anxious and as though they are being intruded upon by outsiders (Allwright & Bailey, 1991). Therefore, it is necessary to gather data using another technique to explore participants' perceptions of classroom practices and what they do in the classroom at a deeper level.

The literature on observational research illustrates two distinct approaches used in the social sciences: structured and unstructured. The former requires pre-developed observation schemes to describe interaction in context (such as interactions between teachers and students in the classroom) and is usually linked to numerical data and tabulation. Unstructured observation, in contrast, relies mainly on qualitative data and is designed to observe cultures, lifestyles and perceptions of certain social groups (Denscombe, 2003). Studies of the washback effects from testing mainly draw on structured observation (see Burrows, 2004; Cheng, 2005; Watanabe, 1996). For example, Hayes and Read (2003) recommended that structured observation should be used to allow the collection of direct data, as well as to reduce any bias resulting from the researcher's personal emotions and background, thus producing more objective observations. It should be noted that this technique does not focus on what motivates a behaviour or the factors that caused classroom events; it is generally used solely to observe overt behaviour and manifest actions, that is, it describes *what* occurs in classroom setting, but not *why* it occurs. Regarding unstructured observation, researchers in washback studies consider this approach helpful to gather data on issues that participants might not freely express their opinions about in interviews and to record information on aspects participants might not be consciously aware of (Wang et al., 2014).

Given the complex nature of washback, this research required a research technique that considered the two distinct views of the pragmatist paradigm (see 4.2), as neither view on its own would be sufficient to explain all the different sources of the multi-directional influences involved in washback. For example, to explain positive or negative influences on classroom teaching, it was essential to understand any intervening factors that might interact with the teaching process due to the new examination. As noted by Miles and Birks (2014), an effective qualitative researcher channels effort into aligning the research technique and his/her own philosophical positions. Using structured observation, the researcher ideally has a clear agenda and a well-designed scheme of what is to be observed during the observation

process and ignores any other factors that may have influenced the variables under observation. In contrast, with unstructured observation, researchers are often not so sure about what is to be observed at the beginning of the research and need to recognise that the focus of the observation may change as experience is gained during the observation. In this latter case, researchers are less interested in quantifying verbal and nonverbal behaviours; the data obtained from this approach usually rely heavily on inferences made by the researcher, who is prepared to assign meanings to the observations. The optimal approach is to employ both means of classroom observation as complementary, resulting in semi-structured observation, which combines the features of structured and unstructured observation. Therefore, semi-structured observation was used in this study to provide a scheme that would allow aspects of classroom practice to be examined in a relatively less systematic manner, not just so that that more data could be collected but also so that the observations would be instructive in validating and complementing the questionnaire data.

McDonough and McDonough (1997) contended that a qualitative researcher should define three important parameters when conducting an observation: observer, goals and procedures. In this study, I was an outsider and not one of the research participants. I decided on a non-participant approach to observe closely the natural occurrence of teaching of the new curriculum in the classroom and record any relevant information, while minimising changes in the subtle behaviours of both teachers and students that might occur due to my presence and allowing the participants to engage freely in class without viewing me as a distraction or threat. Nonetheless, there was a risk that my work in the MoE might inhibit the participants from either expressing their thoughts freely or being willing to participate in the study at all. Thus, I worked to overcome this challenge by omitting any data that might reveal the participants' identities and by implementing privacy procedures, such as the use of pseudonyms for the observed schools and participants. The classroom observation tool employed in this study is presented in the following paragraphs.

Observation scheme

Classroom observation instruments are developed to collect data on what happens in the classroom, to demonstrate the effects of different kinds of interactions in language classrooms and to understand how learning opportunities are created (Allwright & Bailey, 1991). These instruments may differ in purpose and use, areas of analysis, source of the variables and units of analysis. As the classroom observation in this study was semi-structured, a classroom observation instrument was needed to provide a framework. The communicative orientation of language teaching (COLT) observation scheme (Spada &

Frohlich, 1995) has been proven to be an effective tool for observing teaching practices in second language (L2) classrooms. The COLT scheme is divided into two parts. In Part A, the observer records “the activities and episodes that occur during classroom teaching”, including the duration of each lesson. Part B focuses on “the linguistic features of classroom talk between teachers and students and amongst students themselves as they occur within each activity or episode” (p. 13). The researcher codes in real time using the COLT observation scheme as the classroom activities and episodes take place.

This study used the COLT scheme for three primary purposes. First, it is commonly used to identify classrooms that use the communicative approach to language teaching. It was considered appropriate in this investigation because it provides a valuable mechanism for describing the actual activities of teachers and students, including their interaction patterns in class. It can be used to provide detailed descriptions of “the differences in the communicative orientation of language teaching and to determine whether and how this contributes to differences in L2 learning outcomes” (Spada & Lyster, 1997, p. 788).

Another key reason is that the COLT scheme has previously been used successfully in previous washback studies. For example, Cheng (2005) used Part A in her study of teachers’ perceptions and behaviours related to the examination change in Hong Kong (see Chapter 3). Watanabe (2004) used the COLT scheme in his investigation, along with field notes, to develop a specific coding system study called Communicative Orientation for Exam Preparatory Classes (COEPREC). Burrows (2004) also used Part A in her washback study to investigate the influence of assessment classroom practices on teachers in Australia. She found that using COLT assisted “in minimising the effect of variables external to the study and allow for the closer observation and analysis of the teachers’ classroom practices” (p. 51).

In this study, the categories in Part A of the COLT scheme (see Appendix E) were used to describe the significant features of classroom activities in the grade 12 English elective classroom and illustrate episodes of classroom interaction. The scheme in this study consisted of the categories listed below, as defined in the COLT observation scheme (Spada & Frohlich, 1995, COLT Observation Scheme, pp. 13–28):

1. Time: how is time segmented in each activity/episode within the lesson as a percentage of class time?

2. **Activities/or episodes:** this is an open-ended construct – that is, no predetermined descriptors have to be used by the observer. Instead, lesson activities are described as separate episodes, such as drills, games, singing, roleplaying, reading aloud, and so on.
3. **Participant organisation:** this parameter describes three basic patterns of how a lesson is carried out with respect to organisation of classroom interactions. These patterns are established by answering the following questions: “Is the teacher working with the whole class or not?”, “Are the students divided into groups or are they engaged in individual seat work?”, and “If they are engaged in group work, how is it organised?”
4. **Content:** this parameter refers to the subject matter of the activities – that is, what the teacher and student are listening to, writing, reading, or talking about.
5. **Student modality:** This category refers to the skills the students engaged during the lessons—listening, speaking, reading and writing—and other skills used in classroom activities.
6. **Materials used:** this parameter focuses on the type and source of teaching materials used in connection with classroom activities, and for what purposes.

An additional category in Part A, content control, was not relevant to this study. For triangulation purposes, I also used detailed field notes, taken during and after the classroom observation. In these, I detailed the teaching activities as well as the time spent on them. I planned to record them on the day of the classroom observation and to note how the teachers conducted their lessons.

Sampling for the observations

Consistent with the research purpose and methodological considerations of this study, the sampling of teachers was purposive (Patton, 1987), with the selection of teachers based on their experience of teaching the new English elective. The following criteria were used to identify potential participants for the observations:

- School principals willing to grant permission for classroom observation in their schools.
- Teachers willing to allow classroom observation.
- Teachers were willing to have their lessons audio recorded.

Following receipt of consent, three teachers were contacted to arrange for five of their lessons to be observed for the main study. For practical reasons, piloting was conducted with two other teachers from the same school. The school was reasonably accessible and thus relatively convenient and safe for me to visit during the COVID-19 pandemic.

Piloting the observations

The aim of piloting was to avoid possible challenges in accessing school sites, to refine the COLT scheme further and align it with the research purpose and to test the usefulness of the observational data for the main research study. Piloting began by obtaining consent from two teachers (Teacher A and Teacher B) to allow me to attend their classes. Both teachers were observed for one lesson, each lasting 60 minutes. During the observations, the COLT scheme was implemented and additional field notes were taken to record classroom activities and episodes independently as the lessons progressed (see Appendix E). With the agreement of the participants, the sessions were audio recorded to allow ease of retrieval and verification of the data.

I noticed that the teachers did not emphasise the activities prevalent in the exam but instead used the textbook as their main source for teaching, providing students with different types of tasks and activities that reflected grammar, speaking activities and vocabulary not related to the final exam specifications. It should be noted that I was acutely aware of the examination specifications and the teaching content approved in the Omani context as I had witnessed the reform of the examination and textbooks for the grade 12 English elective. However, the behaviours observed could not be assumed to be typical of classes as the piloting was conducted with a small sample within a short period of time (Alderson & Wall, 1993).

I noted that no changes seemed to be needed to the observation instrument as there were no difficulties in using it in the classroom. Other important aspects were that I would need to ask the teacher to provide a copy of the lesson materials before attending the lesson to enable me to grasp fully what was happening in the class, especially given my outsider role. Moreover, it became apparent that it would be necessary to fine-tune the clarity of the recorder: when recording the pilot lessons, background noise slightly affected the intelligibility of the teacher and students.

Administration of observations

Considerable advice on undertaking classroom observation is available in the literature (e.g. Angrosino, 2007; Cohen et al., 2018; Darlington & Scott, 2002). In this case, the procedures were designed to be detailed, including the application of the COLT scheme and taking field notes. The latter aimed to record in a more systematic way various activities, skills and aspects of content taught, while retaining a naturalistic presence in the observations. Moreover, the field notes were used to capture supplementary contextual details in the

lessons through descriptive and reflective comments, for example on aspects such as the teacher's image, the physical environment, specific events and tasks and my own responses (see Appendices F and Appendix G for sample completed COLT schemes and field notes).

All school visits were pre-arranged with the teachers and the respective school principals. The study plan was to observe six different teachers, three males and three females, teaching grade 12 elective diploma classes. Due to the COVID-19 pandemic, all schools were employing a blended learning approach during the period of data collection, i.e. classes in Omani schools were conducted via online platforms in addition to learning through the more traditional face-to-face approach. This mix of online and traditional classes made it difficult to arrange observations with the six teachers over the same period of time. The teachers also contended that it was difficult to manage their timetables as some periods of traditional classroom teaching were given longer than others. Moreover, there were restrictions on visiting the male schools as most of the male teachers contacted happened to be ill at the time. Furthermore, two teachers declined to be recorded using video and therefore three audio recorders were set up in corners of the classroom before the start of the lesson to maintain a natural setting and mitigate any disruption I might cause by recording interactions between the teachers and students. As an observer, I did not participate and maintained social distancing to prevent any risks associated with COVID-19 transmission, as well as to reduce the effect of my presence. In each lesson, I sat in a corner at the back of the class.

The observations were conducted from 1 April until mid-May 2021. The observation period in this study led up to the end-of-year English elective examination, which is conducted from the end of June until July each year. This period of time was appropriate for this study as previous empirical washback studies found that the intensity of washback increases when the test date becomes closer (Bailey, 1996; Cheng, 2005; Freeman, 1996; Wall & Alderson, 1993; Watanabe, 2004). Bailey (1999) referred to this as “seasonality” to show the relationship between time and washback.

4.7.3 Interviews

As mentioned above, interviews were used in this study to complement the data addressing RQ1 and provide answers to RQ2. Some qualitative researchers have described the interview as simply “a purposeful interaction between two or more people focused on one person trying to get information from the other person” (Gay & Airasian, 2003, p. 209). Kvale (1996) offered a similar definition, stating that an interview is “literally an inter-view, an interchange of views between two persons conversing about a theme of common interest”

(p. 11). In this respect, interviews have something in common with questionnaires as they tend to generate data from what people tell the researcher about what they do, what opinions they say they have and what they say they believe.

A main consideration in interviews is the degree of its structure (Denscombe, 2014; Dörnyei, 2007; Oppenheim, 1992; Robson, 2002), which is generally determined by the aim of conducting the interview. In structured interviews, the researcher follows pre-prepared guiding questions and prompts, offering limited options in terms of responses from the participants. The responses that result have various advantages, such as comparability across people or sites, and disadvantages, for example providing limited data, attributes that they share with questionnaires. At the other end of the spectrum, in unstructured interviews the researcher begins by introducing the key area under investigation and then lets the ideas develop in an unpredictable direction, with maximum flexibility in how questions are initiated or answered in relation to the research agenda. The semi-structured interview allows researchers to use predetermined guiding questions and prompts to encourage interviewees to engage in conversation and talk more widely about the area of interest or concern in an exploratory matter.

As outlined in the literature, the structured and semi-structured interview types have typically been used in washback studies. Burrows (2004) employed data from structured interviews to propose a series of hypotheses for the observation phase of her research, as well as to compare the answers of respondents in terms of the type of changes they experienced in teaching and assessment practices resulting from the examination. In other washback studies, semi-structured interviews have been employed to clarify and elaborate on certain answers from a questionnaire. Wall and Horak (2006) took this approach to “gather data that would offer deeper insights than questionnaires could provide, even if it meant working with fewer participants” (p. 26). Unstructured interviews have been employed only rarely in washback studies to allow the collection of open-ended follow-up data after classroom observations (Cheng, 2005; Choi, 2015).

Given the multi-directionality of the washback phenomenon, this research required the application of techniques able to document the different sources of the effects of testing on teaching by both asking about and observing teaching and learning (Bailey, 1999; Cheng, 2005; Wall & Alderson, 1993). While questionnaires explore participants’ perceptions of the implementation of a new examination, interviews can provide answers about the reasons for what teachers do in the classroom, offering deeper insights into the complexity of washback

types and processes and suggesting further potential lines of inquiry. As stated by Richards (2003), interviewing is “a journey within a journey” (p. 65); it is a process of exploring, knowing, and maintaining a conversational atmosphere with members of the society, rather than teasing out something already known or definitive in itself.

Therefore, in this study, conversations with research participants sought to draw on the questionnaire results and discover further aspects of washback resulting from the new examination, as well as to explore the participants’ stories about the relationships between testing and classroom teaching practices. This was important as what might be positive for examination writers or curriculum developers might be negative for teachers because the different evaluators involved in washback have different objectives (Alderson, 1992). Hence, semi-structured interviews were conducted with a sample of participating teachers and MoE decision-makers to allow a range of responses and permit follow-up of unanticipated leads. Although semi-structured interviews use predetermined guiding questions, they can be reworded and supplementary questions can be asked to probe further, which still maintaining a conversational atmosphere. Some of these data were useful in validating questionnaire responses.

Selecting questions for the interviews

For the interviews with both teachers and MoE decision-makers, a range of themes and probes were designed to provide a guide (see Appendix H for the teachers’ interview guide and Appendix I for the MoE decision-makers’ interview guide). The themes were based principally on the research questions and the probes were developed to explore the impact of the new examination, as well as to explore the principles and practices underpinning the change. For triangulation purposes, some themes were the same as those in the research questionnaire, allowing interviewees to confirm, clarify and elaborate on certain answers. In addition, some prompts were used to ensure that the interviewees understood the interview procedures or questions and some of the questions raised in the interviews were evoked simply by my curiosity following the classroom observations.

The interview scheme was originally designed in English. However, the interviews were all conducted in Arabic as this was the first language shared by all the participants and would thus be the language most comfortable for them to communicate in. Using Arabic had the advantage that it would be possible to minimise potential misunderstandings in expressing opinions and ideas (Geisinger, 1994). To enhance the validity of the interview protocols, the Arabic version was translated back into English to look for ambiguities and omissions. I

carried out the translation and it was then read and revised by two of my colleagues, both of whom were postgraduate Omani students studying in UK universities and who shared attributes with the participants in this study (for the Arabic translations, see Appendices H and I for teachers and MoE decision-makers respectively).

Sampling of the interviewees

The interviews were conducted in phases one and three with both teachers and MoE decision-makers. As I had worked as an examination writer in the MoE for more than a decade, a self-selection sampling strategy was used to approach MoE decision-makers: emails were sent that provided a brief introduction concerning the study focus and an invitation to participate in the interviews. Once respondents agreed to be participate, I scheduled interviews with members from the Assessment Measurement Centre and Educational Evaluation Directorate, Curriculum Development Directorate and Supervision Directorate.

In terms of the teacher participants, a purposive sampling technique was used to recruit five teachers from among those who had completed the questionnaire and expressed their interest in being interviewed and recorded. Thus, I was able to target a wide range of perspectives and select “individuals who [could] provide rich and varied insights into the phenomenon under investigation so as to maximize what [could be learned]” (Dörnyei, 2007, p. 126). On this basis, one-on-one interviews were conducted at the interviewees’ school sites, lasting between 30 and 40 minutes each.

Piloting the interviews

After the interview questions and themes had been prepared, I conducted piloting to experience the interview process and familiarise myself with the skills necessary to resolve any problems that might arise during the process. Gillham (2005) described two different phases in piloting interviews. The first is the pre-pilot phase, in which the researcher asks for critical feedback on the format and wording of the questions. The second is the pilot phase proper, which focuses on ensuring the interview scheme is appropriate. In the pre-pilot stage in this study, aimed at checking for understanding (Breakwell, 2012), two colleagues who were involved in the change to the English elective examination were asked to read the questions and report back any doubts or queries they might have about the interview questions, especially to identify any items which were redundant or repetitive. The second stage involved checking comprehension of the interview questions, developing a realistic sense of how long the intended interview would take and noting relevant questions that had

not been incorporated in other phases of the research. In this stage, two interviews were conducted with Teacher A and Teacher B based on their responses in the questionnaire and willingness to participate in the interviews. In addition, one interview was conducted with a decision-maker in the MoE. With the participants' consent, the interviews were conducted on a face-to-face basis at their sites, each lasting approximately 30–40 minutes. All interviews were audio recorded (with their permission) and complemented with written field notes to reflect on the questions and assess what was (not) working. The three interviews were translated into English and transcribed.

The data collected were highly instructive for the main study (phase three), allowing constructive changes to the protocols and procedures for the main interviews. One aspect that became clear from the pilot study was that translating and transcribing interview data demanded a great deal of time and effort. However, by transcribing the data, I gained a comprehensive understanding of the process and was able to extract important information from the transcript. Likewise, the piloting of the interviews enabled me to understand key issues related to the research problem and thus highlighted the need for this research study.

It was apparent that 45 minutes was sufficient time to conduct the interviews with both teachers and MoE decision-makers. This would allow them to communicate their opinions and thoughts freely, which was important as no new information would be obtained from them after the interviews. I also noted that the teachers were hesitant about expressing views that might imply criticism of decision-makers' actions regarding the changes to the examination, possibly because they felt intimidated. According to Atkins and Wallace (2015), such power relationships within an interview setting can be addressed by assuring confidentiality and providing a relaxed atmosphere, making participants feel that they are having a conversation rather than a confrontation. Related to this was the need to conduct the interviews in such a way that the interviewees could respond and expand on issues without any interruption. This would make the interview more effective in obtaining the desired range of responses and remaining open to unexpected perspectives.

A major issue that the pilot study identified was that the extent of the match/mismatch between teachers' perceptions of the curriculum (textbook) change and what they knew about the changes to the examination and its specifications led to discrepancies in the ways they perceived the impact of the new examination on their classroom teaching. Thus, I had to pay attention to how I phrased some questions in the interviews to ensure the teachers would understand and that their responses would address the relevant research question.

Furthermore, piloting the interviews with the MoE respondents led to various reflections concerning comprehensibility, resulting in rephrasing, re-writing, re-ordering and editing some questions and probes (e.g. Q4, Q6, Q8, Q11, Q12 in Part 2). The two teachers' interviews indicated no need for substantial additions or omission of questions. The final versions of the schemes are given in Appendices H and I.

Administering the interviews

Steps in conducting interviewing have been identified widely in the literature (e.g. Creswell, 2013; Denscombe, 2014; Dörnyei, 2007; Gillham, 2005; Richards, 2003; Robson, 2002). In this study, the interview processes in phases one and three benefited from the key techniques outlined by Breakwell (2012). First, the technical aspects of the digital voice recorder were checked to ensure the quality of recordings. Second, I familiarised myself with the interview schedule before the start of each interview. Third, all participants in the interviews were asked the same questions and were given an equal hearing, even in cases where I could predict their answers, to reduce potential biases in data collection. Finally, further probes were often used to elicit elaborations on particular answers and prompts were sometimes necessary when participants did not provide sufficiently specific answers or it was necessary to ensure I had understood their intended meaning.

4.8 Data Analysis

Data analysis in this study consisted of separate analyses of the quantitative data (from the questionnaire) and the qualitative data (from observations and interviews). Analyses of questionnaire data in this study ranged from simple descriptive statistics to complex multivariate procedures. The relationships between categories of questionnaire items were used to build informed interview questions and the observation scheme, which were coded and categorised into themes.

Using the quantitative and qualitative data, I was able to integrate data obtained from the questionnaire and the observations. These results were used to develop interpretations and understandings that would fairly and comprehensively explain the process of the washback phenomenon as related to the study's primary research questions.

4.8.1 Quantitative data analysis

The responses to the teachers' questionnaire comprised the quantitative data for this study. The questionnaires were stored, coded and then analysed using SPSS. Data were initially subject to descriptive analysis, including the frequency distribution, mean, standard deviation, and median and mode. The means and frequency distributions were used to

establish the average values for the questionnaire items related to teachers' perceptions of the English elective examination and its impact on their teaching practices. The standard deviations helped infer the amount of variation in the target population. In addition, the median was calculated to investigate the central tendency in the data. The mode provided the answer most often given for each item in the questionnaire.

Furthermore, data were analysed using inferential statistics to assess whether there were statistically significant differences in responses attributable to certain variables, specifically the demographic characteristics of teachers (RQ4). To this end, non-parametric tests of difference were conducted: Mann–Whitney U , t -tests, one-way ANOVA and chi-squared (χ^2). The statistical rationale for this was that the Shapiro–Wilk test of normality of distribution revealed that the data were not normally distributed. The inferential tests aimed to determine whether factors such as the teachers' gender, qualifications, age, teaching experience, training opportunities and number of teaching hours per week contributed to the perceived washback intensity of the new English elective examination. If the examination produced strong or weak effects only on certain teachers, it would likely be mediated by such factors (see 3.8).

4.8.2 Qualitative data analysis

According to Spradley (1979), qualitative data analysis is a process whereby researchers systematically examine something to determine its constituent parts, identify relationships among those parts and identify each part's relationship to the whole. This process involves searching for and arranging data, synthesising and summarising data, identifying significant patterns, and finding out what has been learned, what is to be learned and what is important (Bogdan & Biklen, 1982). Qualitative analysis can be used to allow the researcher to develop interpretations, derive concepts, explanations, models, or theories, or to pose new questions or offer new directions in the field. Miles and Huberman (1984) contended that qualitative data analysis consists of three concurrent activities – reduction, simplification, abstraction and transformation – and argued that data collection and analysis should overlap to allow the researcher to be open to new ideas and patterns that may emerge. In the case of this research, therefore, the analytic process began simultaneously with data collection to reduce the overload of data and identify significant themes that could be explored later in the time remaining for data collection; hence, for example, data from the analysis of the questionnaire were used to add some questions to the interview protocols.

In preparation for qualitative data analysis, Cohen et al. (2018) stated that researchers should have a clear purpose as this determines the kind of analysis that should be undertaken. One important consideration in this study was identifying themes related to the research questions and representing the “patterning of meaning across the dataset” (Braun & Clarke, 2006, p. 82). The interview data were analysed using Braun and Clarke's (2022) process of thematic analysis. This is a method for “developing, analysing and interpreting patterns across a qualitative dataset, which involves a systematic process of data coding to develop themes” (p. 4). I opted for this approach because a “rigorous thematic method can produce an insightful analysis that answers particular research questions” (Braun & Clarke, 2006, p. 97). The analysis allowed me to capture how the washback phenomenon influenced teachers’ practices in the classroom. This involved investigating the meaning ascribed to the introduction of the new examination by the participants, as well as the meaningfulness of the impact for the participants.

Based on the thematic analysis approach, the core elements of qualitative data analysis were adopted. These processes involve preparing and organising data for analysis, reducing the data into broad themes through the coding process and condensing the applied codes, and finally, for representation and comparison of the data, using figures, diagrams and/or discussion (Creswell, 2013). These processes are useful for working flexibly in terms of collecting initial themes and then defining and categorising the codes of each theme. More specific details of the data analysis are given in the next section.

4.8.3 Observation data analysis

To transcribe the data, it was necessary first to become familiar with them by reading and rereading the notes and transcripts and repeatedly listening to the audio recordings. This process helped me retrieve and identify additional details for the data analysis process (Ritchie & Lewis, 2003). The recorded data were saved in the form of text files to be ready for analysis. The data were not transcribed verbatim; rather, written descriptions of the teachers and classes were organised into a time frame (see Appendix J and Appendix K for sample transcripts). The process of transcription was useful as it prompted the recall of information and generated initial codes and patterns within the classroom observation data; the field notes and materials collected in the lessons were used to inform decisions when the notes from classroom observation alone were not clear.

Following the transcription, the next step was coding. Creswell (2013) asserts that the process of coding “represents the heart of qualitative data analysis” (p. 184). Essentially, I

first used a “descriptive coding” method, one of the 23 coding methods described by Saldaña (2013): this method is most suitable for the characteristics of data obtained from classroom observation. It was used here to identify and label how data related to the washback phenomenon. As the observation aimed to provide full descriptions of what happened in the classrooms, descriptive coding was used to describe the classroom atmosphere and teaching methods to the reader, namely to denote what I saw and heard during data collection (Wolcott, 1994), rather than to scrutinise the nuances of the interactions of the people involved in social action.

To take the analysis of data from a descriptive to an analytic level, the other method of coding used in this study involved categorisation. The data concerning classroom activities and episodes were recorded in Excel files and expressed based on the percentage of overall class time.

4.8.4 Interview data analysis

The interviews with the teachers and MoE decision-makers in phase three were all audio recorded. The focus of analysis was on how the information obtained could be employed to narrate the participants’ experiences with regard to the washback effect. The interview data were analysed according to a three-stage procedure, as employed in the literature (Creswell, 2007; Dörnyei, 2007; Miles & Huberman, 1984): transcription, reduction into themes/patterns through a process of coding and reporting. According to Braun and Clarke (2022), themes are identified through a systematic process of familiarisation with the dataset, coding of data, generating initial themes, and developing and reviewing themes. The procedures used in the analysis in this research largely followed the proposed phases outlined by Braun and Clarke (2022).

The first stage involved transcribing the interviews in Arabic into Microsoft Word text files to prepare them for analysis (Dörnyei, 2007), focusing on the content rather than the form of the verbal data. Hence, the data were transcribed verbatim, without any content selection or editing, because it was not clear at this stage which data might eventually be important. As Miles and Huberman (1994) noted, the meanings of the recorded data “have to be tested for their plausibility, their sturdiness, their ‘confirmability’ – that is, their validity” (p. 11). Thus, the transcribed scripts were sent to the participants for them to check and verify whether the transcripts were compatible with what they had said and the ideas they had tried to express.

The transcripts were then imported into the qualitative analysis software NVivo version 12.1, which has several benefits in terms of its features, such as its high storage capacity and easy indexing (Dörnyei, 2007). As I listened to the audio-recorded data and read through the transcripts for each participant, nVivo enabled me to make “familiarisation notes” (Terry & Hayfield, 2021) concerning the most interesting or puzzling data. This was useful to record “segments of similar or related text” (Fereday & Muir-Cochrane, 2006, p. 4). Once I had familiarised myself with the data and established a firm sense of the content, I coded and categorised the data to confirm their relevance and accuracy in the identification of themes and patterns. The coding process allowed me to review many transcripts and to make effective data comparisons between different participants.

In this study, the coding phase started by labelling meaningful pieces of data that corresponded to key concepts. Saldaña (2016) suggested that “the act of coding requires that you wear your researcher’s analytic lens. But how you perceive and interpret what is happening in the data depends on what type of filter covers that lens” (pp. 7–8). Based on Saldana’s (2016) coding methods, in the first cycle, the coding was guided by the theoretical frameworks and hypotheses representing washback processes (see 3.4) and conceptual framework for this study (3.10.2). When satisfied that the codes generated were aligned with the aim of this investigation, the perceptions of teachers concerning the new examination were studied to understand the washback process and ensure that the related concepts were fit for purpose. Data-driven coding was also used for the interviews with the MoE participants to identify patterns of meaning related to RQ3.

In the second cycle, “pattern coding” was used to sort the nodes in terms of patterns or themes or according to their matching categories. According to Miles and Huberman (1994), “[pattern] codes are explanatory or inferential codes, ones that identify an emergent theme, configuration, or explanation... Pattern coding is a way of grouping those summaries into a smaller number of sets, themes, or constructs” (p. 69). In this cycle of analysis, the patterns were re-coded multiple times by “grouping those summaries into a smaller number of categories, themes, or concepts” (Saldana, 2016, p. 236). During this recurrent re-coding cycle, codes without references were removed for simplicity and repetitive patterns/or ideas were addressed.

The third stage of thematic analysis concerned the development of the themes themselves. The coded nodes were read and recoded on numerous occasions to identify significant border patterns among the themes. In the process of coding the MoE interviews, however, the initial

analysis produced detailed data about the plan and strategies for introducing the new examination. As the goal of my analysis was to explore the intentions in introducing the new examination, I considered that data concerning the implementation were beyond the scope of the investigation. After re-coding numerous times, themes were formed from the interview data.

The last stage of this analysis was carried out to warrant the validity, trustworthiness and credibility of the patterns/themes and codes. Regular meetings were held with an independent colleague to discuss the interview transcripts and the corresponding codes, particularly in terms of discrepancies. Whenever a new code was modified or added, the original coded transcript was checked and recoded accordingly. Once completed, the themes used to guide the presentation of the data were agreed upon (see Appendices L and M for a complete list of codes and themes).

4.9 Validity and Reliability

The criteria for the quality of research in the social sciences is a key issue that needs to be conceptualised and operationalised from the start of a study and through each research step until the publication of the study findings. Given the pragmatist paradigm underpinning this study, qualities such as validity and reliability were especially important (Ary et al., 2010; Babbie, 2004; Bryman, 2004; Cohen et al., 2011). Quantitative research emphasises the importance of validity and reliability, whereas qualitative research addresses these values in terms of quite different criteria, such as those proposed by Lincoln and Guba (1985): “trustworthiness”, “transferability” and “credibility”.

The concept of validity refers to the extent to which an instrument measures what it intends to measure (see, e.g. Cohen et al., 2011). Other definitions state that validity is not only concerned with the instrument itself but with the interpretation of the data obtained from the instrument (Ary et al., 2010). Moreover, validity takes several forms, the main forms of which related to this study are addressed here. One is *content* validity, which refers to whether the research instrument fully and fairly covers the content being investigated (Ary et al., 2010). It is not possible to achieve validity at an absolute level, but researchers often seek to minimise a lack of validity in their research (Cohen et al., 2011). Other forms of validity are *internal* validity (demonstrable causality between two or more variables) and *external* validity (whether the results of a study can be generalised beyond the target population) (Bryman, 2004). The concept of external validity pertains to the degree to which research outcomes can be applied to individuals, situations and periods of time not directly

focused on during the data gathering process. Another important form of validity is *face* validity, which is a subjective judgment of whether a certain measure seems reasonable for obtaining the data the research is intending to gather. Safeguarding the aforementioned forms of validity is therefore discussed further in a following section.

Besides validity, reliability is another important criterion in evaluating research. The concept of reliability is often defined as “how consistently a test measures whatever it intends to measure” (Ary et al., 2010, p. 224). In other words, reliability refers to the stability, dependability and replicability of the measurement. Cohen et al. (2011) suggest that “for research to be reliable it must demonstrate that if it were to be carried out on a similar group of respondents in a similar context (however defined), then similar results would be found” (p. 268). In quantitative research, reliability can be achieved through different measures, for example split-half reliability, item-to-total correlation, Cronbach’s alpha (α) coefficient, test–retest, inter-rater reliability, etc. (Robson, 2002). In this study, Cronbach’s α was used to establish the internal consistency of the questionnaire items and presented high internal consistency, with a coefficient of approximately 0.88.

The meaning of the term *reliability* in qualitative research is still a matter of debate. For example, Cohen et al. (2011) suggested that when the same instrument is used in the same context in terms of the research participants, the results should be the same, whereas LeCompte and Preissle (1993, cited in Cohen et al., 2011), proposed that “the canons of reliability for quantitative research may be unworkable for qualitative research” (p. 332). The interview transcripts were coded by two researchers using the coding framework but intercoder reliability was not examined. However, coders were provided with the relevant information and briefed on the procedures and the purposes of this research. In the case of any mismatch, the particular coding was discussed until common agreement was reached.

To establish the trustworthiness of this research, I adopted a triangulated approach, which Hawkey (2006) describes as an important practice in washback research. As discussed in 4.4.1, triangulation was applied by obtaining data through both qualitative and quantitative methods using different instruments to examine the same research phenomenon. Moreover, I conducted a pilot study to ensure the participants in the main study would be able to understand the interview questions and to ensure that the research instruments were well designed and the data obtained in this phase would be suitable for answering the research questions. In addition, the study drew on various aspects of different studies which had similarities to this research in terms of purpose and research problem to help design the

instruments and guide the research techniques used for data collection. Some of the themes in the teachers' interview scheme were adapted from the previous washback literature (e.g. Qi, 2005; Wall, 2005) to suit the research context and purpose of this thesis.

In addition, I took into consideration Silverman's (2000) argument of the importance of establishing a relationship of trust between the researcher and research participants. Having worked as a teacher in the system helped me establish this sense of trust with the participants (see 1.3 and 2.1). I was also in contact with the participants before and throughout the collection of the interview data. This prolonged engagement entailed contact with research participants even after completion of data collection. Another strategy used in this study was member-checking. The transcribed interviews were checked by the participants to establish the accuracy of the transcripts (see 4.8.4). The translation was also checked for accuracy by a colleague with relevant expertise and back translated (Arabic to English and English to Arabic); a high level of consistency was found between checkers (see 4.8.4).

Furthermore, I sought to ensure trustworthiness by providing detailed descriptions of the research context, thus allowing readers to judge the extent of similarity to other settings and establish whether the findings could be extended to other or similar contexts. These descriptions included information about the study context, such as the places where the participants lived and gained their knowledge and experience, as well as thorough descriptions of the design and administration of data collection and data analysis methods. Thus, future studies would be able to replicate this research in similar contexts. Moreover, I acknowledged and was transparent about my beliefs and knowledge, which could influence my interpretation of the data, as well as providing justifications for my decisions about choice of the research strategies used over other approaches and noting the weaknesses in these strategies. All possible efforts were made to explore instructional practices in the classroom fully within the Omani context and to present the study findings clearly and transparently. The following sections provide more details on how reliability and validity issues were enhanced.

Questionnaire validity and reliability

Although this research adapted a well-established questionnaire from Cheng's (2005) comprehensive washback study, several validity and reliability procedures were utilised in the implementation of the questionnaire in phases one and two of the study. To ensure the suitability of Cheng's (2005) questionnaire for this research, as described above, the questions were revised and analysed to ensure they included items that would address the

research questions and objectives. Technical guidelines for writing questionnaire items were also considered in writing some items in the questionnaire (e.g. Part 3: Q1–Q4). Moreover, the piloting phase of the questionnaire (see 4.7.1) further enhanced the development of the questionnaire in several respects. Various methods were employed in the piloting phase to ensure the content and face validity of the questionnaire. The design of the final questionnaire for the main study was verified by two PhD researchers experienced in language assessment and teaching, who evaluated the questionnaire in terms of the clarity of questions and their relevance to the study objectives.

Observation validity and reliability

One major source of invalidity in observation-based research is the inability to generalise the results to other situations due to the subjective nature of participant observation (Cohen et al., 2011). Another source of invalidity is that the presence of the observer might inhibit participants from freely expressing their thoughts, or may cause participants to alter their behaviours, or discourage participation in the study altogether. Triangulation of data and methods in this research was an essential procedure to overcome any validity and reliability issues in classroom observations. Additionally, I tried to overcome some threats to external invalidity by adopting the role of non-participant observer, by omitting any data that might reveal the participants' identities and by implementing privacy procedures, such as the use of pseudonyms for the observed schools and participants. Furthermore, to control problems that might have occurred during the observation, I focused on the main issues that needed to be looked for in relation to washback from the new English elective examination on classroom teaching practices. This involved piloting the observation instrument to enhance understanding of the research problem and to ensure that the observational categories used for this study were appropriate, clear, specific and discrete, and that they effectively addressed all the research questions.

Interview validity and reliability

Lack of validity in interviews is often expressed in terms of bias (Cohen et al., 2011). Maxwell (2005) noted that interview bias might be attributed to different sources, such as the expectations, attitudes and opinions of the interviewer, preconceived theories or notions, or misinterpretations, misperceptions and misunderstandings on the interviewer's part. Moreover, other causes of invalidity may emerge from biased sampling, poor prompts and biased probes, poor materials, poor rapport with interview participants and alterations to the interview scheme (e.g. changes to the order of the questions) (Oppenheim, 1992). Another major source of invalidity could emerge in terms of face validity – that is, whether the

questions in the interview appear to measure what they are intended to measure (Cohen et al., 2011). Additionally, in studies that conduct multiple interviews, the types of interview questions may be inconsistent. Therefore, in this study, the interviews were piloted to ensure that the questions were appropriate and effectively operationalised the constructs represented in the research questions. Another method of validating the interviews comprised triangulating the data with those obtained from the other instruments (questionnaire and observations). To reduce the risk of inconsistency in the questions, I conducted all the interviews and the core questions remained unchanged from one interview to another.

4.10 My role as a researcher

Silverman (2010) highlighted that the researcher role involves fairness, honesty, integrity and sensitivity during the collection and analysis of data. For Miles and Huberman (1994), confirmability is a key principle, which refers to the extent to which a researcher admits his/her own dispositions and the data and interpretations are not influenced by assumptions or biases on the part of the researcher. Having worked in the MoE and schools in which this study was conducted (see 1.3), I had shared experiences with the research participants who took part at different stages of the study. I thus positioned myself as an insider as I was familiar with the research context and had worked for the MoE for several years. Being an insider researcher played a major role and contributed greatly to my understanding of the contextual issues of this research study, as well as guiding the investigation process. As noted by Ryan (2011), insider knowledge can help give researchers a “surplus of seeing” (p. 642) and confer a higher level of credibility and objectivity than being an outsider.

As a researcher, I was fully aware that my knowledge of and experiences with the research participants and the context might influence or affect my interpretation of the data and introduce bias. Therefore, throughout data collection and analysis, I took care to engage in critical scrutiny to reduce researcher bias, for example minimising the power issues that could have affected the research participants’ perceptions and behaviours. Rather than avoiding my preconceptions about the test and context, I sought to examine them objectively and to be persuasive and empathetic with the participants. I was careful not to ask leading questions that might reflect my own biases and elicit responses affected by my own views and knowledge regarding the introduction of the new elective diploma examination; rather, I asked probing questions only to expand on and clarify what had already been stated by the research participants.

When conducting the teacher interviews, the participants were open and able to speak freely rather than feeling constrained to stress certain opinions or behaviours. I ensured that the interview questions and prompts that started each string of thought, as well as their order and wording, remained unchanged in each interview to better enable comparison and contrast of responses across the research participants. I also showed interest when the participants were discussing their views and expressing their opinions and expressed on many occasions how valuable their participation was to me and the research.

As part of my researcher role, I also adopted an outsider perspective to interpret the data from a critical perspective. This was facilitated by the fact that I had been out of the teaching profession for several years and significant changes had occurred due to the introduction of new curricula and examinations over that time. Rabe (2003) asserted that researchers who have an outsider role will be better able to investigate their research problem and questions with “new” eyes than an insider researcher who may take things for granted. From my outsider standpoint, I derived new insights and knowledge and I viewed the practitioners’ experiences and attitudes from different angles, adopting a fresh critical perspective. However, I also shared my own knowledge and experience to build up mutual trust with the research participants and develop an understanding of the research phenomenon, as advised by Seidman (2006). In particular, it was important to share some of my work experience and background knowledge when research participants asked about a situation that was similar to something they had already described in detail.

Furthermore, I took care to clarify the participants’ responses even though they were pragmatically understood. For example, some of the MoE research participants started their interviews with me using the expression “as you know”, relying on my experiences in the study context. I would then ask them to clarify and explain what they meant by that. I also expressed my own understanding of the situation clearly to make sure the participants and I had shared knowledge. Thus, in this thesis, when articulating my objective role, I identify myself as the “researcher”, but when being reflexive, I use the first person “I”.

4.11 Ethical Considerations

Ethics has been defined as “a matter of principled sensitivity to the rights of others” (Cavan, 1977, p. 810). Ethical principles are an essential consideration in conducting research, from the inception of the research problem to the interpretation and dissemination of the findings. This often entails decisions which “relate directly to the integrity of a piece of research and of the disciplines that are involved” (Bryman, 2012, p. 130). The researcher has an ethical

responsibility towards the research participants to use any means available to preserve their rights and dignity as human beings (Cohen et al., 2018). Some commonly addressed ethical principles are informed consent, privacy, confidentiality and anonymity.

The principle of informed consent comprises four main elements: competence (ensuring that individuals are capable of taking appropriate decisions if they are given the relevant information); voluntarism (only people who voluntarily choose to participate in the research do so and they should have the opportunity to agree to exposure to risks at any stage); full information (participants are given adequate information about what they are consenting to); comprehension (participants are made fully aware of the nature of the research) (Cohen et al., 2018). For Henn et al. (2006), ensuring privacy entails the researcher informing participants about their rights in taking part in the research, especially research that requires them to provide private information (personal or sensitive), such as age, income, or marital status. One way of ensuring that participants' privacy is protected is through taking steps to protect confidentiality. The researcher must not disclose information regarding the research participants or pass information on to others. Another way of protecting privacy and harm is through assuring anonymity, namely that the participants cannot be traced or identified from the information they provide, no matter how personal or sensitive the information.

In this thesis, precautions were taken at every stage to preserve the privacy, anonymity and confidentiality of the participants. First, the University of York Ethical Issues Audit Form was completed (see Appendix N). This is an information sheet related to the data collection methods, detailing the aim of the study, the justification for the research and descriptions of how the data will be processed, along with specific details about data management, particularly with regard to ensuring confidentiality and the safe storage and protection of the data. Participants were assured in the information sheet that all information concerning them would be anonymised to avoid breaches of privacy and confidentiality in the data collection process, except that the questionnaire contained codes linked to teachers' names for only those teachers who agreed to participate in a follow-up interview, thus enabling me to contact those teachers in their schools, as suggested by Brannen (2005). Moreover, pseudonyms were used in the transcriptions of the observations and interviews with the teachers and MoE decision-makers. An email from the university then confirmed that the form had been received and that data collection could be conducted starting from 2 November 2020. Based on receipt of ethical approval from the university, permission to access schools in Oman was granted by the MoE, which stands as the gatekeeper for public and private schools in Oman

(see Appendix O). All the study participants provided informed consent (for consent forms, see Appendix P).

Due to the unprecedented conditions imposed by the COVID-19 pandemic, the Department of Education Ethics Committee at the University of York stated that researchers should not employ face-to-face data collection, unless online or virtual approaches were not possible, to protect the health and wellbeing of the research participants. In this study, however, a detailed record of the participants' practices in the classroom, including the physical setting and the ways in which the different parties interacted, was essential to serve the aims of the research; thus, a virtual approach was not feasible. As I was in Oman to conduct data collection, I followed the guidance in Oman regarding COVID-19 at the time, which was that post-basic education schools were to remain open and only grade 12 students were allowed to study in traditional classrooms as long as they followed COVID-19 regulations, such as maintaining social distancing and wearing a face covering. Therefore, the Department of Education at the University of York approved my request to conduct face-to-face research, adhering to Oman's COVID-19 regulations (Appendix N).

The safety and wellbeing of all involved in the study – the participants and me – were of the highest priority and were safeguarded in this research using systematic and clear procedures. For the interviews, the participants were made aware of COVID-19 safety guidance and practices, such as maintaining social distance, wearing a face cover or mask and not attending the interview if they were experiencing any health issues. To avoid putting pressure on the participants, I avoided conducting the interviews during the teaching time block, when teachers are very busy with their day-to-day teaching and related responsibilities. Instead, the interviews were conducted towards the end of the second semester of the school year, when the participants were free from teaching engagements and could more easily manage their time. Prior to conducting the interviews, I contacted the school leadership and requested a private and well-ventilated room in which to conduct the interviews. Moreover, I took weekly PCR tests from the beginning of the school visits to reduce the threat of virus transmission.

The arrangement of the class and my position during observations was negotiated and discussed with the teachers via email. When observing classes, I sat in a corner at the back. I maintained social distancing, wore a mask throughout the entire lesson and assumed the role of a passive observer so as not to interact with the teacher or students. Moreover, to safeguard all parties, including myself, only schools in the low-density category were

targeted to minimise the chances of transmitting the virus. Furthermore, I checked with the teacher and ensured health measures were in place before the day of the classroom visit.

During data collection, those taking part in the research in any capacity were asked to sign a consent form that addressed all aspects of the ethical principles presented here. The research was conducted on the basis that the data would be anonymised and all participants were told that the data would be confidential to prevent potential distress caused by breaking anonymity. To avoid any identification of the participants, their names and schools were changed and pseudonyms were used throughout the different stages of the study, as they are in this final written thesis.

I informed the participants that the data gathered would be dealt with collectively rather than individually to address the research aims and questions. This approach was intended to facilitate collection of valid data and reduce the possibility of social desirability response bias (Robson, 2002). As the questionnaire respondents were also involved in the interviews, the questionnaires were not totally anonymised but included codes for teachers' names and emails to help me choose a purposive sample of teachers for the interviews.

The participants were made aware that their involvement in the study was completely optional and they could withdraw within a certain timeframe (up to two weeks after data collection concluded) without the need to provide any reason. After the data had been collected, the participants were asked to confirm their agreement concerning the content of the interviews and observations and to ensure that their thoughts were fully represented. They were given a window of one week to respond with their comments and views.

4.12 Summary

In this chapter, after stating the philosophical position underlying the study (pragmatism), the four research questions and the overall research design were presented (4.4). The phases of research, sampling processes and piloting were also described (4.5–4.6). Specific methods of data collection were discussed, along with the rationales, structure, piloting, sampling and administration procedures (4.7). An explanation of how the data were analysed using SPSS for the quantitative data and NVivo for the qualitative data was provided (4.8). Issues of reliability and validity in relation to the questionnaire, observations and interviews were discussed. The chapter concluded by addressing ethical considerations.

In the next chapter, the findings related to the research participants' perceptions of and reactions to the new English elective examination are presented. The findings from the

classroom observations are also presented to illustrate the nature and scope of the washback effects from the new examination on grade 12 teachers' behaviours. The perceptions of the MoE decision-makers (exam writers and curriculum developers) are discussed to shed light on the relationships between their intentions and teachers' practices in class. Additionally, factors other than the examination that influence teaching are delineated to illustrate the extent to which teachers' personal factors and exam-related factors influence the intensity of washback from the new English elective examination.

Chapter 5. Findings

5.1 Introduction

Tests – particularly high-stakes tests used for selection and certification purposes – are widely believed to have a strong impact on educational practice. Over the last decade, the Omani education system has completed a series of national educational evaluations (see Chapter 1). The evaluation results indicated a need to raise the overall English language level to meet the demands of higher education and the labour market in Oman. In an effort to address this challenge, a new English elective diploma examination was introduced in the 2018–2019 school year, based on a newly developed curriculum. Despite concerns that the new examination would cause a shift towards an exam-oriented system, there has been no empirical research undertaken to ascertain how the revised content and the properties of the examination have affected classroom teaching processes. This study aimed to fill this gap by investigating the factors that have shaped the intended washback from the English elective diploma examination as proposed by the MoE decision-makers, with a view to informing policymakers of how washback (positive or negative) operated in the context of the study.

This chapter presents the main findings regarding the four research questions. All the findings were drawn from quantitative data analysis using SPSS for closed questions and qualitative data analysis using NVivo 12 (see 4.8). As mixed methods were adopted in this study (following a “sequential triangulation strategy”), the findings are presented in chronological order (see 4.6 and 4.7). As described by Creswell (2003), this approach enables comparison and supplementation of the study results. The qualitative phase of this thesis drew on the results of the first quantitative phase and explained further aspects of the perceived washback effect from the new examination. Furthermore, qualitative data helped uncover the relationships between what the teachers believed and what they did in the classroom and whether the new examination had affected teaching in the way intended by the MoE decision-makers.

In analysing the questionnaire data, descriptive statistics were used to analyse the teachers’ responses to closed questions. Thematic analysis, following a deductive approach, was adopted for the open-ended questions from the teachers’ questionnaire (see Appendix D). To assess which demographic factors could significantly predict variations in teachers’ classroom teaching and assessment practices, inferential statistics were run. These comprised non-parametric tests as the data were not normally distributed.

For the qualitative data obtained from the classroom observations and interviews, thematic analysis was undertaken in the NVivo 12 software, aimed at systematically identifying interesting and relevant information (themes) emerging from the data (see 4.8.2). The results are reported and discussed according to the order of the research questions. For this purpose, this chapter includes the following four sections:

- Teachers' perceptions of the new English elective examination (RQ1).
- The intended washback from the new English elective examination according to the MoE decision-makers (RQ2).
- The apparent washback on teachers' classroom practices (RQ3).
- The intensity of the apparent washback effect vis-à-vis teacher factors (RQ4).

5.2 Teachers' Perceptions of the New Examination

This section provides results that pertain to the first research question: "What is the nature and scope of the washback effect on teachers' perceptions of the new English elective diploma examination?" This question examined teachers' appreciation and understanding of the intended washback (as set out by the MoE decision-makers) from the new examination on classroom teaching. The overall results of the teacher questionnaire are reported first, followed by the data gathered through interviews. Both are reported in this section using the specific themes stemming from the research aims and questions, as discussed below.

In addressing the first research question, the teachers' perceptions of and reactions to the new examination were analysed by conducting a descriptive analysis of their responses from the questionnaire. As mentioned in Chapter 4, the key characteristics of the teachers taken into account in this study were as follows:

- Demographic characteristics (gender, academic qualifications, age and teaching experience).
- Current teaching situation (grades currently taught, class size and teaching load) and training courses attended relating to the newly introduced English elective examination for 2018–2019 (see Table 4.1).

As previously noted, the questionnaire respondents comprised teachers nationwide who had 1–2 years' experience of teaching the English elective since its implementation in 2018–2019. The return rate of responses was around 92% (193 out of 209 teachers contacted). The questionnaire data in this section were generated using descriptive statistics, namely frequency distributions and percentages. Content analysis was applied to the open-ended statements, based on the categories used for the analysis of the quantitative data.

As most of the questionnaire items were designed with responses given on a five-point Likert scale, expressing agreement (see Appendix D), the teachers' attitudes are expressed as frequency distributions and percentages for Parts 2, 3 and 4 (see 4.7.1). For clarity and simplicity in presenting the results, an abbreviated notation is used to refer to the teachers' questionnaire: TQ1.1 refers to Part 1, item 1, and so on. The results are reported according to the following themes:

- Teachers' reactions to the new examination.
- Teachers' perceptions of the rationale and format of the new examination.
- Teachers' perceptions of the added workload/pressure and the possible challenges involved in teaching the new English elective coursebook.
- Teachers' perceptions of the new English elective curriculum with regard to teaching methods, activities, materials, aspects of lesson planning and the use of mock examinations.
- Teachers' attitudes towards their assessment practices.

The data from the teachers' interviews were analysed using the process of thematic analysis outlined by Braun and Clarke (2006). This approach was adopted as it would complement the data from the questionnaire and observation, helping to interpret the meaning of the interview data from two perspectives: first, from the insights obtained when analysing the questionnaire in phase two, checking whether the data provided were sufficient and answered the research question; second, from a data-driven perspective based on inductive coding. A complete list of the codes and themes is provided in Appendix L. The following section provides the findings related to RQ1.

5.2.1 Teachers' reactions to the new examination

In TQ3.1, the teachers demonstrated their reactions to the new elective exam. The results are presented in Table 5.1.

Table 5.1. Teachers' reactions to the new English elective diploma examination

Item	Response	Frequency	Percentages
It needs further development	Yes	97	50.3%
	No	96	49.7%
I endorse the change	Yes	40	20.7%
	No	153	79.3%
I am sceptical about the need for change	Yes	23	11.9%
	No	170	88.1%
It influences my decisions about which language skills are more important to teach	Yes	64	33.2%
	No	129	66.8%

There was a substantial level of agreement in the teachers' reactions to the new examination. The majority (88.1%) disagreed that they were *sceptical about the need for change* and only a minority (11.9%) agreed with this statement. The second largest percentage (79.3%) represented the teachers who did not *endorse the change*, with only a few (20%) agreeing with it. In total, 66.8% of the teachers did not feel that the new examination *influenced their decision about which language skill was more important to be taught*, while 33.2% of the teachers held the opposite view. Only on the issue of whether the exam *needs further development* was there significant disagreement, with 50.3% of the teachers feeling that it did need further change and 49.7% disagreeing.

The open-ended statements showed that almost all the respondents (21 out of 23) had negative opinions about the content of the test and were in favour of the examination specification undergoing further development, especially in relation to reading and writing. The teachers attributed their negative attitudes to the pressure on their workload caused by the changes to the examination, such as the requirement to teach a new syllabus. They felt adjustments were necessary for their classroom methods to prepare students adequately for the upcoming examination and they recognised the impact this would have on both the students and themselves. Only two teachers held positive attitudes towards the examination change, with one noting that "it is very inspiring and useful" and the other noting that "it is useful for providing feedback".

The negative overall attitudes of the teachers concerning the new examination design were further reflected in the responses elicited from the interviews conducted with the teachers. All five teachers expressed their worry and concern when the exam was initially introduced. They attributed these feelings to the complexity of the new examination and the content of the textbooks for their students. T2 and T5 mentioned that one possible reason for the difficulty of the new examination was that the new textbooks contained too much material, with particular lessons requiring too much time. The teachers reported that teaching the recently introduced *English Insight 2* curriculum in just four weekly lessons per semester was challenging given the extra focus on grammar, vocabulary, listening and speaking exercises. They felt that additional time could be used to help their students prepare for the final examination. In addition to the extra materials in the textbook, the teachers were required to assess the students three times per semester as part of the continuous assessment scheme of work. However, there was no increase in the time available for teaching to accommodate the changes to the materials and the examination. The following excerpt clearly illustrates this issue:

I was really concerned because the book we have to teach is very intensive and there has been a new skill added to the exam – which is listening – compared to the old one. The three skills assessed in the exam were a concern to me as my students always reflect their fears regarding the difficulty level of the new questions in the final exam, but I always remind them that it was their choice to select this subject and they have to like it. (T5)

T1 made the point that the teachers had hoped for a change in the English elective subject. They believed that the change was necessary, as it had been a long time since there had been any new content or features. However, this teacher complained that many teachers lacked access to sufficient supporting materials for certain aspects of the exam, especially in the first year of its implementation. The teacher also reported that they were not informed about the changes beforehand and were not given the opportunity to express their attitudes and opinions about the form the new exam should take.

We were very surprised when this change was announced because we had not expected the change to be very significant, like most other previous changes. Although we had called for a change in the elective subject, we had concerns and fears about how to implement the required change. When we attended the orientation course, most teachers expressed their concerns and fears, especially that this change was introduced in the system without involving teachers in the decision. We felt that it is up to the MoE members to direct examination change. We had nothing to do with this. (T1)

T3 shared the same concerns, reporting that teachers had doubts and confused ideas about the new exam because there was a delay in providing them with sufficient training opportunities. There was not enough awareness about how to teach the new elements in the curriculum. Teachers did not know as much as they needed to about the examination content and its demands. For example, the demands of teaching the grammar content were deemed a challenge for teachers and students alike, as seen in the following comment:

We had fears and concerns about this new change, and we did not know how to deal with it because there was a delay in preparing teachers for the change. The grammar elements in the new curriculum are not only challenging for the students but also for the teachers. Some of the grammar elements in the textbook are worth questioning as to the reason for including them at this point or grade level. They would confuse students rather than help them to use the language. (Teacher 3).

However, even if the teachers' views of the new examination and textbooks were predominantly negative, this does not necessarily mean that they would not change their teaching strategies in the classroom. These negative attitudes and concerns regarding the new English elective emerged during the first year of the implementation of the new examination and changed after the teachers had received specific official publications and frequent training on how to implement the new curriculum. T4 recounted that the

information and examination guides helped them understand what they needed to alter in their teaching practices in the classroom. Furthermore, the new materials and examination were having a positive effect on the way teachers taught compared to previously (T3) and the change was felt to suit the students' level of English proficiency (T1).

Sometime after this change, specifically after the first semester, we attended an orientation course on how to handle this change in the elective diploma. So, we were more relaxed, and we knew how we were supposed to handle it. We now follow what the ministry officials instructed us to do. Specifically, when we saw the new exam specification, we understood what we are required to teach. (T4)

At first, I was having some doubts and fears, but later I really liked this change. So, this change has a positive effect on my way of teaching; when the change came, it took us outside of our "comfort zone". (T3)

After the first semester of the academic year, we felt that this change was positive. Students also adjusted to it, especially when we noticed that this change suited their abilities and overall performance. (T1)

Looking at the results of the questionnaire and the interviews, it is interesting to note the discrepancy in teachers' reactions to the new examination. For example, some teachers believed there was a need to develop the examination further, whereas others had positive attitudes to the change (see Table 5.1) This might relate to the stage at which teachers came across challenges and difficulties arising from implementing the demands and requirements of the new English elective curriculum. This discrepancy may also relate to the little input from teachers about how the teachers in Oman perceived and evaluated the revised versions of the test and thus it remains unclear if the new testing demands and content might be likely to deliver the intended improvement in grade 12 English language teaching and learning. Further discussion of the teachers' instructional practices is presented in the section on classroom observation (see 5.4).

5.2.2 Reasons for change

In Chapter 2 (see 2.2.4), the rationale and objectives underpinning the new examination were outlined. TQ2.A was developed with these in mind, aimed at eliciting teachers' views on the need for the new examination and their understanding of the intentions behind the change. The teachers were asked what they thought were the main reasons for developing the new examination. Their responses were ranked according to the mean scores on a Likert scale of agreement, as shown in Table 5.2.

Table 5.2. The reasons for the examination change as perceived by teachers

Item	Strongly Disagree N (%)	Disagree N (%)	Undecided N (%)	Agree N (%)	Strongly Agree N (%)	Rank	M	SD
To meet the demands of higher education	14 (7.3%)	9 (4.7%)	12 (6.2%)	98 (50.8%)	60 (31.1%)	1	3.94	1.102
To improve students' communicative ability in English	14 (7.3%)	15 (7.8%)	10 (5.2%)	94 (48.7%)	60 (31.1%)	2	3.89	1.149
To use a communicative approach to language learning	9 (4.7%)	10 (5.2%)	19 (9.8%)	110 (57.0%)	45 (23.3%)	3	3.89	0.976
To prepare students for working in an English-speaking academic environment	13 (6.7%)	12 (6.2%)	14 (7.3%)	100 (51.8%)	54 (28.0%)	4	3.88	1.095
To cope with the present decline in English standards	8 (4.1%)	26 (13.5%)	26 (13.5%)	105 (54.4%)	28 (14.5%)	5	3.62	1.025

As can be seen from Table 5.2, the reason for the examination change accorded the most importance by the teachers was *to meet the demands of higher education*. These demands were listed previously (see 2.2) and included achieving the language proficiency level required by higher education institutions. Three items came out with very similar levels of importance: *to improve students' communicative ability in English*, *to use a communicative approach to language learning*, and *to prepare students for working in an English-speaking academic environment*. In addition, most of the teachers felt that the new examination was being implemented *to cope with the present decline in English standards*.

In the space for open-ended responses provided in TQ2.A, half of the teachers (25 out of 50) suggested that preparing students for international exams, such as IELTS and TOFEL, was the most important reason for the change. Moreover, 15 teachers felt that the rationale for the change was to improve specific skills, such as writing and reading, as well as to broaden students' vocabulary. These teachers believed that the emphasis on a wider range of language skills (rather than on all language skills in the new exam: reading, writing, listening, vocabulary) could be another main reason for the change, as the previous exam had only involved reading and writing skills. In all, 10 teachers believed that the examination was introduced to develop higher-order thinking skills and encourage students to be confident in using the language. One of those teachers noted that the purpose of the examination was for

students “to be well-prepared for the future, both for life and work, or for going on to higher academic studies in this country or abroad”.

In the interviews, five teachers mentioned two of the statements from the questionnaire as key reasons for introducing the new examination. Four of the teachers agreed that one of the main reasons for the new examination was to *meet the demands of higher education*. For example, T1 and T3 reported that the new form of the questions helped students cope with the demands and content of the IELTS exam, which is deemed one of the essential requirements for admission to higher education institutions in Oman:

The exam helps students cope with international exams, especially as the reading, writing, and listening skills in the elective exam are similar to the IELTS exam questions. For example, the reading questions now contain analytical skills and indirect questions for assessing students’ comprehension, and also the writing skill involves writing a report about visual diagrams. (T1)

What’s the use of teaching advanced level English as an elective subject if we do not prepare students for these exams? Our role as teachers is to prepare students for life and not just the exam. ... Although it is risky for students to take this new elective as they might get B or C grades, they take it because it prepares them for what they face in the future and for international exams. From my experience, I see that students see that. (T3)

Only T2 mentioned that the new exam was required to *improve students’ communicative ability in English*:

We need to provide our students with 21st-century skills. For example, the kind of writing students have in the new exam includes analysing and comparing charts and graphs, which is very important as a skill for their life and work after school. (T2)

Various other reasons for the introduction of the new examination besides the options given above were also suggested. For example, T2 mentioned that the new examination was having a positive effect on teaching and learning in class and felt that the new content relating to writing skills would help teachers develop their teaching:

Teachers have also developed some skills when teaching this new elective, as they need to be aware of everything included in this change. So, in teaching report writing, I need to be aware of all the types of report, and then I can give my students the gist when teaching them these types. It is a really nice experience, and we are learning many new things. (T2)

In T5’s opinion, the new examination was a natural development from the new textbooks, as it aligned “what is taught in terms of skills and knowledge in the textbook with what is assessed in the exam”.

The teachers felt that the most important reason for the introduction of the new examination was *to meet the demands of higher education*, thus indicating that they implicitly agreed with the principles of the intended washback effect (see 5.3). The other different reasons mentioned by the teachers in both the questionnaire and interviews showed that many acknowledged the importance of the new examination in the Omani education system and the positive effects it was having on their teaching practices. All the teachers generally believed that the new examination had the potential to change teaching and learning for the better.

5.2.3 New examination format

The responses given regarding the changes to the exam specifications for the elective diploma, as compared to the previous examination are shown in Table 5.3.

Table 5.3. Perceptions of changes in the English elective examination specifications

Item	Strongly Disagree N (%)	Disagree N (%)	Undecided N (%)	Agree N (%)	Strongly Agree N (%)	Rank	M	SD
More emphasis on essay writing	6 (3.1%)	13 (6.7%)	10 (5.2%)	104 (53.9%)	60 (31.1%)	1	4.03	0.957
More emphasis on listening skills	5 (2.6%)	22 (11.4%)	16 (8.3%)	117 (60.6%)	33 (17.1%)	2	3.78	0.949
More use of authentic tasks	5 (2.6%)	12 (6.2%)	36 (18.7%)	110 (57.0%)	30 (15.5%)	3	3.77	0.879
More emphasis on reading skills	9 (4.7%)	19 (9.8%)	16 (8.3%)	112 (58.0%)	37 (19.2%)	4	3.77	1.021
More closely related to the objectives of the elective grade 12 curriculum, English Insights 3	12 (6.2%)	17 (8.8%)	28 (14.5%)	110 (57.0%)	26 (13.5%)	5	3.63	1.029
More emphasis on grammatical usage	18 (9.3%)	52 (26.9%)	22 (11.4%)	74 (38.3%)	27 (14.0%)	6	3.21	1.245

The teachers' perceptions of the items mentioned in Table 5.3 were relatively consistent, with most teachers selecting the *Agree* option for all the statements provided. This consistency suggests that the teachers' perceptions of the new English elective were broadly in line with the intentions behind the changes. However, it is interesting to note another discrepancy in teachers' attitudes between the responses given in the table and the findings discussed above in relation to Table 5.1, which relates to teachers' reactions to the new examination change. The differences in teachers' attitudes and reactions towards the new exam clearly suggest a reluctance to make any of the changes in their actual teaching methods that were intended by the MoE decision-makers.

Similar findings were reported by teachers through the open-ended statement in the questionnaire provided in TQ2.B. Most respondents provided additional answers about the changes in the elective diploma exam compared with the old exam. Around 30 teachers stated that the new exam involves more integration of all the different language skills and uses more real-life tasks. However, 10 teachers reflected their concerns regarding the time needed to teach the new curriculum and the level of difficulty of the reading, listening, and writing tasks outlined in the new exam specifications.

The teacher interviews also revealed similar findings, particularly when the teachers were asked for their views on the new test format, where they all provided similar responses. There were a number of comments from all five of the teachers regarding the ways the new elective exam differed from the previous one. These could be divided into two main areas: content and techniques. The first set of comments concerned the content of the exam. All the teachers interviewed believed that the new exam tested different skills (including vocabulary, grammar, and listening) from the old exam. They mentioned that the old exam had mainly tested writing and reading. However, Teacher 1 pointed out that the grammar and vocabulary questions had been deleted from the elective exam paper after the first year of its implementation:

The final exam assesses reading, listening, and writing skills, but doesn't assess grammar, vocabulary, or speaking. In the first year of implementation of the exam (2018–2019), grammar and vocabulary were assessed in the final exam; but then students, even those who are at advanced level, and teachers complained that these parts of the language learning were too confusing and demanding. When it came to questions in the exam, for example, many alternative answers were deemed correct for the same grammar question. (T1)

The teachers mentioned that they were required to teach the grammar and vocabulary tasks in the textbook and that some of these were not covered at all in earlier grade levels. They were also directed by the MoE decision-makers to set some of the grammar and vocabulary tasks as homework for students, especially those tasks repeated from the core textbook or lower grade levels. T4 reported that “this decision aligned with the overall goal of the examination change, as we are supposed to focus on language teaching rather than on analysis of the language”.

Moreover, it was noted that the new examination contained more types of academic writing, such as reports, comparisons, applications and descriptions of a chart (for sample items, see Appendix A). These types of text were newly introduced in the curriculum and had not been studied at previous grade levels. The old examination tested common types of writing that

students had learnt in the lower grades. Another teacher made the point that the new examination prepared students for academic study and international test standards, such as making notes, mind maps, dictionary skills, using visuals and summarising information. As one teacher put it:

The new exam enables students to achieve a high English language standard and equips them for international exams and higher education institutions, especially in terms of the inclusion of academic writing. The listening skill is newly added, which is a very important skill for students to improve their level of English. (T1)

According to another teacher:

The old exam focused on developing two main skills, which were reading and writing. The exam assessed very basic and simplified reading texts besides writing tasks and some basic language tasks for grammar and vocabulary. As for the new exam, it involves a variety of language skills, focusing on four main skills: reading, writing, speaking, and listening, including grammar and vocabulary. So, the new exam is more communication-oriented than its predecessor. (T4)

The second set of comments concerned the type of questions that were covered in the old and new examinations. The items aimed at testing reading and writing contained the same types of questions introduced in previous grade levels, but the new examination also included new types of questions. As T2 remarked, “The listening tasks, for example, involve more authentic and less predictable tasks than those in the previous exam”. The teachers felt that the old examination addressed the different skills separately, while the new one required more integration of the skills: “some new types of questions in the new exam require integration between more than one skill compared to the previous exam, specifically the listening skill, which takes in reading, writing, grammar, and vocabulary skills” (T3). The writing tasks in the new examination encouraged greater thinking and planning than those in the old version, with students being given more choice about what to write:

The new exam requires more thinking and planning, for example, brainstorming ideas for writing topics. The previous exam encouraged students to reproduce memorised essays or short pieces of writing. (T1)

Some teachers commented on the individual skills required by the new examination: students had “to dig deep and think to find the answer for the reading component in the exam, such as matching the texts with the appropriate paragraph” (T5) and they had to write more than for the previous examination (T1). Another teacher mentioned that the new questions related to the listening component targeted advanced-level students: “the new types of listening questions were more challenging than the previous exam. This challenge was more suitable for elective diploma students, who have a high level of English proficiency” (T3).

To summarise, the teachers expressed positive feelings about the new examination, seeming to see it as complementary to the *English Insights 3* textbook. They mentioned several features that illustrated how the new examination differed from the previous version. They welcomed most of these features (new content and techniques) and saw them as changes for the better.

5.2.4 Extra pressure placed on teaching by the new examination

Teachers were asked about the kind of pressures they thought the introduction of the new examination would place on their teaching. Their views are presented in Table 5.4.

Table 5.4. Extra pressures on teaching from the examination

Item	Strongly Disagree N (%)	Disagree N (%)	Undecided N (%)	Agree N (%)	Strongly Agree N (%)	Rank	M	SD
Preparing more authentic tasks	9 (4.7%)	16 (8.3%)	27 (14.0%)	108 (56.0%)	33 (17.1%)	1	3.73	0.996
Arranging more assessment-related activities	13 (6.7%)	23 (11.9%)	16 (8.3%)	93 (48.2%)	48 (24.9%)	2	3.73	1.16
Providing more supplementary teaching materials	6 (3.1%)	33 (17.1%)	27 (14.0%)	92 (47.7%)	35 (18.1%)	3	3.61	1.066
Adopting new teaching strategies	6 (3.1%)	37 (19.2%)	25 (13.0%)	97 (50.3%)	28 (14.5%)	4	3.54	1.056
Teaching a new syllabus: English Insights 3	9 (4.7%)	38 (19.7%)	26 (13%)	84 (43.6%)	36 (19%)	5	3.52	1.142

The teachers expected that the changes would increase their workload and introduce new teaching challenges. They also exhibited anxiety over the examination change. The teachers indicated that *preparing more authentic tasks*, *arranging more assessment-related activities* and *providing more supplementary teaching materials* would place the greatest demands on them. This finding reflected the results regarding teacher’s attitudes towards authentic tasks, with most teachers highlighting the greater use of these in the new examination. The other main worries expressed by teachers were that the change in the examination would result in the need for *adopting new teaching strategies* and *teaching a new syllabus*.

The open-ended section provided in TQ2.C demonstrated that many teachers had serious concerns about teaching the new syllabus. Indeed, half of the respondents (15 out of 30) agreed that the time needed to teach the new syllabus imposed extra pressure on teachers. For example, one respondent mentioned that “the time factor was a pressure point, as books were received very late, and we had so many tasks to cover”. As discussed in Chapter 2 (see

2.2.2), all schools throughout the Sultanate use the same textbooks and they are required to complete the curriculum objectives on time, with the textbook forming the main source of input. There were certain procedures conducted by the MoE stakeholders when designing and implementing the new curriculum that required more time and effort than the procedures for preparing the distribution of the previous curriculum to schools.

Ten teachers also felt that the exam forced them to prepare extra teaching materials for students to match the difficulty level of the new syllabus compared to the standard level of students, thus requiring them to spend much more effort and time on preparation. For example, one teacher stated that ‘a lot of effort needs to be paid to preparing for the classes, and the reading topics are full of challenging vocabulary, whereas the grammar is not important and confusing’. However, three teachers were positive about teaching the new syllabus.

The above findings from the open-ended statement in the questionnaire were reflected in the responses elicited from the interviews conducted with the five teachers, in particular in relation to the level of pressure teachers believed the new examination would place on their teaching. They showed great concern over the difficulty of teaching the English Insights 3 textbook series, which follows the same format as the new examination and is regarded as the main source for teaching the new English elective syllabus (see Chapter 2). One area of extra work and pressure highlighted by the teachers related to the difficulty of teaching the new reading and writing skills. The teachers believed that they needed to explain all the details in the reading texts and felt that this would help students “to find the information that was required in the reading tasks” (T5). Additionally, the teachers mentioned that they had to produce extra materials to use alongside the reading texts included in the book. This was because some reading topics in the textbook required “background knowledge” (T2). T3 said that teaching the unfamiliar reading topics in the textbook made reading difficult as these required a lot of planning. For example:

We have a reading text about “fooling the public”. It was about how the BBC fooled the public with different stories on April 1st. I introduced this topic by showing them a video about flying penguins from a BBC news report. We had a discussion because students were confused about this fact and later, I showed them another video about how the BBC tricked people over this. So, I used the time to introduce the meaning of difficult terms and draw their attention to the background of the topic. (T3)

The types of writing introduced in the textbook were also regarded as quite new and difficult. The textbook included many types of writing. For example, T2 reported “I now usually have to set aside 2–4 lessons to teach the writing tasks in the textbook”. Another teacher stated:

I feel more pressure when I am teaching writing because we do not have enough resources for teaching this skill. This skill takes a lot of time and effort because I need to teach all the types of writing for the final exam, as well as the IELTS writing tasks. I have a big workload. (T1)

Another key concern among some teacher participants was the issue of insufficient teaching time. T2 commented: “We have four lessons per week, the time given to teaching this subject is not enough to teach the topics involved in the textbook”. Other teachers reported similar challenges:

With the previous curriculum, we had sufficient time to teach it in relation to the content. For this new curriculum, we have the content, but we need to utilize this content to present it in the best way. This new change took a long time to sort out. I really had a nice solid approach in place. (Teacher 4)

My concern is about the size of the new curriculum. So, my focus is on how to be selective in teaching this new elective curriculum, especially as it contains many more topics and skills than the previous curriculum. (Teacher 1)

To sum up, it is clear that many teachers were greatly concerned about the methods of teaching required by the new syllabus, geared to the examination. This was because they had to cover the textbook material in a very detailed way. Also, some teachers faced difficulties in relation to the type and the amount of teaching materials they were having to cover. These issues slowed down the teaching process and put more pressure on teachers in terms of planning.

5.2.5 Challenges in teaching

The teachers responded to seven items related to the most challenging aspects of teaching and preparing students for the new examination, as set out in Table 5.5. As can be seen, the students' level of English was felt to pose the greatest challenge to teaching in relation to the new examination. The teachers were also asked to consider six other areas of potential concern: volume of the textbooks, workload, lack of time and learning resources, learning environment and class size. One point to be emphasised here is that class size was perceived by teachers to be the least challenging factor in this context. This could be due to the logistical aspects of teaching English in Omani post-basic schools, in which the average class size is 15–20 students. Such class sizes are also due to the status of English as an elective subject, with some schools having only 2–10 students. One teacher in the open-ended statement under this category mentioned that “sometimes we have small classes of no more than two students, which might hinder some activities requiring group work”.

Table 5.5. Challenges in teaching the new English elective curriculum

Item	Strongly Disagree N (%)	Disagree N (%)	Undecided N (%)	Agree N (%)	Strongly Agree N (%)	Rank	M	SD
Students' current English level	8 (4.1%)	18 (9.3%)	10 (5.2%)	85 (44.0%)	72 (37.3%)	1	4.01	1.085
Volume of the textbooks	4 (2.1%)	29 (15.0%)	23 (11.9%)	86 (44.6%)	51 (26.4%)	2	3.78	1.063
Heavy workload (too many classes per week, arranging online materials, etc.)	8 (4.1%)	33 (17.1%)	21 (10.9%)	82 (42.5%)	49 (25.4%)	3	3.68	1.15
Inadequate lesson time	5 (2.6%)	41 (21.2%)	32 (16.6%)	73 (37.8%)	42 (21.8%)	4	3.55	1.127
Lack of teaching and learning resources	7 (3.6%)	45 (23.3%)	20 (10.4%)	84 (43.5%)	37 (19.2%)	5	3.51	1.151
Learning environment	8 (4.1%)	41 (21.2%)	29 (15.0%)	82 (42.5%)	33 (17.1%)	6	3.47	1.128
Class size	18 (9.3%)	51 (26.4%)	22 (11.4%)	69 (35.8%)	33 (17.1%)	7	3.25	1.275

Similarly, in the open-ended statement provided in TQ2.D, 15 teachers believed that their students' current English level to be one of the main obstacles in teaching the new curriculum. It is also worth noting that 10 teachers perceived the lack of sufficient real-life tasks and authentic materials in the textbook to be another primary concern.

In the interviews with teachers, the students' proficiency was considered one of the most challenging factors in teaching the new curriculum. The teachers felt that the students who selected this subject usually exhibited the necessary motivation and interest to improve their language skills. However, they also talked about the difficulties they had in teaching students with a high level of English proficiency:

When it comes to teaching, the only challenge is that you should have in mind that you have outstanding students, and you need to impress them to catch their attention. I need to hook my student's attention in order to keep them with me. Because, if they get bored, I will lose them. I have to give the students something that really interests them. To achieve this, I have to bring them something at the start of the lesson to make them think, or something vague or mysterious that is connected to my teaching. (T2)

I used to discuss with my students which topic they prefer to learn in my lessons. For example, we have a pyramid about the basic needs of human beings, so this is connected to students' life and attitudes. Also, students at this stage are

knowledgeable and they have their own free reading on different topics. So, I need to connect their knowledge with what I am planning to teach. (T1)

Another serious challenge mentioned in the teacher interviews was the lack of support and training opportunities for teachers. As discussed in Chapter 2 (see 2.2.5), the training of English teachers is supervised and provided by the main training centre in the MoE. Some teachers are nominated for the training conducted by the MoE while others are trained in their regions by their supervisors or their colleagues in their schools. In this regard, all five of the teachers interviewed mentioned that they had not received any training on this new examination. One teacher explained the process of cascading that the MoE decision-makers had used to introduce the new examination. T1 indicated that there had only been a few training opportunities, headed by the MoE decision-makers and members of the publishing company, where teachers were able to discuss the intentions behind the change. Other teachers received information and directions from their supervisors. However, some teachers were only informed about the training by school colleagues who had undertaken the training:

When the MoE announced the new curriculum for the elective diploma, they met with supervisors and trainers in the directorates and provided them with an orientation course, and then after this, these officials cascaded this training to the teachers who are teaching this subject. This orientation is offered annually for teachers, supervisors, and trainers. (T1)

Similarly, other teachers commented in negative terms about the training opportunities offered to those teaching the new English elective. As one remarked:

I would say that we did not get any training programme on teaching strategies for this new curriculum, and even today teachers are not well-prepared or qualified for this change. They only provided us with orientation about this change which was very late, and so not very useful. (T5)

From the above results, it can be argued that the findings regarding the key challenges involved in teaching the revised English elective further explain teachers' reactions to the new examination (see 5.2.1). The most obvious problem was the high level of English language proficiency of the students. Teachers faced great difficulties in preparing teaching materials to reflect the standard of the students in the English elective classes. Moreover, the cascading training process used in this context, with its significant variation in the training opportunities offered to different teachers, meant that there was no guarantee that the same messages about the change reached all teachers in different parts of the country.

5.2.6 New English elective examination design

The purpose of seeking qualitative data in TQ4.1 and TQ4.2 was to generate a wide range of responses, beyond the closed questions and from as many respondents as possible, particularly as the questionnaire was distributed to all teachers who had been teaching the revised English elective over the first two years of its implementation (2018–2019). This section presents the second category of data (TQ4.2) obtained from the teacher questionnaire, as well as further interview findings. In this category, teachers were asked to give their views concerning the format and quality of the English elective examination.

The findings from the open-ended question TQ4.2 showed that teachers had significant concerns about the new exam. Most teachers (86 out of 158) agreed that the examination was an accurate measure of a student's ability to apply the language in real-world contexts. Others noted that the examination results for each language ability allowed them to focus on particular skills that needed more attention. However, 72 teachers found the examination to be inappropriate in some way, largely because it included language skills that differed from the previous examination, such as listening and speaking, and because it tested for overly high levels of academic skills with regard to reading and writing. Teachers who held a negative view of the examination did not find it to be an improvement on the previous version in terms of design and structure. According to these teachers, teaching the new curriculum was more challenging than before. For example, they criticised the content validity of the new examination as speaking skills were not assessed and the questions for reading skills were too difficult for the students.

In the interviews, all five teachers identified specific positive features of the new examination but did not mention any negative effects, confirming their responses when asked about the need for examination change (see 5.2.2). The teachers spoke positively about the design of the new examination because it tested additional skills, rather than just reading and writing (T3 and T4), and it was not mainly based on recognition but challenged the students to use communicative tasks and thinking strategies to a greater degree (T2):

Teaching the previous exam curriculum was very boring, as the outcomes were only based on reading and writing, even when teachers were using a variety of supplementary materials. However, the new exam challenges students more effectively and encourages them to use the language more communicatively, especially with the addition of the listening tasks. (T2)

Another feature that was highlighted was the relevance of the textbook content to the new examination specifications. T1 reported that the textbook included sufficient teaching

material to help teachers prepare their students for the new examination without the need to provide extra materials. This was useful for identifying students' level of proficiency in English and providing them with information about their strengths and weaknesses:

The change in the examination is aligned well with the new textbook, as the skills involved in the curriculum are measured in the exam. I believe that this new exam matches the required skills and abilities outlined in the curriculum. The content and assessment are more aligned with the students' actual level, whether they are A or B students. To me, I think when using this exam, I can differentiate between A+ students or B-level students. (T1)

Another aspect mentioned by three of the teachers was the positive effect of the new examination on the lessons taking place in their classrooms, discussed further in the section on the findings from the classroom observations (see 5.4). T3 noted that the “unseen” reading texts were “a new feature that gave us the idea that we should be teaching reading as a skill rather than the specific reading topics”. T3 noted that the unseen passages promoted the “communicative approach which underpinned the new examination design”. T5 expressed the view that these texts would help students develop their reading habits, especially those who had not yet mastered specific reading skills:

Reading skills are assessed in the new examination context by using “unseen” texts that are sometimes not even related to the topics of the textbooks. This helps student to develop their reading ability, especially those students who really need to further develop their reading skills, such as skimming and scanning, and improve their language outside the classroom. (T5)

However, the teachers did mention two main areas for possible improvement. The first of these was the testing of grammar and vocabulary. T3 explained that grammar and vocabulary questions were deleted during the first year of implementation, but only from the examination paper, not from the textbook. Teachers were told that they should teach these skills by integrating them with the other main skills. They should not skip these skills when teaching because they are important for students' reading and writing. However, T3 argued that grammar and vocabulary should be tested as this would have a positive effect in the classroom as these two skills were still included in the textbook:

So, if we include grammar and vocabulary in the new exam as the first year of its implementation. It would be fair to say that the skills tested in the exam are what attract students' attention during classroom teaching, and they focus on these more than those not included in the exam. (T3)

The second suggestion related to the testing of listening and writing skills. T1 argued for different options in the section on writing, as this would help students “predict the type of

writing text that would appear in the exam”. T2 argued for having a separate paper for the listening question as this would “give students more time for the reading and writing questions in the exam”.

In summary, the new English elective examination was felt to have advantages over the previous version in that it tested more skills and tasks (not just reading and writing), it was better aligned with the teaching activities in the textbook, it encouraged communicative tasks and thinking strategies and it had a positive influence on the classroom. However, several recommendations were made, including expanding the examination to test grammar and vocabulary more directly and giving students more choice regarding the writing tasks they could respond to and the type of paper they sat for the examination.

5.2.7 Decision-making concerning the new examination

TQ3.4 asked teachers for their views on who they felt had made the major decisions about the specifications for the new examination, offering them four options to choose from. The options were based on the formal procedures used by the MOE when introducing a new curriculum (see 2.2.4). The results are presented in Table 5.6.

Table 5.6. Teachers’ perceptions of decision-making concerning the new examination

Item	Response	Frequency	Percentage
Grade 12 English teachers together	Yes	37	19.2%
	No	156	80.8%
English-language technical staff at the Ministry of Education (curriculum developers, exam writers, etc.)	Yes	110	57.0%
	No	83	43.0%
Members of higher education institutions	Yes	28	14.5%
	No	165	85.5%
Decision-making directors from the Ministry of Education (e.g. Director of Examination and Test Administration)	Yes	94	48.7%
	No	99	51.3%

It seems that there was no consensus among the teachers about who decided the examination specifications for the new English elective. The highest support here was for English-language stakeholders (such as exam writers, curriculum writers, supervisors, etc.) and decision-making directors from the MoE, at 57% and 48% respectively. It is interesting to note that a few teachers (19.2%) reported that they had been included in decisions relating to the design of the new specifications to be used in their own teaching, whereas others specifically mentioned that they were not included in the process. As one teacher commented, “I do not have any idea who made the changes, but I do not think it was the teachers’ decision”. It is to be expected that teachers will be involved in such decision-

making regarding examination changes as otherwise they might not be able to implement the changes in their work.

The five teachers who were interviewed reported similar observations, especially when questioned about the procedure for introducing the new examination. As discussed in Chapter 2, the first cohort of students was taught using the new textbooks in 2018/2019. The format of the examination itself was introduced shortly after. The official documentation of the new examination specification that teachers received was a 60-page document, the “*Student Assessment Handbook of English Language: Grades 11–12*”. There were three phases after introducing this change. In the first phase, curriculum development members conducted a series of visits to meet and talk with staff members who were responsible for administering and following up on this change in schools; approximately 70–80% of governorates participated in this phase, including those located in distant parts of Oman. The second phase sought to elicit students’, teachers’ and school principals’ attitudes and feedback concerning this change. This was conducted by interviewing these stakeholders and observing classroom teaching. These two phases were completed at the end of the first year (2018/2019). In the second year, a report on the first phase (stakeholder feedback) was discussed with the publishing company, supervisors and training and assessment team. However, teachers were not involved in this stage of discussion.

The MoE participants gathered stakeholder feedback in two ways: one source was written feedback sent to the MoE supervisory members and the other was school visits conducted by a joint team from the MoE curriculum development department and the publishing company. What my discussions with the MoE participants also included was whether any kind of follow-up research was planned to determine the effects of the examination on teaching. However, the participants believed that three years of implementation was too short to examine whether the intended effects of the change had manifested.

Two months after teaching of the new English elective began, the MoE decision-makers organised a series of meetings and seminars to help teachers implement the revised syllabus:

- Seminars were organised and attended by the curriculum development members, assessment team, the teacher training central team in the MoE and the publishing company’s representatives. The seminar attendees included teachers, senior supervisors and stakeholders from different governorates, such as regional directorate officials and parents. The aim of these seminars was to discuss the change, the rationale, the content

of the textbook and the new examination format. Issues raised by teachers and governorate representatives, including their concerns and recommendations about the change, were discussed. There were lengthy discussions between the teachers about the textbook content and how they would adjust their teaching according to the requirements of the new examination.

- At the micro level, seminars and workshops were organised by senior supervisors in each governorate. These meetings had the largest number of teacher participants, with one teacher selected from each school. Supervisors shared their understanding of the proposed changes as discussed in the previous seminar at the ministry level. They provided examples of how the activities and tasks in the textbook could be carried out in the classroom and demonstrated, in some detail, the teaching plan devoted to the proposed changes.

When the teachers were interviewed, however, they were very negative about the way the new examination specification had been introduced to them. The five teachers explained that they were not involved in decisions regarding its design or form and only heard about it a short time before it was due to be implemented, which made it difficult for them to get to grips with the new textbook, the continuous assessment element and the exam specifications:

We did not give our opinions before the implementation of this change. In the first year of the exam implementation, we found it difficult because it was new to us, and the exam specification also came late. (T5)

I didn't have any idea about it. MoE officials decided the things that should be in the book and what should not be there. I am not sure what should be in the exam or not. I think this is worrying, as it took me a long time to plan my lesson to suit my students' needs as a result of the changes in the classroom assessment, teaching content, and the new exam. (T4)

Moreover, the five teachers revealed that they did not participate in the decision-making committee at the MoE level, or even in the workshops that the MoE members prepared to introduce the examination. This meant they were unfamiliar with the objectives of the change:

Of course [big laugh], it came from the MoE decision-makers, specifically the curriculum officers. The MoE officials just informed us about the change and asked us to accept it the way they had designed it. It was difficult to prepare our students properly for this new change. (T3)

I don't have any idea, actually, but I do not think teachers were involved in this process. When they announced a new curriculum for the elective diploma, they meet with supervisors and trainers in the directorates and provide them with an orientation

course, and so after, they will cascade this training for main teachers who are teaching this subject. We were lost when the exam was firstly implemented. It was a totally new change which demanded a great deal of time and attention. (T1)

Teachers' reactions to the new English elective examination were mostly negative in the first year of implementation. However, their views concerning the reasons for changing the English elective examination reflect the intended washback of the new diploma envisaged by the MoE decision-makers (see 5.3). This suggests that the implementation was relatively successful. In what follows, teachers' attitudes regarding the impact of the new examination on their teaching practices are discussed. The issue of whether teachers would alter their teaching methods to reflect the intended washback of the new exam, as desired by the decision-makers, is discussed further in the observation results (see 5.4).

5.2.8 Washback from the examination on classroom teaching

The results in this section are reported under two categories: (i) perceptions of teaching practices (TQ2.D, TQ3.3, TQ4.1, TQ4.3; (ii) perceptions of classroom assessment practices and evaluation practices (TQ2.F, TQ2.G, TQ3.2).

Changes teachers would like to make to their teaching

TQ2.D asked teachers about the changes they would like to make to their teaching in response to the examination change. The results are set out in Table 5.7.

Most teachers said that they would like to update their teaching approach to reflect the changes in the new specification for the English elective. For example, very many teachers thought they should *emphasise productive skills* (e.g. writing, speaking) in their teaching. Most teachers also agreed on the need to *adopt new teaching methods*, to *teach according to the new English elective diploma specification*, and to *employ more authentic language tasks* in their teaching. Many teachers also said they would like to put more emphasis *on communication practices and receptive skills* in response to the examination change. Thus, the findings indicate that teachers' thoughts about the changes they would like to make in their teaching matched very closely the changes in the examination itself. However, if the examination was to have an influence on the content of their teaching, it would not be to encourage them to pay attention to speaking skills, as these are not tested. This makes it difficult to draw any definite conclusions about the precise influence of the exam in this context.

Table 5.7. Changes teachers would like to make to their teaching

Item	Strongly Disagree N (%)	Disagree N (%)	Undecided N (%)	Agree N (%)	Strongly Agree N (%)	Rank	M	SD
Putting more emphasis on productive skills (e.g. writing, speaking)	3 (1.6%)	7 (3.6%)	6 (3.1%)	117 (60.6)	60 (31.1%)	1	4.16	0.777
Adopting new teaching methods	4 (2.1%)	8 (4.1%)	18 (9.3%)	114 (59%)	49 (25.4%)	2	4.02	0.838
Teaching according to the new test specifications	3 (1.6%)	15 (7.8%)	21 (10.9%)	100 (51.8%)	54 (28.0%)	3	3.97	0.918
Employing more authentic language tasks	5 (2.6%)	11 (5.7%)	20 (10.4)	106 (54.9%)	51 (26.4%)	4	3.97	0.912
Putting more emphasis on communication practices (e.g. group work, debates, role play)	2 (1.0%)	15 (7.8%)	23 (11.9%)	109 (56.8%)	44 (22.8%)	5	3.92	0.866
Putting more emphasis on receptive skills (e.g. reading and listening)	6 (3.1%)	10 (5.2%)	21 (10.9%)	114 (59.1%)	42 (21.8%)	6	3.91	0.9

Similar findings also emerged from the open-ended statement provided in TQ2.D. Half of the teachers (10 out of 20) noted that they would like to focus more on the examination content, especially in relation to the reading, listening and writing tasks. The other 10 teachers reported that the changes they would like to make in response to the new examination would be characterised by a more communicative approach in their teaching and by preparing more materials for their students.

It is interesting to note here a discrepancy in teachers' attitudes between this item and an item in Table 5.1, which relates to their reactions to the new English elective. There was a significant difference in teachers' reactions towards the new elective examination: 33.2% agreed that it influenced their *decision about which language skill is more important to be taught*, while 66.6% disagreed. However, most teachers agreed that they would make use of the new examination changes in their teaching. This discrepancy clearly demonstrates the gap between the teachers' avowed attitudes and what they would do in the classroom, indicating the complex nature of the influence of the examination.

Regarding the interview findings, three teachers reported that to *teach according to the new English elective diploma specification* was a change they had made in their teaching. These teachers generally believed that it was beneficial for students to master the content of the new examination, which mainly assessed reading, writing and listening. This knowledge would be useful if students came across the same skill tested in the final examination, or if they were asked to answer a question about a topic similar to what they had been taught in the classroom:

It is based on students' needs to pass the final exam. I know my students, so I know what they are weak on. I mean what skills they need to obtain. So I focus on the skills that students require for the final exam. (T2)

When we received the new textbook, we looked at how the content and teaching plan were related to the final exam specification. This is because the teaching plan needs to be aligned with the content of the final test. It is very important to achieve this alignment between the content taught and the skills tested in the exam. We should not separate our teaching from the final exam requirements. The final exam specification should direct the teaching plan. (T1)

The main change concerns what we select for teaching. Listening also was added to our teaching plan, so we need to have time for teaching this. I emphasise reading and writing skills in my teaching plan, as they require more time because of the new exam. (T5)

The two other teachers presented different attitudes. T4 suggested a preference for teaching the skills needed by students to master the outcomes that should be achieved at the end of grade 12 (which would involve not only studying the textbook topics but also using other supplementary materials). This was important to meet students' needs and transmit the knowledge and skills they would require going forward:

My focus will be on extended discussions and activating prior knowledge, creating debates and role plays which are helpful for students after grade 12. The topics and language used in the new syllabus will allow me to expand my students' language use to suit the demands of higher education institutions. In the reading skill, for example, I can easily adapt the new materials into my new teaching plan, as there are a variety of new topics. (T4)

I want to prepare students to take international examinations after grade 12, where most higher education institutions require this condition. Our role as teachers is to prepare students for life and not just the exam. (T3)

Changes teachers made to improve learning

Question TQ3.3 comprised four items that asked teachers about the key modifications they had made to enhance students' English language ability. The results are shown in Table 5.8. As can be seen, most teachers (80%) reported they *emphasise the skills which are more likely*

to be tested in the exam. This result is strongly supported by similar findings from the teachers' responses to TQ2.D (see Table 5.7), when they were asked how they would like to change their practice in response to the examination change in terms of *teaching according to the new specifications*. Similarly, 58.2% of the teachers reported that they *do more exam practice*. These findings suggest that teachers understand the need to prepare their students for the skills tested in the examination (see also Table 5.3). However, it is interesting to observe that 42% of teachers said they were paying more attention to the integration of skills, whereas only 36.8% agreed that they would *skip over certain activities in the textbook because they are less likely to be tested in the exam*.

Table 5.8. Teachers' practices for improving students' learning

Item	Response	Frequency	Percentage
I emphasise the skills which are more likely to be tested in the exam	Yes	155	80.3%
	No	38	19.7%
I skip over certain activities in the textbook because they are less likely to be tested in the exam	Yes	71	36.8%
	No	122	63.2%
I do more exam practice	Yes	112	58.2%
	No	81	42.0%
I put more emphasis on the integration of skills	Yes	81	42.0%
	No	112	58.0%

The other responses by teachers to TQ3.3 highlighted the main types of activities that they felt they needed to implement to improve students' learning in response to the revised examination. Sixteen teachers described the learning activities they preferred to engage in as a result, such as role-play, group discussion, task-oriented activities and extra-curricular activities. Thus, the results indicate that the teachers were aware of the positive changes to their teaching brought about by the new English elective examination, which would enable their students to use English more interactively and better perform certain real-life activities.

During the teacher interviews, three teachers commented that the new examination content was affecting their teaching. They seemed to be paying more attention to the skills tested. One teacher, for example, recounted devoting much of the lesson time to writing, rather than to the other skills. This was because writing required more practice and attention:

I do not teach everything in the book. I feel not all of it is important, because I need to prepare students for the exam. I give students speaking tasks as a warm-up, but I mainly focus on writing because we have many types to cover. I do listening tasks and reading sometimes, but my main focus is on the writing skill. (T2)

Another teacher expressed her inclination to not cover the entire contents of the textbook in detail, but instead focus on the skills that would be tested in the exam:

Usually, when I plan my lesson, I consider whether something is related to the final exam, so as to focus on and highlight it in my teaching. I give more time to these aspects in my teaching time. In teaching writing skills, for example, we have lots of formal letter types. I teach all the types introduced in the book (for example, for letter writing: job application letters, formal letters, letters of apology, etc.), but my focus will be on the type of writing included in the final exam. (T1)

Along similar lines, T5 made a similar comment in relation to teaching the reading skill:

In reading, I focus on the type of questions that are involved in the final exam, as we have three different reading texts in the exam, including different types of question. I assess students in my lesson to make sure they are prepared for these questions. (T5)

However, one teacher made at least one comment about the methods adopted in response to the new examination. This teacher reported that she had become more active in employing new ideas to improve students' proficiency levels and to motivate them. She also talked about using teaching strategies of the sort encouraged by the new textbook (role-play, pair and group work, etc.).

I always surf the web for updated strategies. I have used more new ideas, and adapted materials to present the language included in the textbooks. I take what is important based on the required outcomes and then adapt it for teaching. I use a more communicative approach, such as role plays and ways of thinking and discussion, and I push my students to work outside the box. And I also verify my questions by checking students' understanding, answers, and comprehension, as well as the way I deliver my lessons. (T3)

To sum up the above findings, it can be seen that the teachers' perceptions about the changes that occurred in their classroom teaching as a result of the new examination were multifaceted and showed a mixed picture. Some teachers believed that the new examination was having a strongly negative effect on classroom teaching, mainly in terms of the amount of attention they were having to pay to the skills tested in the examination as opposed to those not tested. It also seemed to be influencing their decisions about how much of the textbook to cover and the skills to emphasise in class. However, other teachers talked about the positive effects of the examination on their teaching in a way that reflected the intentions of the decision-makers. The qualitative data (interviews and open-ended questions from the questionnaire) in the next section provide an in-depth exploration of this topic.

The washback effect on teaching methods

TQ4.1 was an open-ended question in the teacher questionnaire, which asked teachers to give their opinions concerning the effect of the new English elective examination on their medium of instruction in the classroom. Most teachers (133 out of 155) who responded to this question agreed that the content and properties of the new examination would, to a

certain extent, influence the way they taught, especially in terms of reading skills. The teachers assumed that the new examination would require them to alter the content of their teaching rather than their way of teaching. A few teachers regarded adopting a more communicative approach to teaching the most suitable way of responding to the changes.

The teacher interviews yielded more detailed information than the open-ended questions in the teacher questionnaire concerning the impact of the new examination on teachers' classroom instruction. The five teachers provided a wealth of information regarding the emphasis they placed on the new examination specification in their teaching, their views on their teaching plans and the techniques they employed for teaching. The teachers were asked to describe their use of the new examination specification in their teaching to see whether they were paying attention to all the material in the syllabus or whether they were leaving anything out because of the new specification. They were asked about how much they used the examination specification and why they made certain decisions in their teaching. All five teachers acknowledged the importance of the new examination specification, but they used it for different purposes. Two teachers said that the exam specification made them familiar with the goals and content of the exam and the rating scales that were used to mark each section. Two teachers mentioned that the examination specification was the main source they had for information about the new components:

The information about the final exam is important. So once you know about the assessment goals and content, and the rating scales for each skill, you build your teaching on the basis of this. (T4)

I only read it at the beginning of the academic year. I just used it to acquaint myself with the syllabus, such as the outcomes I have to assess, or any other updates to the syllabus. But all the parts relating to the continuous assessment procedures and tools are not so interesting parts for me to read, and especially for experienced teachers, such as how to keep records of students work. (T3)

The other two teachers commented that they usually referred to the examination specification in their teaching. The problem of covering all the English Insights 3 textbook series in a single year has already been mentioned (see 5.2.4) and in the interviews the teachers stressed that they could not complete the syllabus without extra teaching time. These problems led the teachers to use the textbook in a similar way to the examination specification document:

It helps me to focus on the most important aspects of the syllabus. Similarly, this is also important for students, as they used to ask about the exam specifications. This specification and the information about exams are a priority for me when teaching.

Based on this, I am able to plan well in accordance with my students' needs, and to provide extra materials to add fun to my teaching. (T1)

Yes, I usually read this document and become familiar with it at the beginning of the academic year. This is important because I need to decide on my teaching plan. I need to know the final exam specification to focus on the right parts in my teaching. For example, there are 8 types of writing in the course book, but I need to know which of these are included in the final exam. (T2)

The second point that emerged concerning washback on classroom teaching was the use of the teaching plan as prescribed by the teachers' guide, or whether it could be adjusted or adapted to suit the requirements of the examination specification. All five teachers reported using the teachers' guide in ways that were different from the MoE curriculum designers' intentions. For example, T1 commented that if a particular type of writing appeared in the exam, her teaching plan would change in a way different from the steps outlined in the teachers' guide:

I usually teach writing skills in one lesson, but I will spend more than one lesson if the writing concerned is tested in the exam. If using one lesson, I will explain the writing type, analyse samples of written work, and then ask my students to complete the writing topic at home for homework. But if the text is tested in the exam, I will do the same steps as before, where I will explain, analyse a sample of written work, and discuss mistakes/errors and techniques regarding these samples, and then I will get them to do some writing as a group work. Groups then share their writing and discuss each other's writing. I give them another topic to practise the same type of writing for homework. (T1)

Another teacher reported having to change her teaching plan as outlined in the teachers' guide to match her students' outstanding English language proficiency level as the tasks included in the guide were not suitable:

Our students are really outstanding; they tend to get bored quickly if we follow similar steps to the guidebook. For example, one of the units in the book contains many vocabulary tasks, and the second part of this unit contains many speaking tasks. This is not good for our students. We worked on the organisation of this unit. As a result, we modified these tasks, and so we arranged it to cover vocabulary rules which we then use to talk about a certain topic. (T3)

Furthermore, there were a few comments about the methods that the five teachers used in teaching the new syllabus. Three teachers talked about trying to follow the methods for teaching set out in the teachers' guidebook, but they did not seem able to do so, either because of the problem of covering all the *English Insights 3* material, or because of the heavy workload involved in teaching the new syllabus. One talked about teaching reading skills to get students used to the strategies and skills required by the examination and another spoke of being unable to cover all the material in the textbook in the allotted time:

For instance, the book includes four units per semester; each unit includes plenty of different reading types. I spent time and effort selecting from this reading material to teach the students what they are supposed to achieve in the classroom, based on the exam requirements (such as skimming, guessing the meaning of words in context, and getting the gist). (T1)

Our preparation for teaching this new syllabus also changed, especially in that we have only four lessons per week to teach 8 units in the book. So, we are facing a big challenge in terms of the size of the textbook and the amount of teaching time available. So because of this challenge, I don't have enough time to cover everything in the book. In this sense, I am forced to choose and select what I should leave out and keep in my teaching plan. (T2)

The third teacher mentioned that they might neglect some parts of the textbook's approach if considered not useful for the final examination. The teacher also explained that she paid more attention to teaching writing in her teaching plan than to the other skills set out in the textbook because it was given more weight in the examination specification. For example, the teacher used the following steps in teaching students to write emails:

I feel we should focus more on writing because there are many different types of writing tasks in the final exam. For example, I used to teach email writing at all grade levels by first informing students that this type is called interactive writing. I then ask them to identify the information in each paragraph. Then, we discuss each paragraph by explaining the language and vocabulary needed for this type, and how they can make their writing interesting for the reader. I also focus on the language and style. (T5)

The other two teachers spoke about how they were now using more creative and communicative activities in teaching the new syllabus. They tended to employ a variety of class-based tasks and activities (such as using Smart TV or Padlet, watching videos to reinforce vocabulary and synthesise information across texts, preparing PowerPoint presentations, listening to authentic texts, and note-taking). The teachers emphasised how the changes in the elective syllabus had helped them to integrate different skills or employ activities that helped synthesise information from speaking, listening and reading into writing, specifically where these activities were not already included in the textbook materials:

I use a more communicative approach, such as role plays and ways of thinking and discussion, and I push my students to work outside the box. And I also verify my questions by checking students' understanding, answers, and comprehension, as well as the way I deliver my lessons. (T3)

T4 provided a particular example of how she taught writing skills in the new elective syllabus:

For example, the opinion writing includes different styles, such as expressing an opinion for or against some writing. So, if I just explain the differences between these two types, students might mix the two types up and get confused. But I can, for example, explain these types by giving them OREO chocolate and asking them to state the reasons for their opinion. Thus, they will never forget this type of writing and it will stick in their minds for ever. I also ask groups to share their writing and discuss it as a whole class. I also ask them to perform their writing as a story and I use their writing to design videos. In their free time, they watch these videos they have written. (T4)

Alongside this material, question TQ4.3 explored how many teachers carried out the 17 teaching activities listed in Table 5.9 in their classes. The table shows how often teachers tended to carry out these teaching activities in Omani post-basic schools. They are arranged in descending order, rather than in the order used in the questionnaire, for ease of discussion. It is apparent that activity A1, do activities similar to mock exams, was carried out by the highest number of teachers. This activity was closely followed by A2, practise skimming and scanning skills, together with activities A3 to A6. Activities A7 to A11 were carried out by teachers by slightly fewer teachers. A12 to A16 were not commonly used in the classroom. The last activity, A17, memorise model samples for writing, was ranked lowest among the items. A comparison of A17 and A1 suggests that teachers tend to concentrate on teaching activities geared towards the requirements of the new English elective examination, thus supporting previous results.

Table 5.9. Teaching activities

Item	Frequency	Percentage
A1. Do activities similar to mock exams.	135	69.95%
A2. Practise skimming and scanning skills.	130	67.36%
A3. Tell the students the aim of each lesson.	122	63.21%
A4. Watch movies or news in English.	115	59.59%
A5. Organise real-life language activities (e.g. mock interviews, sketches).	113	58.55%
A6. Organise group work or discussions.	108	55.96%
A7. Practise notetaking.	98	50.78%
A8. Read newspapers, magazines, or books written in English.	94	48.70%
A9. Memorise useful expressions for writing text in the final exam.	93	48.19%
A10. Study grammar rules.	89	46.11%
A11. Practise model samples for speaking tasks.	81	41.97%
A12. Demonstrate how to carry out particular language activities.	78	40.41%
A13. Discuss textbook exercises.	68	35.23%
A15. Practise pronunciation.	67	34.72%
A16. Memorise vocabulary lists.	64	33.16%
A17. Memorise model samples for writing.	57	29.53%

During the interviews, the teachers were asked about the attention they paid in their lessons to each of the skills areas tested in the new examination. With regard to teaching reading,

the five teachers suggested that the new examination content and strategies were having an impact on their approach. One of the clearest impacts was that the teachers tried to cover all the matching reading items (see 2.2.4), which they said was because of the exam. As T1 remarked, “I usually give students practice in matching paragraphs to the topic sentence question and discuss with the whole class what they need to do to answer this type of question”. Three teachers considered that they generally used the passages and questions just as they appeared in the book but said that they added extra questions similar to those in the examination, either of their own design or from past examination papers, to supplement the Insight 3 series. The following extracts illustrate this:

If the reading text in the course book includes similar question types to the final exam, I spend time doing these in my lesson. I allocate more time for these tasks in my teaching plan. I also allow students to work individually (instead of in groups) when they have reading text types similar to the exam. I sometimes take the opportunity to measure students’ performance when they do these tasks, such as once every two weeks. (T5)

The textbook includes similar reading texts to the final exam, but I usually employ extra reading tasks similar to the final exams to draw students’ attention to these tasks. I discuss with students the methods and strategies for answering these questions. (T3)

T4 mentioned that the teachers in her school used to modify the reading questions in the textbook, because:

Some reading texts in the book are written across two pages and include only one task or question, while other questions are not well written. So, we add exam-like questions to train students on these questions and assess their comprehension. Thus, it is a win–win situation. (T4)

The above findings on the teaching of reading indicate that the new exam had an important influence on the amount of attention the teachers were paying to reading and the type of material they used in their classes. It is important to note, though, that the teachers focused on practising the skills and strategies that appeared in the examination, rather than on the content of the reading texts.

In relation to teaching writing, all five interviewed teachers discussed the steps involved in how they taught this skill. Three teachers noted that the writing types in the exam were covered in the textbook, as seen in the comment, “The good thing about this change is that the topics and types of writing in the exam are mostly found in the book” (T4). All teachers said that they covered all the writing types in the book, but they were paying extra attention

to the types of writing that appeared in the examination. Their comments on this included the following:

I think it is important to focus on the writing included in the exam because these types are important for students' life, study, and work. This does not mean I skip the other types which are not included in the final exam. We teach them but we do not focus on them as much as those included in the final exam. (T1)

I focus primarily on the elements included in the exam because these are the most important for the students to learn, and it keeps them relaxed. I have this target from the start of the academic year, to prioritise the elements students need for the final exam and the class tests. (T5)

When teachers were asked about their teaching of listening skills, they all said that they paid more attention to these than to the other skills. Three teachers believed that some listening questions in the exam were new (as opposed to the examinations at lower grade levels), such as “gap-filling tasks” and “note-taking items” (T2 and T3). One teacher commented that they usually added extra questions to the listening texts that appeared in the textbook because the textbook included questions at too low a level:

I arranged my teaching plan to teach students the listening task from the book for 20 minutes, and then give them some practice from the international exams. I usually modify the questions for the listening task in the elective book to make them more challenging and interesting, and get them used to the listening question for the final exam. For example, I skip some of the listening tasks in the book because they only ask students to practise pronunciation, so this is a waste of time for the students. (T3)

Another two teachers revealed that they covered all the listening texts in the book, but one said that she allowed students to listen to the text at home and paid much more attention to practising the listening questions that appeared in the examination:

I also do the listening tasks introduced in the students' book, but I ask students to listen to the text at home before attending the listening lesson and check the answers later with them. (T1)

Another teacher mentioned that she used extra materials to practise the listening topics that appeared in the textbook:

I teach students listening tasks from different units or themes in the book. I give them extra practice for the listening skill from different resources, such as web-based links. (T5)

In summary, regarding teachers' perceptions about their teaching methods, most teachers when asked in the questionnaire and interviews believed the examination specification had a significant impact on the content of their teaching. In addition, the interview findings

revealed that the five teachers did not cover the teachers' guide thoroughly because they believed some topics in the book were not tested in the examination, the book content did not match their students' high level of English proficiency and it was difficult to complete the book in the allotted teaching time. Furthermore, while the teachers did not explicitly state that they had altered their teaching methods, there was a clear tendency to focus more on the content and skills tested in the new examination. Specifically, teachers were inclined to focus on the activities directly related to the requirements of the new English elective examination, as discussed further in the reporting of the results of the classroom observations (see 5.4).

Teacher classroom assessment practices

Regarding TQ2.F, six items were listed under this category to explore teachers' views of the primary function of the mock examinations in grade 12 post-basic schools in Oman. The results are shown in Table 5.10.

Table 5.10. Role of mock examinations as perceived by teachers

Item	Strongly Disagree N (%)	Disagree N (%)	Undecided N (%)	Agree N (%)	Strongly Agree N (%)	Rank	M	SD
To prepare students for the real exam	5 (2.6%)	6 (3.1%)	7 (3.6%)	76 (39.4%)	99 (51.3%)	1	4.34	0.893
To give feedback to students	4 (2.1%)	5 (2.6%)	7 (3.6%)	116 (60.1%)	61 (31.6%)	2	4.17	0.786
To motivate students	3 (1.6%)	12 (6.2%)	18 (9.3%)	97 (50.3%)	63 (32.6%)	3	4.06	0.899
To encourage students to study regularly	3 (1.6%)	13 (6.7%)	17 (8.8%)	106 (54.9%)	54 (28.0%)	4	4.01	0.884
To identify areas for re-teaching	5 (2.6%)	13 (6.7%)	25 (13.0%)	100 (51.8%)	50 (25.9%)	5	3.92	0.943
To get students to pay attention in class	2 (1.0%)	27 (14.0%)	29 (15.0%)	96 (49.7%)	39 (20.2%)	6	3.74	0.971

Observing the patterns in the mean scores of the items, it can be seen that *to prepare students for the real exam* was rated as the most important function of the mock examinations. In post-basic schools in Oman, high-stakes tests clearly play an important role. Indeed, as reported in Chapter 1, and in line with the findings from the questionnaire, teachers regarded preparing students for the end-of-year exam as one of their main responsibilities. Moreover, *to give feedback to students* was ranked second by the teachers. The large number of teachers viewing the mock examinations as a tool for providing feedback to students is supported by the results elsewhere (see Tables 5.8 and 5.10). The teachers also felt that mock examinations would motivate the students and encourage them to study regularly. The functions of

identifying areas for revision and getting students to pay attention in class received the lowest support.

When interviewed, the teachers commented on the function of examination preparation lessons, their teaching plan for employing them and the content and sources of mock examinations. Three teachers believed that examination preparation lessons were needed to familiarise students with the content and requirements of the final examination and to give them focused practice on exam-style questions. Their comments included:

They [the mock exams] are important because students will be made aware of the kind of questions that will be included in the final exam. We did not receive any mock exams for the grade 12 elective this year, but we have them for other grades. So, students feel, worried because they should practise the actual exam. (T5)

I usually employ mock exams to help students practise for the actual exam, and I think they need more focus on the new form of the questions. (T1)

One teacher noted that they prepared students for the types of writing text in the examination more than any other feature. This teacher commented that the textbook did not include extensive practice for the writing component of the exam, thus “students feel stressed because they should get more practice on the type of writing before the actual exam” (T5). Another teacher prepared her students for the listening component of the examination by discussing the requirements of each question and then asking the students to answer the questions for a whole lesson. Some teachers also provided students with extra listening practice for homework. One teacher explained why she preferred to set the listening as homework rather than doing it in class:

I ask students to do the listening at home before attending the listening preparation lesson to reduce the lesson time spent on listening. I only check the students’ answers when practising the listening component of the exam. (T2)

For three teachers, examination preparation took place during their regular lesson time, with one lesson often taken up by each component of the examination:

I start with one lesson, 45 minutes, for listening. Then I use the next lesson for the reading questions, and then another lesson to practise the writing questions. I usually include exam preparation lessons whenever I teach something related to the exam. (T2)

During lessons preparing for the examination, teachers used past papers, model questions from web-based English language materials (T2 and T5) or books published commercially (T1):

I browse different websites, such as the British Council, for reading and listening questions, past exam papers, past mock exams. I give students these exams during my lesson teaching time. (T1)

I provide preparation for each skill separately in different lessons. I also teach them techniques for answering the question types. I also discuss and share feedback with the whole class about each question in the exam preparation class. (T2)

I distribute the papers, and I ask them to answer the questions individually or sometimes as group work. (T5)

T3 and T4 worked their way through their teaching plan throughout the year without feeling the need to give specific lessons for examination preparation or mock examinations. They employed the textbook and other supplementary materials they designed themselves in a systematic manner:

For me, I don't think exam preparation is important, as I already prepare my students throughout the year. So I do not waste time on preparation, it is covered in my teaching plan. By the end of the year, students know what they are going to face in the final exam. We prepare them from the beginning of the year, and they are prepared for this stage. (T3)

Actually, I do not need to worry about exam preparation as I spend the whole year teaching them different questions similar to the exam questions and demands, and I have taught them how to use the language effectively. (T4)

To summarise, the role of the examination preparation lessons was to prepare students for the new requirements and to give them more practice in answering questions. The teachers used examination papers from previous years and other extra materials besides the textbook. They had to spend lesson time covering each component of the examination and they had different ways of giving their students practice in these components. A few teachers did not devote time to teaching exam-like questions but concentrated instead on teaching the content of the syllabus thoroughly, in a systematic manner.

In TQ3.2, teachers were invited to identify which methods they used to assess students' learning in their schools. The assessment methods listed in Table 5.11 are outlined for teachers in the specification of the new English elective diploma examination. The purpose of this category was to determine the assessment method that most affected teachers' daily teaching.

Most teachers reported using *written work* and *daily observations*, followed closely by *quizzes*. This seems reasonable because these three methods of assessing students' learning are commonly employed in everyday classroom teaching. *Class tests* came out slightly lower as a method for assessing students' learning. The fact that teachers listed *presentations* as the

assessment method least used in their teaching suggests that the more important an assessment type is, the more attention they pay it. In the case of the assessment specifications of the English elective diploma examination, presentations do not feature.

Table 5.11. Assessment methods most frequently used by teachers

Item	Response	Frequency	Percentage
Day-to-day observation	Yes	131	67.9%
	No	62	32.1%
Presentations	Yes	89	46.1%
	No	104	53.9%
Written work	Yes	132	68.4%
	No	61	31.6%
Quizzes	Yes	119	61.7%
	No	74	38.3%
Class tests	Yes	101	52.3%
	No	92	47.7%

In the interviews, all five teachers talked about the class tests in the new English elective syllabus that are used for gathering information about students' attainment of the learning outcomes. In the grade 12 English elective, teachers are required to write three class tests per semester, with each test being awarded a total of 10 marks. The teachers talked about the design of the class tests they had written, which were influenced by the examination content. They referred to the influence the question types had on the way they chose or found a question or a specific text from past exam papers, the textbook, or other official examination publications. For example, teachers spoke about the features which resembled the examination design:

I think it is a bad practice to give students a quiz or test that contains different questions to the final exam. For example, there are four important reading text types: narrative, informative, interactive, and evaluative reading. So, I must give students practice on all these text types, even if they are not required by the exam. (T1)

The continuous assessment practices lead to success in the final exam. Continuous assessment is helpful because it prepares and equips students for the final exam. In my class tests, I give students practice in unseen and authentic reading texts. I give students tips and guidelines for writing any type of text or, if I have assessed them on grammar in the continuous assessment, this helps them in writing the final exam, even if grammar is not included in the exam. (T4)

All five teachers talked about testing reading, listening, or writing without being asked directly. However, they did not talk about testing speaking as this is not required by the revised syllabus. One teacher commented that, "Even if the listening skill was not required for the continuous assessment but assessed in the final exam, we gave priority to reading, listening, and writing when it came to testing speaking" (T4). Another teacher said that "As the skills tested in the exam and continuous assessment were different, what we need is to

give equal attention to all skills when teaching because we do not want to assess content as outlined in the textbooks, rather assessing the skills” (T3).

In summary, the results indicated that the examination seemed to be having a strong effect on the continuous assessment element, mainly in terms of the choice of assessment method, the design and types of test used and the amount of attention teachers paid to listening, reading and writing compared to speaking.

In TQ2.G, teachers were asked about their perceptions regarding a list of factors that influenced their teaching. Their responses were listed according to their mean scores, as shown in Table 5.12.

Table 5.12. Factors influencing classroom teaching

Item	Strongly Disagree N (%)	Disagree N (%)	Undecided N (%)	Agree N (%)	Strongly Agree N (%)	Rank	M	SD
Teaching experience	6 (3.1%)	10 (5.2%)	7 (3.6%)	106 (54.9%)	64 (33.2%)	1	4.1	0.922
The need to obtain satisfaction in teaching	1 (0.5%)	5 (2.6%)	27 (14.0%)	112 (58.0%)	48 (24.9%)	2	4.04	0.735
Teaching beliefs and attitudes	5 (2.6%)	8 (4.1%)	19 (9.8%)	105 (54.4%)	56 (29.0%)	3	4.03	0.889
Textbooks	3 (1.6%)	8 (4.1%)	14 (7.3%)	130 (67.4%)	38 (19.7%)	4	3.99	0.76
Professional training	5 (2.6%)	14 (7.3%)	23 (11.9%)	86 (44.6%)	65 (33.7%)	5	3.99	0.992
Learners' expectations	3 (1.6%)	6 (3.1%)	25 (13.0%)	126 (65.3%)	33 (17.1%)	6	3.93	0.75
Final exam results	5 (2.6%)	18 (9.3%)	29 (15.9%)	111 (57.0%)	30 (15.5%)	7	3.74	0.922
Parents' expectations	7 (3.6%)	27 (14.0%)	51 (26.4%)	92 (47.7%)	16 (8.3%)	8	3.43	0.956
Supervisors' expectations	5 (2.6%)	29 (15.0%)	56 (29.0%)	82 (45.1%)	16 (8.3%)	9	3.41	0.932

As the table shows, the teachers regarded *textbooks*, *professional training*, *learners' expectations*, *final exam results*, *parents' expectations* and *supervision expectations* less important than the other factors. An interesting discovery here is that textbooks, which are the most commonly available teaching materials in the Omani school system (see Chapter 2), were not seen as more significant than teaching experience, beliefs and satisfaction in teaching. Another key finding was that teachers perceived the final examination results to be an avenue of support for teaching.

In the interviews, the five teachers mentioned some of the above factors when questioned about the issues that most influenced the way they conducted their lessons and affected whether they could react to the demands and design of the new examination as intended by the MoE decision-makers. Other factors identified included students' attitudes towards the new examination, supervisors' expectations and the role of the principal.

Four of the teachers reported their students' fears and concerns about passing the examination. T4 talked about students' being fearful of the new exam because they did not have enough information about its demands and the level of expertise required:

Overall, students had some fears and concerns regarding the new exam. Some students wrongly believed that this new elective is suitable for those who need further support in English. But I assured them that this new elective change is for students who want to develop their abilities to become more skilful in using the language. (T4)

Similarly, T5 explained that students had fears and concerns about the new examination design and requirements because of the change in the difficulty of the questions compared to the previous version. This teacher also described how students felt about the listening component, which was newly added to the examination:

They had some concerns and fears, especially about the addition of the listening skill. They choose it because they think that this subject is a "remedial teaching plan" used to raise students' language level and prepare them for grade 12. So after grade 11, they mostly drop it, as they feel it is very demanding and requires lots of time and effort. So, students were very shocked when they first selected this subject, as it requires a higher English language level than they were expecting. (T5)

Another teacher believed that students would practise all the types of writing that appeared in the textbook if they thought they could all be part of the writing component in the examination. This finding is interesting because such changes in the examination could be causing the whole educational system to react to students' exam-related demands:

My students were not happy about the new changes to the writing skill, because there are many styles they need to learn which are not tested in the exam. My students wanted to know what type of writing they would get in the exam, and they were upset if I was not able to prepare them for what was coming in the writing test (T1)

Another interesting finding was that only one teacher commented on students' motivation in relation to the new format of the examination.

We usually give students a brief about this subject and why they should take it. So, they were aware of the change to this subject. They liked this change, and they were eager to be tested on the types of questions that might appear in the exam. (T3)

The second classroom-related factor to emerge was the importance of supervisors in guiding and supporting the teachers who were trying to cope with the new examination. Four of the teachers spoke positively about their supervisors. They said that even if their supervisors did not have enough information about the objectives and requirements of the new examination, they still held regular discussions and guidance sessions with them about their teaching and their ability to cope with the changes:

Our school supervisors really support us a lot. From the beginning of this change, they were, like us, not fully aware of the changes involved. Later, they became very supportive. Based on their visits to the new elective classes, they noted that the teaching methods used by teachers in these classes were different from those used in the main grade 12 core classes. (T4)

They support us, but like us, they felt lost, especially at the beginning of this change. They were not involved in, part of, or even consulted about this change. (T5)

Supervisors were supportive, they had regular visits to our classes and provided us with sufficient feedback and guidance. (T2)

One teacher, however, indicated that the role of the supervisor was not always effective. This teacher explained that the supervisor in their school was only concerned with the students' level of performance and checking up on the teacher's lesson plans, rather than providing support or discussing any challenges with the teacher:

My school supervisor was concerned about the students' ability level, the content we teach, and whether or not we are following the teaching plan. (T1)

Moreover, the role of the principal as a factor influencing teaching practices in classroom emerged in the teachers' interviews. All five teachers complained about the role of their school principal in relation to this new change. One teacher revealed "I realised that our school principal did not know that the new elective syllabus was more intensive and covered different skills from the previous syllabus" (T2). Another teacher commented that her school principal did not know "why I had made so little progress in teaching the new syllabus" (T5). Also, T4 remarked that "My school principal did not know until late that the students needed to put in more effort and attention if they were to succeed in the new elective exam".

To sum up, the effects of the new elective examination on teachers' practices are intricate. First, the findings indicated negative reactions among teachers in the first year of the introduction of the examination. For example, while the majority (88.1%) of teachers were not sceptical about the need to change the old elective examination, a sizable proportion (79.3%) of them did not endorse the new elective diploma examination. These attitudes were

related to challenges and difficulties in their teaching arising from the new content and format of the examination. However, the interviews showed that the negative feelings changed for the better some time on from the implementation of the exam. Moreover, teachers were aware of the need for the new examination and had better knowledge of the new format and the changes made to the English elective diploma. Interestingly, moreover, most of the responses to the questions seem to be mildly positive, responding “agree” rather than “strongly agree” on the five-point scale (see Tables 5.2, 5.3, 5.4, 5.5, 5.7, 5.10 and 5.12). This might indicate that the teachers saw little difference between the two options, or be a typical feature of questionnaires, namely that people are less likely to opt for extreme points of a scale (Burchell & Marsh, 1992; Rolstad et al., 2011).

Second, in terms of washback on teaching practices, most teachers agreed that they would prepare their students for the new examination design and requirements. Teachers generally believed that it was beneficial for students to master the content of the new examination, which assesses reading, writing and listening. It also seemed to be influencing their decisions about how much of the textbook to cover and the skills to emphasise in their classroom teaching. There was a tendency for the teachers to pay more attention to activities which were directly related to the requirements of the new examination. However, a few other teachers talked only about the positive effect of the examination on their teaching methodology, in a way that reflected the intentions of the decision-makers.

Third, the findings concerning washback on assessment practices in the classroom showed that teachers devoted lesson time to teaching exam-like questions and the examination seemed to be having a strong effect on their continuous assessment practices. Therefore, it seems that explaining the changes teachers would like to implement in their teaching as a result of the new examination requires further investigation and this will be explored further through the in-depth interviews with the MoE decision-makers (5.3) and detailed classroom observations (5.4).

5.3 Intended Washback

This section reports on the intended washback that the MoE decision-makers (exam writers, supervisors and curriculum developers) aimed to achieve from the new English elective diploma examination. It addresses the second research question of this project: “What is the new elective diploma examination's intended washback as perceived by the MoE decision-makers?” The purpose of this is to analyse the responses to the specific expectations and objectives of the new examination as perceived by the MoE decision-makers and to

determine whether the introduction of the new examination has been as effective as the MoE decision-makers hoped it would be.

In discussing the findings in this section, excerpts from the four MoE decision-makers are presented. The names of the curriculum and assessment developers and supervisors were replaced with the pseudonyms MoE-1, MoE-2, MoE-3 and MoE-4. All the interviews were conducted in Arabic and translated into English. A detailed explanation and description of the process of analysing the interview data was presented in Chapter 4 (see 4.8.4). The codes and themes generated from the interviews were grouped under three overall themes: (a) enhancing students' language skills; (b) providing continuous assessment for learning; (c) improving student-centredness and teacher practices. Table 5.13 shows one example of how the data were coded based on semantic and latent meanings and then categorised accordingly.

Table 5.13. Sample coding of data

Theme One	Sub-Themes	Examples of Participant Responses
Enhancing students' language skills	Rationale for the change	“Students and teachers felt that the old exam paper was a burden because it mainly focused on different types of writing.”
	Inadequate coverage of language skills	“We need to improve the outcomes in grade 12, where students struggle to meet the IELTS standard levels required to study at higher education institutions or for the workplace.”
	Low IELTS exam scores	
	Negative impact of the old exam on teaching	“The lack of learning strategies and the failure to specify the required methods for teaching language skills in the old version of the elective were evident problems.”

As illustrated in the table, one of the insights from the first theme, “enhancing students’ language skills”, relates to what the previous examination looked like before the attempt to introduce new skills in the English elective diploma. This theme concerns whether the introduction of the new examination was perceived as appropriate for the stakeholders (such as teachers, students, parents, etc.) and whether the reform was deemed as effective in influencing classroom teaching as the MoE decision-makers had hoped. It is important to address both the intentions of the MoE participants in making changes to the examination and it compares to the previous examination that it replaced. A complete listing of the codes and themes is included in Appendix M.

First theme: Enhancing students' language skills

The first theme, “Enhancing students' language skills” (referring to speaking, reading and writing, in addition to grammar and vocabulary knowledge), was considered one of the main intentions of the MoE decision-makers in introducing the new examination. MoE-2 and MoE-4 noted that the decision to change the old examination content was taken by all the relevant parties involved in English language teaching in the MoE, including those working in supervision, assessment, curriculum development, training, teaching, etc. These participants also mentioned that the decision was shared with other external bodies, such as the Ministry of Higher Education and the University of Sultan Qaboos, among others. The following comment illustrates this:

The decision was made at a committee level. All the parties involved in English Language teaching were represented in this committee. The committee at the MoE level included members from curriculum development, assessment, training, and supervision. Apart from these people, we consulted external bodies, such as the Ministry of Higher Education, SQU, the Educational Council, private higher education institutions, etc. The external bodies were used as consultants to monitor the quality assurance of the process and the overall decision. (MoE-4)

It is clear from this that decisions about the change to the English elective diploma were initially made at the level of the policymakers within the Omani educational system. The involvement of different consultants with the MoE members regarding this change illustrates this and also helps demonstrate how the Omani educational system is structured. To meet the MoE's requirements for substantive changes to the English elective examination, the MoE policymakers first had to identify the problems within the existing curriculum. Thus, the rationale for the change in the examination is discussed in the following paragraphs.

Rationale for the change

The participants make several comments about the need to introduce the new English elective examination. These could be divided into three sub-themes: inadequate coverage of language skills, low IELTS scores, and the negative impact of the old examination on teaching. With regard to the first point, the MOE participants remarked that the previous examination only tested reading and writing and neglected speaking and listening:

The previous exam only included reading and writing, which was not enough to develop the students' language skills. (MoE-1)

Students may have difficulties in listening, so why do we not give them the option to develop their listening skills? (MoE-3)

It is good to cover the main skills, reading and writing, but at the same time, students need to develop other skills. (MoE-4)

MoE-4 emphasised the need to focus on skills other than reading and writing to meet the outcomes expected by higher education institutions or workplace requirements. Overall, all participants had negative opinions about the content of the old examination in that it ignored several language areas, such as listening, grammar and vocabulary. Moreover, MoE-2 pointed out that the focus of the old examination was more on writing skills than on reading skills, “because the old textbook was mainly a writing course rather than an English language teaching coursebook.” This participant further noted that the focus on writing meant skipping straight to the highest stage of evidence for language learning, without addressing the skills required to get to this stage. The participants also suggested further possible improvements, including expanding the new examination to test oral skills (speaking and listening) and giving students more choice over the passage they read or listened to, the level of the paper they sat and the writing tasks they had to respond to.

In addition, the participants reported that another significant factor in the decision to introduce the new examination was the concern among various higher-level policymakers in the country (including the Oman Academic Accreditation Authority, Sultan Qaboos University, MoE, and other higher education institutions) about the inadequate English language skills of grade 12 students. They mentioned that after grade 12, students were struggling to meet the language demands of tertiary studies. All four participants observed that the previous examination failed to prepare students sufficiently for international English language tests, specifically IELTS. For example, MoE-2 said:

Grade 12 students would be unable to cope with the demands and level of the IELTS exam because they did not speak English sufficiently well or because they lacked the skills and strategies required for the exam. (MoE-2)

MoE-1 discussed how their work on the new examination specification sought to accommodate the IELTS standards in terms of the elements of writing, reading and listening, stating that “we focused on the elements of the IELTS exam rather than other international exams because we in the Omani context are following the British approach to teaching”. The participants stressed the importance of IELTS, which provides access to higher education and employment in Oman. Because of this, the revisions to the old examination included the expansion of content in line with the demands and requirements of IELTS, for example in relation to the development and application of English language skills and the assessment of academic study tasks.

Furthermore, it was recognised that the previous curriculum was having a negative impact on the way students were learning English. The participants discussed its effect on students' general use of the language when they graduated from high school. The lack of oral and aural language skills (speaking and listening) affected students' ability to use English as a language of communication. Students would study listening and speaking only if they knew that they would have to study these two skills as part of the examination. Such an approach was not only an issue in relation to grade 12 students. Indeed, the whole educational context was organised around students' exam-related demands. This situation had deep roots in the country's culture and suited the students as they saw it as an appropriate way to prepare for the final examination. For example, MoE-3 reported:

This is not good practice because students in grade 12 mainly focus on the content of the exam. So if grammar and vocabulary are not going to be assessed, students will not pay attention to them. Students care about their marks in the exam, especially in grade 12. In my country, students are not studying for the sake of learning; rather, they study to get good grades and achieve the requirements set by higher education institutions. So we need to reconsider this situation. (MoE-3)

In addition to these factors, other comments about the problems associated with the previous examination pointed to the need to introduce a new one. MoE-3 and MoE-4 reported that the examination itself had a negative impact on teaching methodology:

A number of learning strategies which could help students develop their English language level in the outside world and in the classroom were not explicitly covered in the textbook. One of these was the teaching of skimming and scanning skills as part of the reading curriculum, since these skills are essential for students to develop effective reading habits. (MoE-3)

The exam specification focused primarily on writing. The types of writing assessed in the continuous assessment gave teachers the notion that they should teach specific writing styles rather than the underlying skill itself. In my opinion, this undermined the communicative approach that the English language teaching in the Omani educational system is trying to promote. (MoE-4)

These two experts explained the specific features of the old examination that could prevent students from developing their language ability in the outside world as well as in the classroom. One of these was the lack of coverage of skimming and scanning strategies in reading, which MoE-3 felt were beneficial techniques for students who were still developing their reading habits. The other issue regarding the old exam was that it paid too much attention to writing. This had a negative effect on the teaching and learning that was taking place in the classroom.

It is interesting to note that there were a few comments about the extent to which the examination could influence curriculum change or teachers' behaviours. This relates further to how decisions about the new elective examination were made by the MoE English language decision-makers and those within the Omani educational system. For example, more teachers were involved than previously in making decisions about changes to the English elective examination during its implementation phases, especially in the second year, known as "the customisation period" (MoE-4):

A large sample of teachers from different governorates participated in the decision-making committee at the MoE level, which made them familiar with the objectives of the examination change, while several teachers participated in the workshops that the MoE members prepared for introducing the exam. (MoE-3).

The following section provides more details about the changes encapsulated in the new English elective examination.

Development of general and academic language skills

One way of identifying the new features included in the examination content is to compare the two versions, old and new. In this respect, the point most frequently made by the participants was that the new examination measures a wider range of skills since it includes listening in addition to reading and writing:

The topics tested in the new exam are more relevant to everyday life than those in the former exam, particularly with regard to the writing and listening questions. (MoE-1)

The new exam tests more skills in ways that are more interesting than had been done previously. (MoE-4)

The new examination includes various types of question to develop reading skills, and the format of these is new to this context. The old examination tended to feature mostly multiple-choice and short-answer questions. (MoE-2)

The new types of question relating to writing skills require students to use synthesis, contrast and comparison, statistical knowledge, and language skills. (MoE-3)

Other comments related to the specifications of the new examination paper. MoE-1 stated that "students had to employ different reading strategies to make their way through the reading texts", MoE-4 mentioned that "they had to write more for all of the skills, not just for the writing skills", and this was also emphasised by MoE-2, who reported that "they were given different types of optional writing questions". Participant MoE-3 pointed out "the listening elements tested in the new exam are more interesting and realistic than those in the old exam".

It is clear that the development of the three core linguistic skills in the new examination was felt to be far better compared to the previous one. The new version included an emphasis on everyday skills, a focus on listening, opportunities for a more student-centred approach and the need to use higher-order thinking skills that require active involvement rather than just recognition. The new examination questions as reported by MoE-1 “go beyond simple language learning and teach skills that will support success in all areas, testing such skills as writing formal letters, providing descriptions of data, form-filling, expressing opinions, etc.”. Moreover, the participants mentioned that the new examination was specifically designed to cover the academic skills included in IELTS (such as note-taking and summarising), which are essential for higher education:

Students are struggling when they travel abroad, and they face challenges and have to spend a long time passing an English language foundation programme, so the development of the academic skills was essential for our grade 12 students. (MoE-2)

The new format of the writing questions includes news reporting tasks or assessments and formal report writing, which are similar to the IELTS writing tasks. These tasks help students to learn new skills and competencies that were not included before in the other exam. (MoE-3)

The relative advantages of the features of the new examination were considered indications of a successful change. The principal intended washback from the changes was to foster the English language skills students would need for academic study in higher education institutions and in the workplace. By including tasks that assess higher-order thinking skills, the new examination was likely intended to develop critical thinking, problem-solving and interpersonal abilities among students. (See Appendix A for the new examination specification).

Second theme: Providing continuous assessment for learning

In addition to the end-of-year examination, another feature of assessment under the English elective diploma curriculum is teacher classroom assessment conducted throughout the school year for formative and summative purposes. According to MoE-1, continuous assessment aims at “providing a fairer, more balanced picture of students’ attainment”. In similar terms, MoE-4 believed that continuous assessment is more important for learning than the end-of-year examination because “It provides useful information about students’ English language development”.

The continuous assessment element incorporates various formats for gathering information (such as semester tests, presentations, day-to-day observations, written work and projects,

portfolios and quizzes), enabling a comprehensive evaluation of the students' English language attainment (SAH, 2021). All four participants pointed out that these tools were designed to help teachers to conduct assessments throughout the year so that the teacher could give every student a chance to perform and interact with each other. For example, MoE-4 commented:

The various tools and strategies used to assess the speaking skill in the current change are more varied than before. Speaking can be assessed as one way, two ways, and interactively. There is an emphasis on students' interaction with each other in the class, and to give students' chance to perform better under relaxed environment. (MoE-4)

The main skills assessed through continuous assessment are speaking, reading and writing. MoE-1 and MoE-4 revealed that the marking distributions for these skills were based on the content and priorities of the English elective textbook, "*English Insights 3*":

The weightings in the continuous assessment are different than before because the outcomes expected from the students have changed. Grammar and vocabulary in the continuous assessment are integrated with other skills rather than assessed separately. Also, speaking and writing are combined in order to be more useful. (MoE-2)

The continuous assessment element complements the assessment provided by the final exam. The distribution of the overall mark is based on continuous assessment (30%) and the final exam (70%). (MoE-1)

The teachers may teach grammar and vocabulary implicitly or integrate them with the other main skills. They cannot skip these components, even though they are not tested in the final exam, because they are important for students' reading and writing skills. So, students would not be able to write accurately without understanding grammar or learning new vocabulary. (MoE-4)

There are differences in the skills tested in the new examination and those assessed using continuous assessment. The latter covers speaking and grammar and vocabulary (not assessed in the final examination), which are combined and integrated in the teaching and assessment of the other skills, such as writing. Hence, continuous assessment was intended as an aspect of positive washback, leading to a focus on skills not tested in the final examination. However, the greater weight of the final examination tended to lead to a focus on the skills tested to improve students' attainment and increase their confidence in taking the examination, not, as intended, a focus on speaking, grammar or vocabulary.

With regard to writing skills, the four participants reported that the type of texts assessed in the final exam are not the same as those assessed in continuous assessment. For example, in

the continuous assessment element, students are required to write a description of data (from a graph/figure), a type of writing that is not included in the final exam:

We cannot assess all outcomes in the exam; we need to balance the content among the two assessment tools. For writing, if we have four types of writing, we assess students on two types in the exam and on the other two in the continuous assessment. Students need to understand all the writing types because they will be assessed either in the continuous assessment or the exam. (MoE-3)

Writing skills are assessed differently in the continuous assessment than in the exam. Students' attainment levels vary, so in the continuous assessment we give students more time to submit drafts, and they have three drafts to develop and practice their writing skills. Students are given feedback during the continuous assessment on their writing performance. Assessment of writing skills in the class allows students to expand their knowledge about the topic involved before they write about it. (MoE-2)

The above excerpts illustrate how the new changes directly affected students' learning. Students would learn all the styles of writing included in the textbook even if they knew they would not have to apply them all in the final exam. For example, students would learn how to write descriptions of data and biographical and literary essays in the classroom but might be asked to evaluate opinions on particular topics in the examination. Thus, teachers would not have to "teach to the test". The focus could be on learning rather than test preparation, which could help better develop the students' English language ability.

Another issue that surfaced from the analysis was the mock examination preparation period. MoE-2 believed that "the mock exam preparation period was useful, but it should not be the focus of the teaching". Similarly, MoE-3 said that "we encouraged teachers to view this period more positively, rather than treating it purely as a mechanical period of teaching". Based on my experience in this context, MoE decision-makers have clear goals and purposes for teachers regarding the use of mock examinations in classroom teaching. MoE decision-makers recommended teachers strive for a balance between teaching and testing, avoiding an excessive focus on coaching for the examination or narrowing the curriculum to addressing tested skills. More importantly, they encouraged teachers to develop efficient methods for resisting pressure from students to allocate a disproportionate amount of teaching time to mock examination activities and focusing on mock examination questions at the expense of communicative activities. Teachers were advised to use these examinations as a diagnostic tool for providing information and feedback concerning learning, rather than equating students' language learning with a score in the examinations. All four participants reported that the purpose of the mock examination preparation period was to give students

practice in answering test-like questions and to familiarise them with the requirements of the real examination, such as the question types and the level of difficulty:

This was to get schools familiar with the new exam layout when it was first implemented and to see whether or not students were well prepared for the exam. (MoE-1)

It is a useful tool to provide practice for the exam and familiarise students with the exam questions. The mock exam allows students to practice the actual exam atmosphere and assess the time required for each part of the exam paper. It gives students experience of the types of exam questions. (MoE-4)

I don't think they are useful for students. This is because students who choose this subject are already at an advanced level; but if they help them be prepared, there is no harm for students to use them. (MoE-3)

The participants revealed that implementation of the mock examination was optional for teachers, i.e. not compulsory. Only one sample of the mock examination was produced when the new examination was first implemented:

We do not write any mock exams now because schools are mostly prepared and aware of them. We also do not need to do these types of exams anymore, as students are already familiar with the questions since they have a sample of the exam from the first year of its implementation. (MoE-1)

It can be seen that the most important use of mock examinations was to familiarise students with the new format. This meant that the examination did not seem to have a strong effect on classroom teaching practices. The participants did not regard preparing students for the new examination questions as one of their major responsibilities, believing that the mock examinations should not direct students' learning for the final examination but rather provide students with practice in the new format.

According to the participants, the main content of the preparation lessons for the examination always consisted of reading, listening and writing. Teachers worked their way through the model questions in the textbook, past examination papers and other materials they developed themselves when they began intensive preparation for the examination. The students were also encouraged to work through past examination papers or practice activities in the textbook on their own:

Students and teachers should have their own resources relating to the final exam questions. Also, teachers could use the final exam templates provided in the assessment specification to design an actual mock exam. (MoE-4)

The textbook allows students to practice some questions relating to the exam specification. Also, we have asked the publishing company to list all the questions in

the textbooks, and these are similar in type and format to the final exam questions. This list is sent to the schools. (MoE-3)

Moreover, teachers were advised to offer IELTS practice for the three sections of the examination, operating under the principle that “if students could answer a sufficient number of IELTS questions correctly, they could achieve high marks in the real exam paper” (MoE-2).

In summary, through their responses, the four participants indicated that the intended washback from continuous assessment in the context of the change to the examination was largely that it would serve students’ learning. Changing the format of the examination would not change the emphasis on assessing students’ attainment in the classroom. The continuous assessment element was used to adapt students’ learning activities according to their needs. Teachers also used the different assessment tools to adjust their teaching practices in the classroom to identify students’ weaknesses and strengths (see 2.2.3).

Increasing student-centredness and improving teacher effectiveness

According to MoE-3 and MoE-1, the new elective exam was intended to promote student-centred activities and independent language learning, as the following comments suggest:

I would also like to emphasise that students are intended to become more independent learners. Teachers were advised to give students the opportunity to take ownership of their own learning, and to emphasise student-centred instruction. For example, students are given individual tasks to discuss their views in front of the whole class, or they are given assessments for a certain length of time as homework or a self-study task, so that they do not feel pressured to complete the task within the period of the lesson. (MoE-3)

The elective textbook provided students with practice activities and self-study tasks after each unit. Students were encouraged to develop their skills further in different areas of language learning. (MoE-1)

Furthermore, MoE-2 pointed that:

We advised teachers to give students homework for the activities that do not require teaching in the class. This is an elective subject, so students who select it usually exhibit the necessary motivation and interest to improve their language skills at a higher level. (MoE-2)

Additionally, there were other comments about classroom teaching in terms of both the content (“what to teach”) and the method (“how to teach”). Regarding the “what” issue, MoE-2 commented that some teachers focused on the skills and knowledge specified in the final examination because it determined the students' futures after they completed grade 12.

Those teachers felt that their students needed to develop these skills so that they could score impressive marks in the final examination:

If the writing exam assessed students on persuasive writing, teachers would teach the other types of writing included in the textbook, but their focus would be on the style concentrated upon in the exam. So, they gave more emphasis to the type of writing included in the final exam paper. (MoE-2)

By way of contrast, the other three participants mentioned that they aimed for teachers to cover all the teaching content, which involved studying the textbook and other teaching and learning resources:

We have encouraged teachers to reduce students' attention towards and their fears about the exam results, and to give instead much more emphasis to improving students' abilities. The new exam level was not above the students' expectations; it was designed to suit the outcomes required by the curriculum. (MoE-1)

Teachers are encouraged to teach the skills and not to teach just for the purpose of the exam. We also make teachers aware that this change aims to equip students with the skills required after grade 12. (MoE-4)

The good thing about this new change is that it contains a clear introduction and descriptions for all parts of the teaching content, including the exam parts. The role of the supervisors is to follow the implementation of this change ... so we give the teachers flexibility in teaching as long as their curriculum newsletters are clear and specific. (MoE-3)

The above comments show that the new exam was intended to have positive effects in terms of teaching content. Teachers' choice of content could be based on their students' needs and interests rather than on the specifications and demands of the examination. However, the term "flexibility" used by MoE-3 is interesting, as it could be variously interpreted and might entail aspects that could facilitate or hinder certain teaching practices (see Chapter 6, sections RQs1&3).

All participants reported that the teaching content of the new English elective was intended to motivate students by covering a variety of topics not just limited to Omani students' existing knowledge but designed to expand their knowledge of other cultures:

The reading materials include topics relating to university life in the UK and Europe, art in public places around the world, museums and galleries, etc. (MoE-2)

The teaching content helps increase students' motivation to use English outside of the classroom situation. (MoE-3).

Content knowledge alone is not enough to develop students' language skills, they also need appropriate language. For example, when students are taught about a new

topic, they find it a challenge to speak about this topic if they cannot demonstrate the required speaking skills as well as their content knowledge. (MoE-4)

The above examples illustrate that a major advantage of the new English elective curriculum was its relevance to the everyday world in terms of the topics and materials addressed and the skills covered in the textbook. The new topics covered in reading would help students learn about different cultures worldwide and the skills targeted would help them in future. Furthermore, according to MoE-4, students could be given multiple opportunities to demonstrate their understanding through speaking, again indicating an intended emphasis on a student-centred approach in classroom teaching.

Regarding teaching methods, MoE-4 reported that the teachers were encouraged to use more “active” or “newer” techniques than in previous times or even compared to the core diploma classes. Based on his visits to schools, he noticed that teachers were becoming more akin to “facilitators” than “dominators”, but he did not give examples to support this view, so it was unclear how this was reflected in the classroom. Similarly, MoE-3 reported that “teachers now used more communicative language tasks and urged their students to communicate and collaborate in the classroom more than under the previous curriculum”. MoE-2 noted that “professional learning communities had developed among teachers, supervisors and schools regarding teaching practices for the new elective diploma curriculum”. Other positive changes in how teachers were teaching as a result of the introduction of the new English elective examination were as follows:

We noticed a significant improvement in teaching methods after implementing this curriculum. For example, teachers are becoming more creative in their different uses of technology in classroom teaching. The use of technology in this new change aligns with students' learning preferences and the younger generations of teachers. (MoE-4)

The exam change affects what teachers do in the classroom. We noticed a big difference in the strategies used by teachers and in students' level of attainment from in the core subject. This was obvious if you compared students' results in the final exam between the first and second years of implementing the new elective. (MoE-3)

As discussed above, the new English elective examination contained features which aimed to have positive effects on classroom teaching: new topics for reading, including a variety of authentic tasks, and a focus on language skills rather than language form. There were many activities which encouraged student-to-student interactions rather than the traditional pattern of teacher explanation. However, the fact that the participants perceived a positive effect of the new examination on teaching practices does not necessarily mean that the teachers understood the underlying principles of the new English elective curriculum or that

they were able to implement them in the way that these participants intended. The next section concerns whether or not the intended changes were being implemented in classroom teaching.

5.4 Washback on Teachers' Classroom Practices

Bailey (1999), Taylor (2005), and Barnes (2017) all agree that the content and properties of a test can determine whether its washback is positive or negative, predominantly depending on whether they are in line with the principles and practices of CLT or traditional teaching (see 3.7). According to Taylor, "it is unlikely that a test based on an automated theoretical construct will lead to positive washback since, in many parts of the world, a narrow view of linguistic competence has been replaced by a broader perspective on communicative competence" (p. 276).

Given that the new elective diploma exam and textbook were designed to incorporate communicative teaching in the classroom (see 5.3), this study employed classroom observation to address the third research question: "What is the nature and scope of washback effects on teachers' classroom practices as a result of the new English elective diploma examination?" The observations aimed to examine the extent to which the MoE decision-makers' intentions in introducing the new examination were met, namely the extent to which the examination influenced (or did not influence) what was taught and how teachers conducted their teaching. It was deemed equally important to observe teachers' behaviours in the classroom and to investigate their perspectives through questionnaires (see 5.2) to address all the study aims.

Observation was conducted using Part A of the COLT scheme (Spada & Frohlich, 1995) to record the methods employed in classrooms and note the application of teaching materials in real time (see Appendix E). As discussed in Chapter 3 (see 3.8.2), Part A of the COLT scheme focuses on classroom practices, procedures and materials in second-language classrooms. The scheme combines real-time coding with transcriptions of audio recordings of educational lessons (teachers and students) to describe classroom activities. The analysis using provides information on four main categories:

- participant organisation
- lesson content
- student modality
- materials

The COLT scheme has its roots in CLT and focuses on the time spent teaching English language communicatively in terms of the four dimensions. The CLT literature considers classroom interaction and a wide range of linguistic, communicative and group work activities important factors in the development of communicative competence (Allen et al., 1984). In the context of this study, applying the scheme aimed to identify whether the English elective examination affected the “what” (content) of teaching, the “whom” (participants) in terms of interaction, and “how often”, i.e. the frequency of the teaching activity (Genesee & Upshur, 1996) as intended by the MoE decision-makers. This analysis is followed by a further analysis of the classes observed for each teacher, including how they conducted their lessons (see Appendix E).

As discussed in Chapter 4, this study examined 14 elective diploma lessons at three schools: five lessons conducted by Teacher A, five lessons conducted by Teacher B and four lessons conducted by Teacher C (see 4.6.1 for teacher descriptions). The observations were conducted in the classes of teachers who had responded to the questionnaire. I decided to observe the lessons of multiple teachers to observe whether the influence of the new examination varied between teachers and to gain a better understanding of the skills they taught and the methods they used. Notably, time constraints resulting from COVID-19 restrictions in schools during the data collection phase limited the classroom observations to four or five lessons per teacher (see 4.7.2).

The different skills the lessons focused on depended on the teachers’ lesson plans on the observation days. The observations of Teacher A and Teacher B took place in the middle of the second semester of the year (early April), when consistent teaching was already underway. Analysis of those lessons revealed similarities in terms of the content of the lessons and the textbook materials used. The observations of Teacher C took place at the end of the semester (mid-May), when intensive examination practice typically begins. This schedule was arranged based on the teachers’ instructional plans and the COVID-19 restrictions imposed on schools.

5.4.1 Analysis of classroom observations

The lessons observed were coded according to the five aforementioned COLT (Part A) categories and recordings were made of classroom activities during the lessons (see Appendix E). During the classes, I sat at the back of the classroom, making observations and categorising them.

The main units of analysis for COLT (Part A) are “Activities & Episodes”, that is, individual units that form the lesson’s teaching segments. Separate activities include one or more episodes that can correspond to different COLT (Part A) categories. For example, three episodes of one activity might be the teacher introducing a reading text, the teacher reading new words in the text, and individual students reading parts of the text aloud. To address the first COLT (Part A) category (participant organisation), calculations of the time invested in activities and episodes produced a percentage of the total lesson time that teachers and students spent on each category, enabling comparison of the average time spent on each category. This comparison enabled exploration of the three teachers’ practices in teaching the new English elective to determine whether these reflected the intended washback from the new examination.

Table 5.14 shows an example of the analysis of the classroom observations for the three teachers, representing the total time spent on each activity and/or episode for each classroom visit from the viewpoint of participant organisation.

Table 5.14. Percentage of time spent on different communicative modes (Teacher A)

Participant organisation	Lesson 1	Lesson 2	Lesson 3	Lesson 4	Lesson 5	Average
	%	%	%	%	%	%
Teacher to student/class	45	0.00	6	76	68.86	39.1
Student to student/class	0.00	50	0.00	0.00	0.00	10
Choral	0.00	0.00	0.00	0.00	0.00	0.00
Group	0.00	50	65	24	31.14	34.1
Individual	55	0.00	29	0.00	0.00	16.8
Total	100	100	100	100	100	100

Table 5.14 uses Teacher A’s lessons as an example and provides the raw data for all five lessons. For instance, during Lessons 1 and 3, the teacher spent substantially different proportions of time – 45% and 6% – addressing the class.

The following sections present the findings from the observed lessons for all three teachers, delving into the results for each COLT (Part A) category and noting any similarities and differences between the teachers.

Participant organisation

The first COLT category concerns whether the focus in the classroom is on the teacher, the whole class, a group of students or individuals. Participant organisation thus provides information regarding the nature of teaching, describing the different configurations of communication between teachers and students. Table 5.15 presents the means for the time spent on each category of participant organisation expressed as percentages.

Table 5.15. Average participant organisation as a percentage of total lesson time

Participant organisation	Teacher A Lesson Average (%)	Teacher B Lesson Average (%)	Teacher C Lesson Average (%)
Teacher to student/class	39.1	74.0	54.5
Student to student/class	10	0.00	7.9
Choral	0.00	0.00	0.00
Group	34.1	4.4	13.70
Individual	16.8	21.6	23.9
Total	100	100	100

Table 5.15 shows that type of organisation most prevalent in Teacher A's lessons was teacher to student/class, indicating a focus on teacher instruction for most of the lesson time. Moreover, the teacher spent just over one third (34.1%) of the lesson time working with students in groups, including speaking, listening and note-writing tasks. Therefore, teacher-centred and group-work activities represented over half of Teacher A's total lesson time. About a quarter of Teacher A's time was dedicated to individual work and student-to-class activities, including reading or reviewing answers to reading, vocabulary and listening activities from the textbook.

Teacher B's lessons clearly focused on teacher-centred communication (74% of the total class time), with students spending little time working together on tasks in groups (only 4.4%) compared to Teacher A's students. Individual work, which always involved reading or listening, accounted for 21.6% of Teacher B's lessons.

Teacher C followed a similar pattern, speaking to students for more than half of their lessons (54.5% of total class time). This instruction primarily focused on discussing various practice tests for the final listening examination and leading reading and writing tasks. However, the teacher devoted parts of class time to individual work (23.9%), group work (13.7%) and pair work (7.9%). In groups or in pairs, students were engaged in speaking and writing tasks. When working individually, they read or reviewed answers related to listening practice tasks.

Teacher A and Teacher C both included several activities in which students presented their work in front of the whole class. Notably, the students in both classes were more engaged in individual work than group work, potentially restricting their opportunity for engagement in discussions or debates with their peers as advocated by CLT. This might discourage students from participating in classroom communications.

Opportunities for group work activity were substantially more extensive in Teacher A's lessons, allowing students to negotiate the teaching content, discuss their work and use the

target language in meaningful ways. However, despite the expectation that Teacher B and Teacher C would follow the main principles and objectives stipulated by the new curriculum, teacher-centred communication comprised (64%) of their total lesson time during the observed lessons.

Figures 5.1 and 5.2 show the seating arrangements in the classrooms of Teacher A and Teacher C (Teacher B did not give consent).



Figure 5.1. Classroom observation: Seating arrangement (Teacher A)

Figure 1 shows that the students were involved in group-work activities. During the third observed lesson, the teacher allocated 10–20 minutes to students listening to and discussing their assigned listening task in groups before sharing their responses with the rest of the class. In this manner, Teacher A encouraged the students to talk about and practise the target language.



Figure 5.2. Classroom observation: Seating arrangement (Teacher C)

Figure 5.2 shows that the students were engaged in individual work. An example from Teacher C's first lesson involved the teacher discussing the listening questions in the final examination paper before asking students to read and review answers corresponding to the practice listening examination. During this lesson, the students mostly provided short answers, sometimes limited to "yes" or "no". Other aspects of interaction between the three teachers and their students in individual activities are discussed in 5.4.2.

Content

The second COLT (Part A) category identifies different types of teaching and learning activities and measures whether primary focus of teaching is meaning, form or a combination of the two. Observing classroom activities made it possible to establish whether teachers used traditional teaching approaches focusing on the form of the language or the communicative approach that focuses on the form and meaning of English, as intended by the MoE decision-makers.

This category includes subcategories related to classroom management and language issues and an additional subcategory that distinguishes between content related to the immediate classroom environment and student experience (narrow) and content encompassing topics beyond the student environment and classroom (broad). Table 5.16 reports the findings related to this category as a percentage of total class time.

Table 5.16. Content types as a percentage of total lesson time

Content	Teacher A Lesson Average (%)	Teacher B Lesson Average (%)	Teacher C Lesson Average (%)
Procedure only	0.00	0.00	0.00
Form – Pronunciation only	0.00	0.00	0.00
Form – Vocabulary only	15	13.6	4.89
Form – Grammar only	0.00	5	0.00
Form – Spelling only	0.00	0.00	0.00
Form – Vocabulary and grammar	0.00	4.4	0.00
Function only	11	5	4.4
Discourse	15.6	10	31.0
Narrow	8.0	0.00	9
Broad	50.4	62.0	50.71
Total	100	100	100

Teacher A focused on broad topics (50.4% of the class time), which included international celebrations, wedding traditions, life stages and formal letter writing. The teacher limited time spent on narrow topics to a brief discussion regarding a writing activity about student life timelines in Lesson 3 and descriptions of leisure time and holidays in Lesson 4. Language instruction constituted the bulk of Teacher A’s lessons. Vocabulary teaching was particularly important, accounting for 15% of class time. During this part of the class, the teacher and students spent time finding the meanings of words, phrases and collocations when encountering a new reading text. Information about written discourse represented a similarly significant language focus (over 15%). This work typically related to features of coherence and cohesion in reading activities and discourse markers in writing tasks.

Teacher B’s classes also centred on broad topics (62% of lesson time). This significant amount of class time included discussions of topics extending beyond immediate classroom concerns: traditional occupations, the economy and social entrepreneurship. The next most predominant category for Teacher B (38%) was language subcategories, with vocabulary-centred episodes the most significant among these, followed by discourse, which ensures coherence and cohesion during writing and reading activities. Other language aspects addressed during Teacher B’s lessons occupied less total class time: grammar (5%), function (5%) and vocabulary and grammar (4.4%).

Similar to Teacher A and Teacher B, discussing broad topics featured heavily in Teacher C’s lessons (50.71% of the total lesson time), especially in the form of writing, reading and listening activities (24%, 40% and 45%). Broad topics were less integral to speaking activities (11.78%). Teacher C spent only 9% of total class time on narrow topics. Examples of narrow topics included writing reports on school events and reviewing practice elective

diploma exams. The discourse category was the second-largest category (31%), followed vocabulary-centred episodes (4.89%).

All three teachers emphasised communication of meaning (“Other topics”) (59%) more than other language features (40%). Discussion of broad topics figured prominently (53%) in the lessons of all three teachers. Narrow topics were found in the lessons of Teachers A and C (6%) but absent entirely from Teacher B’s lessons. This was to be expected given that students who opt for the English elective are usually proficient users of English, precluding teachers from having to teach narrowly focused material, such as language meaning. This can also be attributed to the intention of the MoE decision-makers in introducing the new curriculum to motivate students and expand their knowledge via a variety of authentic and communicative listening, reading and writing tasks (see 5.3).

The language subcategories did not receive substantial attention (41%), although Teacher A included these more frequently than the other teachers. This suggests that the teachers invested more time in providing information about written discourse markers and vocabulary than other language subcategories. Concerning discourse teaching, Teachers A and C spent more time talking to students about written discourse than Teacher B. Vocabulary instruction demanded considerably more lesson time in Teacher A’s and Teacher B’s classes than in Teacher C’s classes. Regarding the other language features of COLT (Part A), while Teacher B addressed the grammar and vocabulary-and-grammar categories to a small degree, Teacher A and Teacher C did not expose their students to them. Based on detailed observation of the lessons, this can be explained by new topics mostly being introduced and discussed with an emphasis on unknown vocabulary, without reference to grammar or vocabulary-and-grammar elements. The students in the English elective class were also studying the core English curriculum, a compulsory subject that tests vocabulary and grammar. However, these language components are not included in the English elective examination paper, potentially leading teachers to focus less on language “form” because this is covered as part of the core curriculum.

Additionally, in all cases, the students were only exposed to a small amount of teaching of language functions. This supports the finding that students were exposed to individual work much more than collaborative activities. Notably, although there were no explanations of activity procedures in any of the lessons, the results for this subcategory do not suggest the predominance of teacher-centred activities indicated by the analysis of participant organisation.

Student modality

The fourth COLT (Part A) category, student modality, refers to the four main skills that students practise via classroom activities – listening, speaking, reading and writing – as well as other skills used in the classroom (e.g. acting and drawing). Thus, this aspect of the observation aimed to identify whether the lessons observed covered the range of student modalities intended by MoE decision-makers.

Table 5.17 presents student modality as a percentage of total class time for each of the three teachers and shows that listening was the skill most practised in Teacher A’s lessons (27.6%). The students primarily engaged in listening to broad topics during discussions of reading, speaking and writing activities and checking answers to activities in their textbook. Reading was the second most common modality for Teacher A (20.8%) and the students practised speaking skills for an average of 15.8% of total lesson time. Speaking activities involved discussions of both written discourse and broad topics.

Table 5.17. Student modality as a percentage of total lesson time

Student modality	Teacher A	Teacher B	Teacher C
	Lesson Average (%)	Lesson Average (%)	Lesson Average (%)
Listening (L)	27.6	16.2	21.75
Speaking (S)	15.8	26.2	10.75
Reading (R)	20.8	28.2	43.5
Writing (W)	0.00	0.00	24
Other skills only	0.00	0.00	0.00
L+S	11.2	0.00	0.00
L+R	0.00	0.00	0.00
L+W	3.6	0.00	0.00
S+R	11.8	29.4	0.00
L+S+W	9.2	0.00	0.00
L+S+R	0.00	0.00	0.00
Total	100	100	100

The students also performed activities that required an equal focus on speaking and reading. Such activities accounted for 11.8% of class time. Several other skill combinations appeared in Teacher A’s lessons. For example, listening was combined with speaking activities (11.2%) in lessons involving the students taking notes and presenting information to the whole class. Listening was combined with speaking and writing activities (9.2%) when the students were listening to and summarising information in group activities. Listening combined with writing activities included opportunities for the students to listen while taking notes and also to exchange information.

The most significant modality in Teacher B’s classes was the combination of speaking and reading, which represented an average of 29.4% of total lesson time. This combination

reflects how much time the students spent on reading and discussing language (such as grammar, vocabulary and discourse markers) and broad topics. Reading was the second most common modality (28.2%), closely followed by speaking (26.2%). An average of 16.2% of class time was allocated to practising listening skills, with listening activities often involving using a textbook or the teacher giving instructions and information concerning a broad topic.

Reading was the skill most practised in Teacher C's lessons, representing 43.5% of total class time. During reading activities, Teacher C spent time introducing new topics and discussing the overall ideas of texts. Writing was the second most practised activity in Teacher C's lessons, occupying an average of 24% of class time, in which it was conducted as a group-work activity. Listening accounted for an average of 21.75% of total lesson time. Although some of this involved the students considering the coherence and cohesion of report writing, they primarily listened to explanations of listening and writing skills relevant to the practice examinations and checking answers to practice examination materials. A small proportion (10.75%) of total lesson time was allocated to speaking.

Thus, reading and listening (31% and 22%) were generally more prominently practised by students than writing, speaking or combinations of skills. However, the percentage of time spent on reading and listening skills varied between the three teachers. For example, reading activities completed in Teacher C's class accounted for 44% of class time, whereas Teacher A and Teacher B allocated only 21% and 28% of time to this skill respectively. This could have been due to the different language abilities of students in the classes, which might have affected the teaching time spent on different classroom activities. Alternatively, it could have been due to differences between the teachers in terms of approach. For example, in Lesson 4, Teacher C spent 23% of the time focusing on a reading text about "changemakers", whereas Teacher B spent only 13% of total class time on the same text in Lesson 5. This suggests that the amount of reading that occurred in the lessons observed also related to the choice of reading activity and the different ways in which different teachers handled the same activity.

It should also be noted that although speaking activities appeared in all lessons (18%), speaking often corresponded to a pre-lesson activity or featured in combination with other skills (i.e. reading, listening or writing) instead of receiving overt attention. Another main finding was that less time was allocated to student writing than any other skill, representing only 8% of total lesson time. Furthermore, writing was absent entirely from Teacher B's lessons, a finding not indicative of the amount of writing typically expected of these classes.

An explanation might be that writing tasks were often given as homework, as in the case of the example from Teacher B's class provided in Figure 5.3. Homework is not included in the COLT scheme and this study did not include a record of any homework exercises assigned to students during the observed classes. Hence, the time spent on speaking and writing might be a positive indication of the influence of the new examination: it does not test speaking, but it does test writing, which can be practised through homework activities. This indicates that the amount and type of homework can significantly impact students' language learning (Hayes, 2003; Hughes & Greenhough, 2004).

Overall, the students in Teacher A's lessons used the widest variety of skills and these occurred in combinations more evenly than in Teacher B's and Teacher C's classes. However, although the student modality category provides useful information about the time devoted to particular skills and combinations of skills in classroom activities, it does not directly address the degree to which the skills employed in the classroom teaching context have a communicative basis.

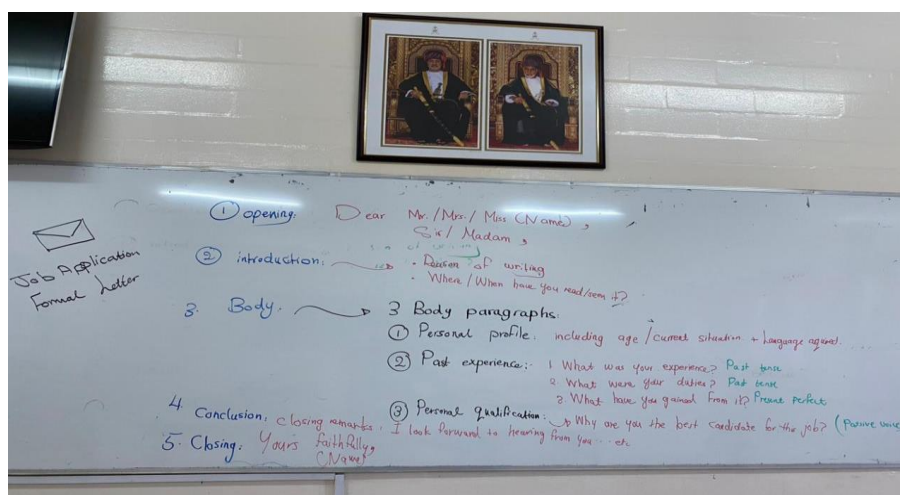


Figure 5.3. Observed writing activity (Teacher B)

Figure 5.3 depicts a writing activity about job application letters. The teacher spent the whole lesson explaining each part of the formal letter, using the whiteboard to summarise the main components of letter writing as they were discussed verbally. At the end of the lesson, students were given a letter-writing task for homework.

Materials

The final COLT (Part A) category records the different types of classroom materials used during the lessons observed in terms of the text type and the source of the material. Because the development of discourse competence relates to whether students are exposed to extended written texts rather than short standalone sentences, the text types were subdivided

into “minimal” and “extended”. Audio and video text types were also identified. The source of the material indicates the intended audience of the text. For example, source material might be an authentic text for native speakers (NS) or an instructional text designed for non-native speakers (NNS) learning a second language. Any adaptation of native-speaker materials for teaching non-native speakers was considered part of this category. This study used this category to report on the influence of implementing the new examination on teachers’ use of teaching materials.

Table 5.18 presents the mean percentages of total class time spent using different materials during each of the three teachers’ lessons.

Table 5.18. Source material types used as a percentage of total lesson time

Type of material	Teacher A Lesson Average (%)	Teacher B Lesson Average (%)	Teacher C Lesson Average (%)
Minimal L2-NNS	0.00	0.00	0.00
Extended L2-NNS	34.4	71.1	74.5
Minimal + Extended + L2-NNS	0.00	0.00	0.00
Minimal + Audio + L2-NNS	0.00	0.00	0.00
Extended + Audio + L2-NNS	7.28	10	0.00
Audio + L2-NNS	0.00	0.00	13.75
Audio + Visual + L2-NNS	45.47	3.8	0.00
Minimal + Visual + L2-NNS	0.00	0.00	0.00
Minimal + Student-made	0.00	0.00	0.00
Extended + Student-made	0.00	4	0.00
Minimal + Visual + L2-NS	0.00	0.00	0.00
Visual + Student-made	10	0.00	8
Visual + L2-NNS	0.00	0.00	0.00
Total	97.15	89	96.25

Teacher A’s lessons most frequently used audio and visual texts designed for non-native speakers (Audio + Visual + L2-NNS). Such texts accounted for 45.47% of total lesson time. Examples from this category included listening activities and discussions of broad topics. The use of extended written materials designed for non-native speakers (Extended L2-NNS), which typically involved activities such as reading passages, deciphering word meaning from long texts and working on written texts, took up 34.4% of total lesson time. Although only one example of student-made material (Visual + Student-made) appeared in Teacher A’s lesson, students worked with this type of material for 10% of the total lesson time. This was slightly more time than they spent on Extended + Audio + L2-NNS materials (7.28%).

Teacher B’s classes most commonly used extended written materials designed for non-native speakers (Extended L2-NNS). This accounted for 71.1% of total class time. This type involved using textbooks for reading and speaking activities. However, the use of extended

audio materials (Extended + Audio + L2-NNS) was observed for only 10% of total lesson time. This included listening exercises and discussion questions. Teacher B's lessons presented the same combinations of material types as observed in Teacher A's lessons. For example, both teachers included materials coded as Audio + Visual + L2-NNS. However, in Teacher B's lessons, this material category was dedicated to "listening" and the videos included information about report writing in Lessons 2 and 3 (see Figure 4.4 for a visual of one of Teacher B's lessons). Time spent with Extended + Student-made materials accounted for only 4% of total class time, with the only example a summary of a reading text about traditional occupations in Lesson 3.

Contrasting with Teacher A and Teacher B, Teacher C did not use a combination of materials in class. Instead, Teacher C most used extended written materials designed for non-native speakers (Extended L2-NNS). Such materials accounted for 74.5% of total lesson time and were used for oral discussions and reading activities. Audio + L2-NNS, observed in discussions of listening activities from the textbook, occupied 13.75% of class time, and Visual + Student-made occupied 8% of lesson time, represented by the use of a video about listening practice tests.

In broad terms, extended written materials from the subject textbook (*English Insight 3*) and a range of extended written sources (standard materials for teaching English as a second language) were most used by Teacher B and Teacher C. These materials are test-oriented, following the same format as the new examination. Moreover, the materials used in these classes were selected to meet the requirements of the examination, that is, teachers ensured they included material on making notes, mind maps, dictionary skills, using visuals and summarising information. However, combinations of material types appeared more often in Teacher A's lessons (see Figure 5.5 for a visual of one of Teacher A's lessons) than in those of Teacher B and Teacher C. This indicates that Teacher A used a wider variety of materials to integrate and practise different language skills, incorporating authentic materials (e.g. audio and video materials) from different sources. Most frequently, Teacher A asked the students to read and listen to different sources and utilise these in their textbook writing tasks. Hence, it appears that Teacher B and Teacher C drew on a more restricted range of teaching materials. This further confirms that teachers contribute substantially to mediating washback. This issue will be discussed in greater detail in Chapter 6.



Figure 5.4. Writing lesson combining material types in Teacher B's classroom

Figure 5.3 shows a writing lesson on report writing that took place during one of Teacher B's lessons. The teacher discussed structuring a report and then asked students to watch a video featuring report-writing tips. Next, students had to read a sample of a report and discuss the key information in the text.



Figure 5.5. Writing lesson combining material types in Teacher A's classroom

Figure 5.4 shows a writing lesson about writing an application letter that took place during one of Teacher A's lessons. The teacher discussed structuring an application letter using a different video for each component of the letter. Next, the teacher distributed a sample application letter and asked each group in the class to write a response to the letter. The students presented their writing in groups to the whole class.

Although language use is not mentioned by the COLT scheme, all the lessons featured materials and teaching in English. I noted that the students in the Teacher B's and Teacher C's classes were more likely to use their first language, especially during group-work activities, than Teacher A's students. Different teachers have different beliefs regarding students using their first language in the classroom, beliefs usually reflected in a teacher's

class management. However, this first-language issue might also be explained by several other factors. For example, Teacher B's and C's classes were larger than Teacher A's, providing more opportunity for students to use their common language to express and share their ideas. Additionally, larger groups are more difficult to control, meaning that even if a teacher desires an English-only classroom, monitoring first-language use is challenging (compared to small-group contexts). It is also plausible that the students in Teacher B's and Teacher C's classes possessed varying language abilities, prompting some students to use their first language to enable them to follow the lesson.

5.4.2 Further analysis of the observations

Given the complex nature of washback within the teaching context, many aspects of classroom practice require further observation. Accordingly, this section reports findings from the analysis of several significant activities that took place during the observed lessons, activities not identified by the COLT scheme, but from the field notes (see Appendix E). This analysis provides insights into the nature and scope of classroom practices arising from the implementation of the new examination in grade 12. These observations generally pertain to two main categories: exam information and strategies and teacher–student interactions.

Exam information and strategies

The provision of information about the examination and elaboration on strategies during classroom teaching produces useful insights in several ways. First, the practice indicates the degree of washback intensity in each lesson (see 3.5.1). Second, it specifies the amount of time spent introducing the exam and helping students make their test-taking more efficient and effective. Recent research studies on washback have analysed references to tests to investigate how much class time teachers spend training students on strategies directly related to the test and giving students information about test characteristics (e.g. Alderson & Hamp-Lyon, 1996; Choi, 2015; Hayes, 2003; Hayes & Read, 2004; Shohamy et al., 1996). According to Ritter and Idol-Maestas (1986, p. 350), mastery of test-taking strategies is vital for students to improve their test scores. Amer (1993) provided examples of such strategies, advising students “to read the instructions carefully, to schedule their time appropriately, to make use of clue words in the questions, to delay answering difficult questions, and to review their work in order to check their answers” (p. 71).

In the lessons observed, Teachers A, B and C referred to the examination in two different ways, both providing students with detailed information about the examination and its characteristics and giving them specific test-taking strategy tips. The analysis calculated the

time teachers spent talking about the examination or explaining strategies as a percentage of total class time during the lessons observed. Table 5.19 presents the results for each of the three teachers.

Table 5.19. References to the examination as a percentage of total class time

Reference type	Teacher A Lesson Average (%)	Teacher B Lesson Average (%)	Teacher C Lesson Average (%)	Total (%)
Exam information	0.33	1.6	5.4	2.44
Exam strategies	16	24	20.24	20.08
Total	17	25.6	26	23

Teacher A provided students with specific information about the examination for an average of 0.33% of total class time and Teacher B spent 1.6% of total class time on the topic. For a specific example, Teacher A spoke about the different writing question types in the examination while some students took notes and others asked specific questions about the examination. In other instances, Teacher A drew a link between class textbook-based activities and the examination, raising the students' awareness of examination tasks that would require reading and writing skills.

The students in Teacher B's lessons also spent a larger percentage of class time receiving tips about effective test-taking strategies than Teacher A's students (16% in Teacher A's classes, 24% in Teacher B's classes). Examples included giving tasks under examination conditions, identifying skills and strategies and evaluating student performance on exam-like tasks. These findings align with those produced by the COLT analysis, which showed Teacher B leading teacher-centred lessons (teacher to student/class) and controlling the selection of classroom activities.

Teacher C dedicated more time to providing information about the examination (5.4%) than the other two teachers. This was unsurprising given that the observations of Teacher C's classes took place closer in time to the final examination. Nonetheless, the time that Teacher C dedicated to teaching test-taking strategies (20.24%) was below (but close to) that of Teacher B. Notably, in one of Teacher C's lessons, she used her first language to provide students with tips for the listening and writing questions in the examination, demonstrating the degree of focus on providing students with exam-related information (rather than simply practising English).

Table 5.19 makes it apparent that all three teachers invested more time in teaching exam-taking strategies than providing information about the examination. These strategies warranted deeper analysis. A total of 15 exam-taking strategies were identified during the

observed lessons for the three teachers. These strategies were categorised according to the skill emphasised by the teaching content and the percentage of total class time spent discussing each strategy was calculated. The strategies were identified by their type on a lesson-by-lesson basis before being compared to determine which strategies, if any, were commonly used across the lessons of each teacher. This cross-referencing enabled me to compare the lessons of the three teachers. For example, if a teacher frequently used a strategy, the data could indicate whether that strategy was mentioned during another teacher's lesson and if so, the percentage of total class time that teacher spent discussing it.

Tables 5.20, 5.22 and 5.24 present the most observed test-taking tips as an average percentage of total class time for each teacher and Tables 5.21, 5.23 and 5.25 provide examples of use. The former demonstrate that although some strategies were common to the lessons of all three teachers, the emphasis on teaching these strategies varied. Furthermore, in most cases, especially for Teacher A, many strategies were not discussed at length or not mentioned at all.

Table 5.20. Focus on examination strategies: Teacher B (cross-referenced with Teachers A and C)

Strategy code	Strategy description	Skill	Teacher A	Teacher B	Teacher C
			Lesson Average (%)	Lesson Average (%)	Lesson Average (%)
S1	Identify main ideas	Reading	0.00	15.7	0.00
		Speaking	0.00	7.14	0.00
		Listening	9.63	10.52	0.00
S2	The importance of openings	Writing	0.00	7.47	0.00
S3	Memorise useful expressions	Reading	0.00	4.28	11
S4	Memorise sample answers	Writing	5	15	11
S5	Make an outline	Writing	15.25	7.47	17
S6	Use a range of languages	Writing	5	14.18	0.00
S7	Skim the text	Reading	6	7.14	0.00
S8	Predict the meaning of new vocabulary	Listening		3.5	0.00
		Reading	9.25	0.00	9.25
S9	Scan the text	Reading	0.00	28	9.25

Teacher B spent 24% of the total class time on nine strategies during the five lessons observed. The strategy that the teacher spent the most time discussing during class was scanning the text to develop reading skills. This was followed by identifying the main ideas to develop multiple skills at once (reading, speaking and listening). Teachers A and C differed in terms of their focus on strategies. For example, Teacher A taught S1 (identifying main ideas) to enhance listening skills, but this strategy was not used in Teacher C's classes. Memorising sample answers, making outlines of essays (writing tasks) and predicting the

meaning of new words (listening and reading tasks) were strategies commonly used by all the teachers. More concretely, each teacher advised their students to memorise and plan a structure for writing formal letters. Although Teachers A and B addressed using appropriate language in the context of teaching writing skills, Teacher C did not mention this strategy.

Table 5.21. Example of “the importance of openings” (Teacher B, writing task)

Speaker	Excerpt	Note
Teacher	What do you have in the part of opening?	Teacher B had distributed a formal letter text and asked students to read it silently.
Student 1	“Dear”.	
Teacher	We cannot use “hi” or “hello”. The only formal and associated word here is “Dear”. After that, what have you noticed? What do we have after “Dear”?	
Student 2	Mrs.	
Teacher	What else do you notice?	
Student 3	A comma after the name.	
Teacher	What do you notice from the name?	
Student 4	There is the first letter in the name is capital.	
Student 5	He is not using his first name but his sir’s name.	
Teacher	Whenever you write a formal letter entitled the name of the person Mrs, Ms, Mr, etc. and then we have to put the name. Ok ... after that, what is after the name?	
Student 6	Comma.	
Teacher	What if the letter doesn’t mention a name, to who I am going to write, which words can I use (can you remember we took this last year)?	
Student 7	Madam or Sir.	
Teacher	Yes, if don’t know to whom I am sending I will write Sir or Madam, and then comma after. So, any question[s] before I leave the opening part?	

Teacher A dedicated less time to test-taking strategies than the other two teachers. Instead, Teacher A focused on making outlines for writing tasks and predicting and identifying the meaning of new vocabulary for reading tasks. She also incorporated the notion of practice as a basic listening strategy, which introduced students to ideas about approaching listening tasks in the exam. Teacher A addressed several strategies that were absent from the classes of Teachers B and C. For example, she encouraged students to brainstorm writing topics and use their imagination to formulate information for essay writing, strategies neither Teacher B nor Teacher C addressed. However, the other strategies were common to all teachers. For

example, all the teachers advised their students to plan their essay before starting the writing process and encouraged them to memorise samples of writing tasks.

Table 5.22. Focus on exam strategies: Teacher A (cross-referenced with Teachers B and C)

Strategy code	Strategy description	Skill	Teacher A Lesson Average (%)	Teacher B Lesson Average (%)	Teacher C Lesson Average (%)
S10	Practice	Listening	5.45	0.00	8.3
		Reading	0.00	0.00	9.25
S1	Identify main ideas	Listening	9.63	10.52	0.00
		Reading	0.00	15.7	0.00
		Speaking	0.00	7.14	0.00
S11	Predict missing information	Listening	5.45	0.00	0.00
		Reading	0.00	0.00	4.34
S12	Identify the meaning of new vocabulary	Reading	8.33	0.00	0.00
S13	Read and understand the rubric carefully	Reading	4	0.00	0.00
		Listening	3.7	0.00	0.00
S7	Skim the text	Reading	6	7.14	0.00
S5	Make an outline	Writing	14.19	7.47	17
S8	Predict the meaning of new vocabulary	Reading	9.25		9.25
		Listening	0.00	3.5	0.00
S14	Brainstorm the writing topic	Writing	1.63	0.00	0.00
S6	Use a range of languages	Writing	5	14.18	0.00
S4	Memorise sample answers	Writing	5	15	11
S15	Use imagination to formulate information for essay writing	Writing	5	0.00	0.00

Table 5.23. Example: Identifying the meaning of new vocabulary (Teacher A, reading task)

Speaker	Excerpt	Note
Teacher	We will [now have] reading. Who is going to read the first passage?	The topic of the reading lesson was about the different approaches to marriage.
Student 1	(Reads the first passage aloud)	
Teacher	What about the second passage? Who is going to read it?	
Student 2	(Reads the second passage aloud)	
Teacher	I want you now to work in group[s] to find out the meaning of key words, [and] then we [will] discuss the main information about wedding traditions.	The teacher distributed iPads to each group to allow them to find the meaning of the words in bold while they read. Three minutes were given for the group work.
Teacher	I will [now check] the meaning of the words.	The teacher wrote the word definitions on the board.

Teacher C suggested nine strategies to her students and devoted a substantial amount of time to them, with these discussions occupying 26% of the total class time. Essay planning and practice represented the number one strategy in all four of Teacher C's classes. Similar to Teacher A, Teacher C emphasised the importance of practising and predicting missing information. Notably, Teacher B made no mention of these strategies. Teacher C also presented the importance of openings, memorising useful expressions and memorising sample answers as basic writing strategies, using these discussions to encourage students in

the writing process. Only the strategies of managing time and working quickly were unique to Teacher C. Interestingly, these strategies were considered in the context of several skills but particularly listening. Teacher C focused on the strategies of scanning and predicting the meaning of new vocabulary to teach reading skills.

Table 5.24. Focus on exam strategies: Teacher C (cross-referenced with Teachers A and B)

Strategy code	Strategy description	Skill	Teacher A	Teacher B	Teacher C
			Lesson Average (%)	Lesson Average (%)	Lesson Average (%)
S10	Practice	Listening	5.45	0.00	8.3
		Reading	0.00	0.00	9.25
S16	Manage time and work quickly	Listening	0.00	0.00	2
S11	Predict missing information	Reading	5.45	0.00	4.34
S5	Make an outline	Writing	15.25	7.47	17
S2	Importance of openings	Writing	0.00	7.47	11
S3	Memorise useful expressions	Reading	0.00	4.28	0.00
S4	Memorise sample answers	Writing	5	15	11
S9	Scan the text	Reading	0.00	28	9.25
S8	Predict the meaning of new vocabulary	Reading	9.25	0.00	9.25
		Listening	0.00	3.5	0.00

Table 5.25. Example: Making an outline (Teacher C, writing task)

Speaker	Excerpt	Note
Teacher	Choose one topic and think of the basic information for writing a report.	Teacher distributed a worksheet (a pre-writing plan)
Students	(List positive/and negative events, reasons behind them happening and recommendations)	
Teacher	Good. The report in the final exam will be about an event. So, you have now to do a practice on report writing about one of the optional events you have in the paper.	
Students	(In groups write their ideas for each topic)	
Teacher	Now, let's check your ideas for each topic.	Teacher explained and summarised the main structure of a report, including using sub-headings for each paragraph, using appropriate language and organising each paragraph appropriately
Teacher	I want each of you now to write a report and send it to me for checking.	

Teacher–student interactions

The analysis of the COLT scheme (see 5.4.1) coded activities as an exclusive/primary focus or combinations with equal focus. However, according to Spada and Frohlich (1995), other activities may also require attention:

It is important to note once again that the secondary focuses (that is, check marks which do not indicate either a primary/exclusive focus or combinations with equal focus) were ignored in our calculations above. We have not included these because in our work with COLT we have been more interested in those categories which are more prominent in different classroom settings. Depending on the goals of the research, it may be important to take note of these secondary emphases. (p. 116)

Several washback studies have explored common elements of teacher–student interaction, including teacher discourse (Alderson & Hamp-Lyons, 1996; Burrows, 2004; Glover, 2014; Turner, 2001), the influence of student participation and teacher feedback on student performance (Choi, 2015) and teacher’s explanations of concepts to students (Burrows, 2004; Glover, 2014; Stecher et al., 2004; Watanabe, 1996). However, this study’s purpose and context prompted teacher guidance of individual students in classroom activities to be considered a secondary focus of interaction between teachers and students. Such secondary interaction, which includes monitoring student performance or responding to student questions, has also been considered in previous washback studies to examine differences in teaching methodologies and approaches (Burrows, 2004; Hayes & Read, 2004; Wesdorp, 1982). Table 5.26 shows the average amount of this type of interaction for each of the three teachers.

Table 5.26. Teacher–student secondary interactions as a percentage of total class time

Teacher guidance	Teacher A Lesson Average (%)	Teacher B Lesson Average (%)	Teacher C Lesson Average (%)
Teacher guidance in classroom activities	6.4	0.85	0

Teacher A often spent time guiding students while they were doing class tasks individually, in groups and in pairs. Although this form of teacher–student interaction focused on issues related to language and class activities, the guidance Teacher A gave to students varied across the five observed lessons according to the type of class activity in question. Teacher A spent the most time helping students while they worked on writing tasks. In most of the writing activities, the teacher asked the students to work in groups and attempted to attend to each group as she walked around. A small amount of assistance was given during reading activities, with the teacher clarifying the reading tasks, answering questions from individual students and explaining language aspects as they occurred in the lessons.

Teacher B occasionally assisted students while they worked on class activities, whether in groups, in pairs or individually. Teacher B provided guidance and assistance for reading tasks for only 0.85% of the total lesson time. During the reading activities, the teacher clarified the tasks and corrected the students when they made pronunciation or grammatical errors. She also asked the students questions to help them discover the answers to her questions. This approach indicates that the teacher was more likely monitoring her students individually than on a group or pair basis. The teacher's questions helped students who seemed unsure what to talk about or those who had run out of ideas.

Although both Teacher A and Teacher B helped students during classroom activities in various ways, Teacher C's classes featured no interaction of this type (i.e. the teacher did not monitor students while they worked on class tasks in groups, in pairs or individually). The students did not have significant access to their teacher during any of the four observed lessons. When Teacher C's students had to work on a task, they primarily worked in groups and did not ask their teacher for explanations or assistance. This aspect of the classroom environment may be attributed to the teacher's instructional style or other factors, including the lesson topic and plan, the class seating arrangement, the time of the lessons observed and the number of students.

To sum up, there were notable differences in activity types used by each teacher. Teacher B and Teacher C spent substantial time delivering information about the exam and giving tips for taking the exam, meaning teacher-centred activity accounted for a large proportion of lesson time. However, Teacher A's lessons covered a broader range of activities and materials related to language development, with test references and test-taking strategies playing a very limited role. Teacher A was observed to interact frequently with her students while they worked individually or in groups, a mode of interaction less evident in Teacher B's classes and very rare in Teacher C's classes. As discussed, the new English elective examination was introduced to encourage a communicative, learner-centred classroom teaching methodology. However, this outcome was only observed in the case of Teacher A. From this perspective, the effect on the teaching methodologies of Teacher B and Teacher C was more negative. Thus, these findings illustrate the complexity of the washback effect on teaching. Accordingly, the next section investigates other factors that might determine an examination's washback effect, especially individual-teacher-related factors.

5.5 The Washback Effect vis-à-vis Teacher Factors

The analysis addressing RQ4 examined the washback intensity of the new examination based on teacher-related factors – namely, their gender, age, academic qualifications, years of teaching experience, in-service training, whether they were currently teaching the English elective and the number of hours per week they spent teaching the new grade 12 English elective, as these were suggested by the literature on washback to be the main factors mediating the process of washback (see 3.8). This phase of analysis assessed the effect of the stated independent variables (IVs) on two sets of dependent variables (DVs) – teachers’ teaching practices (TQ3.3, TQ2.D, and TQ4.3) and teachers’ assessment practices (TQ2.F and TQ3.2). The relationships between the variables were analysed using inferential statistics to determine whether they might contribute to the washback intensity. If the examination produced strong or weak effects only on certain teachers, the effect would likely be mediated by certain teacher factors.

This section only reports the significant results from the inferential statistics and disregards the insignificant ones, with alpha set at .05 (see Appendix Q). The following tests were used in this investigation:

1. The Mann–Whitney U test was used to compare the differences between two independent groups as the sample distributions were not normally distributed and the sample sizes were small ($N < 30$). The dependent variables were all ordinal. The Kruskal–Wallis H test was employed to determine whether the medians of two or more groups were different.
2. Independent sample t -tests were conducted to identify whether there were statistically significant differences in the mean scores for the two categorical independent groups on the IVs (i.e. gender, academic qualifications, whether currently teaching the English elective and training attended) in relation to their effect on the main DVs stated above. This test was conducted because the DV data were normally distributed for each group of IVs. Moreover, the DVs were measured on a continuous scale.
3. One-way analysis of variance (ANOVA) was employed to test the effect of the IVs with more than two levels on the DVs. This test was used to identify the overall statistically significant differences between the groups of IVs. The mean scores were compared to determine significant differences in the DVs across the stated groups.

4. Chi-squared (χ^2) tests were used to determine whether there were statistically significant relationships between two categorical variables.

The following sub-sections present the findings in response to RQ4 based on the above tests.

5.5.1 Teachers' teaching practices

In TQ4.3, independent-samples *t*-tests and one-way ANOVAs were conducted to determine if there were differences in the total number of selected learning activities based on gender, academic qualification, whether the participant was currently teaching the English elective, whether the participant had attended previous training, age and years of teaching experience. Homogeneity of variance was assessed using Levene's test of homogeneity and where there was no homogeneity of variance, the modified Welch's *t*-test and Welch's ANOVA were used. Surprisingly, there were no statistically significant differences in the total number of learning activities across all the demographic variables (all variables $p > 0.05$; see Table P1 in Appendix Q). This means that there was no effect of teachers' related factors on their classroom practices in selecting the stated teaching activities as a result of the new elective examination.

With regard to TQ3.3, a bivariate analysis using chi-squared tests of association were conducted between specific teaching practices aimed at improving students' learning and demographic variables; gender, academic qualification, whether currently teaching the English elective, previous training, age and years of teaching experience. The *p*-value indicates if there was a significant/or non-significant relationship between the two categorical variables. In what follows, the percentages of teachers using each teaching practice to improve students' learning are presented along with the results of the chi-squared tests (see Tables 5.27–5.29).

As shown in Table 5.27, there was a small, statistically significant association with gender ($\chi^2(1) = 3.968$, $p = 0.046$, Cramér's $V = 0.141$), with a significantly higher proportion of females who responded that they emphasised the skills more likely to be tested in the examination. In addition, there was a moderate, statistically significant association with age, ($\chi^2(3) = 7.954$, $p = 0.047$, Cramér's $V = 0.199$). This association refers to younger teachers aged 20–30 years old. However, there were no statistically significant associations across the remaining demographic variables (all $p > 0.05$). This indicates that teachers who were female and aged 20–30 appeared to be slightly affected in teaching the skills tested in the examination, while there was no effect from the other characteristics.

Table 5.27. Emphasising skills likely to be tested in the examination

Variable	Category	No		Yes		χ^2 test			Cramér's V
		N	%	N	%	Value	df	p	
Gender	Female	21	17.6	98	82.4	3.968	1	0.046*	0.141
	Male	24	29.6	57	70.4				
Academic qualification	Bachelor's	36	21.6	131	78.4	.516	1	0.472	0.051
	Postgraduate	9	27.3	24	72.7				
Age	20–30	3	17.6	14	82.4	7.954	3	0.047*	0.199
	31–40	21	17.8	97	82.2				
	41–50	15	28.3	38	71.7				
	> 50	6	50.0	6	50.0				
Currently teaching elective	Teaching	16	21.9	57	78.1	.022	1	0.881	0.011
	Not teaching	29	22.8	98	77.2				
Training events	Training	17	27.0	46	73.0	1.061	1	0.303	0.073
	No training	28	20.4	109	79.6				
Teaching experience	0–6	4	22.2	14	77.8	4.565	3	0.207	0.151
	7–10	11	29.7	26	70.3				
	11–15	10	14.3	60	85.7				
	> 15	20	26.7	55	73.3				

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 5.28. Skipping activities in the textbook

Variable	Category	No		Yes		χ^2 test			Cramér's V
		N	%	N	%	Value	df	p	
Gender	Female	61	51.3	58	48.7	22.493	1	0.001***	0.335
	Male	68	84.0	13	16.0				
Academic qualification	Bachelor's	106	63.5	61	36.5	.466	1	0.495	0.048
	Postgraduate	23	69.7	10	30.3				
Age	20–30	8	47.1	9	52.9	6.130	3	0.105	0.175
	31–40	76	64.4	42	35.6				
	41–50	34	64.2	19	35.8				
	> 50	11	91.7	1	8.3				
Currently teaching elective	Teaching	44	60.3	29	39.7	.897	1	0.344	0.067
	Not teaching	85	66.9	42	33.1				
Training events	Training	41	65.1	22	34.9	.013	1	0.908	0.008
	No training	88	64.2	49	35.8				
Teaching experience	0–6	10	55.6	8	44.4	1.752	3	0.625	0.094
	7–10	23	62.2	14	37.8				
	11–15	49	70.0	21	30.0				
	> 15	47	62.7	28	37.3				

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

A bivariate analysis using chi-squared tests of association was conducted between whether participants skipped over certain activities in the textbook because they were less likely to be tested in the examination and selected IVs: gender, academic qualification, whether

currently teaching the English elective, previous training, age and years of teaching experience. There was a strong, statistically significant association with gender ($\chi^2(1) = 22.493$ $p < 0.001$, Cramér's $V = 0.335$), particularly for females who answered “No”. However, there were no statistically significant associations found across the remaining variables (all $p > 0.05$). This means that teacher-related factors were not related to whether teachers “skip over certain activities in the textbook” as a result of the new examination or not, except for female teachers.

The same analysis for “I do more exam practice” showed no statistically significant associations were across the variables (all $p > 0.05$) (see Table P2 in Appendix Q). This indicates that there was no effect for doing exam preparation practices in classroom in relation to teacher-related factors (IVs).

Table 5.29. Putting more emphasis on the integration of skills

Variable	Category	No		Yes		χ^2 test			Cramér's V
		N	%	N	%	Value	df	p	
Gender	Female	73	61.3	46	38.7	.415	1	0.520	0.046
	Male	46	56.8	35	43.2				
Academic qualification	Bachelor's	10	61.1	65	38.9	1.046	1	0.307	0.072
	Postgraduate	2	51.5	16	48.5				
Age	20–30	9	52.9	8	47.1	8.980	3	0.030*	0.212
	31–40	79	66.9	39	33.1				
	41–50	23	43.4	30	56.6				
	> 50	8	66.7	4	33.3				
Currently teaching elective	Teaching	39	53.4	34	46.6	1.761	1	0.185	0.094
	Not teaching	80	63.0	47	37.0				
Training events	Training	31	49.2	32	50.8	4.044	1	0.044*	0.142
	No training	88	64.2	49	35.8				
Teaching experience	0–6	11	61.1	7	38.9	5.375	3	0.146	0.164
	7–10	24	64.9	13	35.1				
	11–15	47	67.1	23	32.9				
	> 15	37	49.3	38	50.7				

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

A bivariate analysis using chi-squared tests of association for whether some participants put more emphasis on the integration of skills than others showed a moderate, statistically significant association for younger teachers ($\chi^2(3) = 8.980$, $p = 0.030$, Cramér's $V = 0.212$). Additionally, there was a small, statistically association with having undertaken previous training ($\chi^2(1) = 4.044$, $p = 0.044$, Cramér's $V = 0.142$). No statistically significant associations were found across the remaining demographic variables (all $p > 0.05$). Overall,

the data show that younger teachers and those who had training were more likely to emphasise the integration of skills in the context of the new examination than other teachers.

With regard to TQ2.D, Mann–Whitney U and Kruskal–Wallis H tests were conducted to determine if there were differences in the level of agreement with each of the six “significant changes in teaching in response to the introduction of the English elective diploma examination” items based on gender, academic qualification, whether the participant was currently teaching the English elective, whether the participant had attended previous training, age and years of teaching experience. The results showed no statistically significant differences in agreement levels across the IVs and the items “*Teach according to the new test specifications*”, “*Adopt new teaching methods*”, “*Put more emphasis on communication practices*”, “*Put more emphasis on receptive skills*” and “*Put more emphasis on productive skills*”. For the item “*Employ more authentic language tasks*” there was a statistically significant difference in levels of agreement only between participants who were teaching at the time of collecting data for this study (mean rank = 107.09) and those who were not (mean rank = 91.26), at $U = 3598.500$, $p = 0.036$. Overall, the data show that the teachers who were teaching in the year of data collection were employing more authentic language tasks than those who were not in response to the introduction of the examination, while there was no relation between other teacher-related factors and the five significant changes in teaching (see Tables P3–P8 in Appendix Q).

To sum up, the above paragraphs have addressed the washback intensity of the new examination by considering the relationship between teacher-related factors and the teachers’ stated teaching practices. The analysis was based on the teacher questionnaire items related to their attitudes to their teaching practices and their response to the introduction of the new examination (see TQ3.3, TQ2.D and TQ4.3). The data show no statistically significant differences between teacher-related factors and the learning activity items that teachers believed were important in teaching the English elective classes. Teaching the skills tested in the examination, skipping over certain activities in the textbook and emphasising the integration of skills appeared to show statistically significant associations at varying levels with teacher-related factors. Nonetheless, these were less likely to result in washback intensity manifest through an emphasis on exam-related aspects in classroom teaching (see Tables 5.27–5.29). Finally, those teachers who were teaching in the year of data collection reported employing more authentic language tasks as a change in their teaching, but there were no statistically significant differences in relation to the other significant changes in teaching as a result of the new examination.

5.5.2 Teachers' classroom assessment practices

Regarding TQ2.F, Mann–Whitney U and Kruskal–Wallis H tests were conducted to determine if there were differences in the levels of agreement for each of the six “function of mock tests” items based on gender, academic qualification, whether the participant was currently teaching the English elective, whether the participant had attended previous training, age and years of teaching experience. The results are presented separately for each of the six items and Tables 5.30-5.35 provide a summary of the participant responses to the six items and respective inferential tests.

Table 5.30. Giving feedback to students

Variable	Category	Median	Mean rank	U	z	p
Gender	Female	4	101.35	3937.500	-1.547	0.122
	Male	4	90.31			
Academic qualification	Bachelor's	4	97.01	2443.500	-0.006	0.995
	Postgraduate	4	96.95			
Currently teaching elective	Teaching	4	111.94	3259.500	-3.233	0.001**
	Not teaching	4	88.50			
Training events	Training	4	99.77	3789.500	-0.528	0.598
	No training	4	95.78			
Variable	Category	Median	Mean rank	χ^2	df	p
Age	20–30	4	94.64	4.337	3.000	0.227
	31–40	4	97.33			
	41–50	4	102.82			
	> 50	4	69.45			
Teaching experience	0–6	4	96.63	.232	3.000	0.972
	7–10	4	99.31			
	11–15	4	94.93			
	> 15	4	97.85			

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

As can be seen from Table 5.30, there were statistically significant differences in the levels of agreement between participants who were currently teaching the English elective (mean rank = 111.94) and those who were not (mean rank = 88.50), at $U = 3259.500$, $p = 0.001$. This indicates that this factor affected teachers' views of the function of mock examinations as a means of giving student feedback. There were no statistically significant differences in agreement levels across the other variables (all $p > 0.05$).

With regard to the function of the mock examinations related to the item “*To motivate students*”, the results show no statistically significant differences in agreement levels across

the teacher-related variables (all $p > 0.05$) (see Table P9 in Appendix Q). Hence, there was no link with teachers' view of the mock examination as a way of motivating students.

Table 5.31. Preparing students for the real examination

Variable	Category	Median	Mean rank	U	z	p
Gender	Female	4	95.06	4673.000	0.668	0.504
	Male	5	99.99			
Academic qualification	Bachelor's	5	97.60	2348.000	-0.385	0.700
	Postgraduate	4	93.77			
Currently teaching elective	Teaching	5	110.29	3375.000	-2.780	0.005**
	Not teaching	4	89.44			
Training events	Training	5	103.19	3588.000	-1.139	0.255
	No training	4	94.28			
Variable	Category	Median	Mean rank	χ^2	df	p
Age	20–30	4	83.93	2.08	3.000	0.556
	31–40	4	96.58			
	41–50	4	103.34			
	above 50	5	88.73			
Teaching experience	0–6	4	86.31	2.872	3.000	0.412
	7–10	5	91.31			
	11–15	5	94.85			
	> 15	4	104.33			

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 5.31 addresses differences in the level of agreement for the function “*To prepare students for the real examination*”. There was a statistically significant difference in the level of agreement between participants who were currently teaching the elective (mean rank = 110.29) and those who were not (mean rank = 89.44), at $U = 3375.000$, $p = 0.005$. There were no statistically significant differences in agreement levels across the remaining variables (all $p > 0.05$). This indicates that only those teachers who were teaching the elective in the year of data collection appeared to be affected by the view that the function of the mock examination was to prepare students for the final examination.

Table 5.32. Identifying areas for re-teaching

Variable	Category	Median	Mean rank	<i>U</i>	<i>z</i>	<i>p</i>
Gender	Female	4	100.97	3981.500	-1.336	0.182
	Male	4	90.89			
Academic Qualification	Bachelor's	4	93.43	3027.500	2.259	0.024*
	Postgraduate	4	116.42			
Currently Teaching	Teaching	4	108.13	3526.000	-2.277	0.023*
	No teaching	4	90.67			
Training Events	Training	4	100.64	3738.500	-0.654	0.513
	No training	4	95.40			
Variable	Category	Median	Mean rank	χ^2	<i>df</i>	<i>p</i>
Age	20–30	4	78.32	4.058	3.000	0.255
	31–40	4	99.29			
	41–50	4	101.14			
	> 50	4	77.27			
Teaching experience	0–6	3.5	90.56	1.370	3.000	0.713
	7–10	4	89.77			
	11–15	4	99.28			
	> 15	4	99.99			

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 5.32 presents the results of the analysis identifying differences in the level of agreement for the item “*To identify areas for re-teaching*”. Again, there was a statistically significant difference between the participants who were currently teaching the elective (mean rank = 108.13) and those who were not (mean rank = 90.67), at $U = 3526.000$, $p = 0.023$, as well as between those with a Bachelor’s degree (mean rank = 93.43) and those with a postgraduate degree (mean rank = 116.42), at $U = 3027.500$, $p = 0.024$. The reason teachers’ qualifications would be a factor in decisions about re-teaching could be that teachers who hold higher qualifications in Oman have longer teaching experience and more opportunities to participate in training courses. More experienced teachers (those with higher levels of qualifications and more training) would be more likely to understand the functions of mock examinations as proposed by the MoE decision-makers than less experienced teachers (see 3.8.1). There were no other statistically significant differences in agreement levels across the remaining variables (all $p > 0.05$). This indicates that the teacher’s level of qualification and whether they were currently teaching the elective affected their views of the purpose of mock examinations being to identify areas for re-teaching.

Table 5.33. Getting students to pay attention in class

Variable	Category	Median	Mean rank	<i>U</i>	<i>z</i>	<i>p</i>
Gender	Female	4	103.55	3680.000	-2.175	0.030*
	Male	4	86.92			
Academic qualification	Bachelor's	4	97.94	2291.000	-0.590	0.555
	Postgraduate	4	91.87			
Currently teaching elective	Teaching	4	104.99	3746.000	-1.613	0.107
	Not teaching	4	92.46			
Training events	Training	4	111.92	3072.500	-2.652	0.008**
	No training	4	90.43			
Variable	Category	Median	Mean rank	χ^2	<i>df</i>	<i>p</i>
Age	20–30	4	84.32	1.801	3	0.615
	31–40	4	100.20			
	41–50	4	92.19			
	> 50	4	101.45			
Teaching experience	0–6	4	83.78	3.849	3	0.278
	7–10	4	91.31			
	11–15	4	106.22			
	> 15	4	94.15			

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 5.33 concerns differences in the level of agreement for the function “*To get students to pay attention in class*”. The level of agreement was statistically significantly different between females (mean rank = 103.55) and males (mean rank = 86.92), at $U = 3680.000$, $p = 0.030$, as well as between those who had received previous training (mean rank = 111.92) and those who had not (mean rank = 90.43), at $U = 3072.500$, $p = 0.008$. There were no statistically significant differences in agreement across the remaining variables (all $p > 0.05$). This shows that gender and training were the only factors that affected teachers’ views of the use of the mock examinations to attract students’ attention in class.

Table 5.34. Encouraging students to study regularly

Variable	Category	Median	Mean rank	<i>U</i>	<i>z</i>	<i>p</i>
Gender	Female	4	97.49	4389.000	-0.167	0.867
	Male	4	96.25			
Academic qualification	Bachelor's	4	97.01	2444.000	-0.004	0.997
	Postgraduate	4	96.97			
Currently teaching the elective	Teaching	4	110.76	3341.500	-2.867	0.004**
	Not teaching	4	89.17			
Training events	Training	4	102.46	3631.000	-1.000	0.317
	No training	4	94.60			
Variable	Category	Median	Mean rank	χ^2	<i>df</i>	<i>p</i>
Age	20–30	4	93.79	.132	3.000	0.988
	31–40	4	96.83			
	41–50	4	98.69			
	> 50	4	95.05			
Teaching experience	0–6	4	101.72	.239	3.000	0.971
	7–10	4	95.15			
	11–15	4	95.99			
	> 15	4	97.85			

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 5.34 provides the results concerning differences in the level of agreement for the function “*To encourage students to study regularly*”. The only statistically significant difference in levels of agreement was between participants who were currently teaching the elective (mean rank = 110.76) and those who were not (mean rank = 89.17), at $U = 3341.500$, $p = 0.004$ (all other $p > 0.05$). Hence, the only factor that contributed to teachers’ view of mock examinations as a way of encouraging students to study was whether they were teaching the elective classes in the year of data collection or not.

For TQ3.2, a bivariate analysis using chi-squared tests of association was conducted to establish whether participants were more or less inclined to use specific classroom assessment methods because they were less likely to be tested in the examination based on the following variables: gender, academic qualification, whether the participant was currently teaching the elective, whether the participant had attended previous training, age, and years of teaching experience. The results revealed no statistically significant associations between the reported use of day-to-day observation, presentations and quizzes as assessment tools in classroom teaching and the demographic variables (see Tables P10, P11 and P12 in Appendix Q; all $p > 0.05$).

With regard to written work, however, there was a moderate, statistically significant association with gender ($\chi^2(1) = 6.18, p = 0.010$, Cramér's $V = 0.182$). Additionally, there was a moderate, statistically significant association with age ($\chi^2(3) = 14.813, p = 0.002$, Cramér's $V = 0.272$). There were no statistically significant associations across the remaining demographic variables (see Table P12 in Appendix Q; all $p > 0.05$). There were no statistically significant associations for quizzes (see Table P13 in Appendix Q; all $p > 0.05$). However, there was a small significant association between class tests and whether or not the participant had attended previous training ($\chi^2(1) = 4.305, p = 0.038$, Cramér's $V = 0.147$) (see Table P14 in Appendix Q). There were no other statistically significant associations (all $p > 0.05$).

To sum up, the above findings concerned the effect of teacher-related factors on teachers' reported views of assessment and their practices. Levels of agreement regarding the functions of the mock examinations (except to motivate students) differed significantly in relation to several factors (see Tables 5.30–5.35). Moreover, certain teacher characteristics affect their uses of class tests and written work as main tools in classroom teaching, but not their use of classroom observation, presentations and quizzes in the classroom (see Tables 5.27–5.29).

5.6 Conclusion

This chapter has focused on the findings from phases two and three of the study, obtained from a variety of different sources. The findings have been presented based on four primary themes related to the research questions posed by the study, as follows:

- Teachers' perceptions of and reactions to the new examination.
- The intended washback of the new English elective diploma examination proposed by the MoE decision-makers.
- Washback on teachers' classroom practices.
- The nature and intensity of the washback effect vis-à-vis teacher factors.

First, the findings show that identifying teachers' perceptions of and reactions to the new examination is complicated. The questionnaire and interview findings revealed substantial differences in teachers' reactions. Although the questionnaire findings indicated overall negative reactions among teachers towards the new examination design, the interviews showed that these feelings changed for the better a while after the introduction of the examination. Their initial negative reactions were due to the difficulty of processing so much information concerning the change initially, when there was not enough guidance, practice,

or teacher training available. Moreover, there were concerns and worries over the methods of teaching the new syllabus, the form and volume of the teaching material and the students' current levels of English. Teachers also mentioned that the process of introducing the exam was not communicated sufficiently at the micro level; they only received news about the examination after it had been introduced, which made it difficult for them to prepare their students for the change. Notwithstanding these difficulties, as the teachers became more familiar with the contents and format of the new examination, their perceptions of its quality increased and viewed the new examination in a more positive light.

It seemed that the teachers were more willing to change their teaching content and activities than their teaching methods. The examination seemed to be having a strong effect on teaching activities, particularly in terms of the amount of attention teachers were paying to the skills tested in the examination as opposed to those not tested. Moreover, there was a tendency for teachers to focus on activities directly related to the requirements of the new examination. When aspects of classroom assessment practices were explored from the teachers' perspectives, it seemed that they preferred to give students extensive practice in examination questions in their classes. The examination seemed to be having a strong effect on continuous assessment, especially on the choice of assessment method, the way this was designed and the amount of attention teachers paid to listening, reading and writing over speaking skills.

Second, the participants suggested that the inclusion of English language skills more broadly in the new examination would have favourable washback on teachers' practices in the classroom, mentioning several ways in which the new examination differed from the old one. They saw most of these features (content, question types, examination requirements and difficulty level) as positive changes. Moreover, the MoE participants reported that the intended washback from introducing the examination was to promote the practice of continuous assessment for learning. This was intended to develop students' learning strategies and their ability to use English effectively. The analysis also showed that the MoE participants understood the examination preparation period to be essential for students' success in the final examination, but they did not wish there to be too great an influence on what teachers did in the classroom.

The changes in teachers' classroom behaviours as a result of the introduction of the new examination were divided into two kinds: the "what" (content) element and the "how" (methodological) element. Regarding the former, some of the MoE participants believed that

teachers would likely pay more attention to the content that might be directly related to the new examination, whereas others suggested that teachers could and did use a wider variety of materials to integrate and practise different language skills by incorporating authentic materials from different sources. In terms of teaching methods (“how”), the view was that these represented an improvement compared with the previous examination. Additionally, the findings from the MoE participants indicated a greater focus on a student-centred approach arising from the examination change.

Third, the findings demonstrated the washback effect on classroom activities and interactions, showing that the participants’ organisation of classroom interaction shared similar characteristics across the different classes observed, generally comprising teacher-centred activities. However, Teacher A’s lessons were less teacher-controlled and consequently featured more significant student activity than the classes of Teacher B and Teacher C. Segmenting the observed lessons into content revealed that all three teachers predominantly emphasised the communication of meaning rather than other aspects of language use. Although reading and listening were the skills most prominently practised, the time spent on them varied between the three teachers. In terms of the materials presented in the classroom, Teacher B and Teacher C drew heavily on written materials from the textbook, whereas Teacher A supplemented the textbook with audio and visual materials from various authentic sources. Moving beyond the COLT scheme, further analysis of the observations indicated that although all three teachers addressed the English elective examination specifically, they spent more lesson time discussing exam-taking strategies more generally. Regarding teacher–student secondary interactions, both Teacher A and Teacher B spent some time asking students whether they needed help or assistance when they were working on class activities, but this interaction pattern was completely absent in Teacher C’s classes.

Finally, the results reported in this section revealed that teacher characteristics could not be entirely considered intervening variables in mediating washback and may not influence the direction and intensity of the washback effect of the new English elective examination. There does not seem to be a strong linear relationship between teacher-related factors and their practices in the classroom vis-à-vis the washback effect of the new examination. Indeed, there were no statistically significant differences in the relationships between the independent variables and the main dependent variables, but there were some teacher-related factors that mediated and interacted with the examination resulted in a washback effect on classroom teaching practices (see 5.5). Interestingly, inferential statistics showed some differences in the teachers’ classroom practices according to their characteristics.

The analysis showed that although identifying the washback effect from the new English elective examination was complex, triangulating the data from various sources and methods contributed to unveiling what is happening in the Omani classroom setting, namely what kind of washback the new English elective examination has induced and how it operates. The following chapter further synthesises the key findings from both the qualitative and quantitative data strands and relates these to the pertinent literature.

Chapter 6. Discussion

6.1 Introduction

This chapter reviews the research findings within the context of the existing literature and established knowledge regarding the impact of introducing a high-stakes examination on teachers' perceptions and their classroom practices. It synthesises and interprets the findings reported in Chapter 5, with a particular focus on the four research questions:

- RQ1. What are teachers' perceptions of the new English elective diploma examination?
- RQ2. What is the intended washback from the new English elective diploma examination according to MoE decision-makers?
- RQ3. What are the nature and scope of the apparent washback effects resulting from the new English elective diploma examination on teachers' classroom practices?
- RQ4. How does the intensity of the apparent washback effects from the new English elective diploma examination differ according to teachers' personal characteristics?

These questions sought to investigate the washback effect of the new English elective examination in Oman by: (a) exploring teachers' perceptions of and reactions to the new examination; (b) investigating the extent to which the perceived effects of the new examination have created a washback effect on teacher's classroom practices; (c) identifying the intended washback of the examination as envisaged by the decision-makers and test developers in the MoE; (d) examining the ways in which teacher characteristics (e.g. teaching experience, gender and age) potentially affect the intensity of apparent washback from the examination.

The following sections discuss the key findings of this research through the lens of the conceptual framework adopted (see 3.10.2) and previous empirical studies of washback reviewed in Chapter 3 (see 3.6–3.9). These sections outline the contribution of this research to knowledge, theoretically and methodologically, documented in the existing literature on the washback of testing. The limitations and potential consequences of the research are also discussed, as well as the implications of the interpretation of the research findings.

6.2 Synthesis of Research Findings

6.2.1 Teachers' perceptions of the new examination

Empirical research has provided compelling evidence of washback on teachers' attitudes, often indicating a conflict between what teachers think are effective teaching methods and how they feel they are required to "teach to the test" (Smith et al., 1991, p. 41). However, as

noted in Chapters 2 and 3, there is an apparent overlap in washback studies between the way teachers teach and their views about how to teach since research has commonly relied on self-report data as evidence of teaching practices (Pan & Newfields, 2012; Shih, 2013; Stecher et al., 2004; Wall & Alderson, 1993; Watanabe, 2004; Yu, 2018), rather than as evidence of attitudes. In this study, therefore, different approaches to triangulation were used – data and methodological – to adequately address the research questions and increase the representativeness and quality control of the project (see 4.4.2).

The purpose of investigating teachers' perceptions in this study was identify the potential relationship between teachers' beliefs about teaching and learning and their classroom teaching practices, as suggested by the literature (e.g. Cheng et al., 2004; Norris, 2004; Winke, 2011) and the framework for this study (see 3.10.2). What teachers think and believe about a specific examination and how familiar they are with what it aims to assess relate closely to how and what they teach, and vice versa (Barnes, 2017; Borg, 2005; Chappell et al., 2015; Cheng, 2005; Choi & Lee, 2017; Dammak et al., 2022; Hamp-Lyons, 1997; Lai & Waltman, 2008; Spratt, 2005; Watanabe, 2004). Therefore, teachers' perceptions of their teaching in relation to a new examination ought to be assessed in isolation from their actual practice in the classroom. The findings pertaining to RQ1, concerning the impact of the new examination on teachers' perceptions, addressed through the questionnaire and interviews conducted in phases two and three, also made it possible to answer of RQ2, RQ3 and RQ4 (see 6.2.2, 6.2.3 and 6.2.4). Three major themes were explored in the teacher questionnaire and interviews, as follows:

- Teachers' perceptions regarding the new English elective examination.
- Teachers' perceptions regarding their classroom teaching behaviours.
- Teachers' perceptions of their assessment practices.

Teachers' perceptions regarding the new examination

Reflecting on this first theme, it can be argued that the new examination had a positive impact on teachers' attitudes. The results suggest that the teachers felt the intended goals of the new diploma were helpful for students' learning and several teachers talked about the positive influences of the examination on their classroom teaching. The findings support advocates of the concept of MDI (Popham, 1987), arguing that tests perceived as important as a result of the consequences associated them would achieve educational reform goals by driving teaching and learning (see Frederiksen & Collins, 1989; Fournier-Kowaleski, 2005). Popham (1987) argued that high stakes testing would affect teachers' attention in a way that

made the tests serve as a “curricula magnet” (p. 680). Popham’s (1987) argument for MDI was that it is important the constructs comprising the goals and objectives of the curriculum, or the knowledge and skills expected to be achieved in a course, are what the tests direct teachers to teach when preparing for the tests. However, the findings of this study suggest that teachers' views of the beneficial aspects of the new examination, while significant, are not sufficient evidence of teachers' actual practice. Although the teachers said that the new examination had many positive features, making it an improvement on the previous version, it was not clear that it would encourage them to pay equal attention to all the language skills and make full use of the teaching strategies introduced in the new textbook, particularly because speaking, grammar and vocabulary were not tested directly, but only covered in the textbook.

This is further supported by the findings of previous washback studies, which have shown that the materials available to teachers for teaching and the goal-oriented nature of a textbook play a significant role in how teachers teach (e.g. Alderson & Hamp-Lyons, 1996; Barnes, 2017; Cheng, 2005; Kılıçkaya, 2016; Read & Hayes, 2004; Tsagari, 2011). Barnes (2017) found that teachers' reliance on TOEFL iBT preparation courses and textbooks, which only covered test content, led to a teacher-centred approach to teaching that focused on individual responses. This was appropriate for test-taking but not for developing language proficiency. In this research, the results suggest that any changes brought about in teaching might not simply be due to the new examination since it does not test all the skills covered in the new textbook (speaking, grammar and vocabulary). The potential positive effects on teaching may also be due to the new textbooks and the promotion of certain knowledge and skills in introducing the new English elective curriculum. Considering the essential role that textbooks have played in teaching English in this context (see Chapter 2), it is likely they affect the process of MDI and “produce varying and non-uniform instructional consequences” (Airasian, 1988, p. 8).

Regarding teachers' reactions to the new English elective examination, the results of the questionnaire showed significant levels of agreement. The most interesting finding was that while most teachers (88%) were not sceptical about the need for a new examination, the majority (79%) did not actually endorse the change. The findings from the questionnaire were further explored in the responses elicited from the interviews with the five teachers, in which their negative reactions to the new examination were seen to relate to the initial stage of implementation when they encountered various challenges (see 5.2.4).

A significant change to the new curriculum is that reading, writing, speaking and listening are incorporated in the teaching plan. Teachers tend to rely on textbooks and other curriculum materials, such as workbooks or teachers' guides, for their lesson plans. The textbooks published by the MoE are frequently used as curricula, guiding teachers on what to teach. The new English elective examination was meant to reflect the learning outcomes identified in the textbook and to reinforce them; however, if teachers found it difficult to understand what they were supposed to teach, the examination would not make this clearer. These findings are in line with the expected effects of the changes to the examination, as it was anticipated that these would make teachers feel uncomfortable and under pressure when they were first introduced. This corroborates Fullan's (2015) view that the process of change takes a long time and that there will always be many different interpretations of the intentions behind a change when not enough time has passed for it to assume its final form.

To understand the teaching context more fully, it was also necessary to examine teachers' perceptions of the possible challenges arising from the implementation of the new examination design and demands. Findings from the questionnaire and the interviews revealed some difficulties caused by the new elective exam requirements. One of the main challenges was the focus on authenticity, both in terms of aligning this approach with teachers' abilities in employing teaching strategies to enhance students' communication and the perceived value of using authentic materials in classroom teaching. Green's (2007) washback model suggested that the greater the overlap between the target skills or focal construct and the real-world demands a learner faces, the greater the potential for positive washback (see 3.4.4). The teachers in this study were aware of the overlap but expressed opposing views on the need for authenticity. This was because the exam necessitated extra authentic materials for reading and writing, which increased their workload and created extra pressure.

Fifteen of thirty open-ended responses and five teacher interviews revealed the difficulty of teaching the new textbook as mandated by the MoE. The teachers felt overwhelmed by their already packed schedules, extra duties and the increased workload caused by the change. They were concerned that there was insufficient time to cover the elective textbook as instructed by the MoE decision-makers since the textbook is considered the main source of instruction in this context (see Chapter 2), although it was unclear if this was only a time issue or if other factors were involved (see 5.2.4 and 5.2.5). This is confirmed by other washback studies, which have found that teachers can have negative attitudes towards a new test and how to teach the new textbooks, resenting the extra administrative work entailed by

the test (Hawkey, 2006; Wall, 2005) and the time pressure they experience when teaching to the test (Cheng, 2005; Ferman, 2004; Shih, 2009; Shohamy et al., 1996; Wall, 2005). The pressure of insufficient teaching time suggests a degree of washback anxiety in terms of what teachers do and how they teach, as has been pointed out by several washback studies: teachers feel anxious about covering the material needed for a test (Cheng, 2005; Sevimli, 2007; Sukyadi & Mardiani, 2011; Tsagari, 2009; Turner, 2009; Wall, 2005; Yıldırım, 2010), they invest extra time and effort to increase their students' scores (Ferman, 2004; Kılıçkaya, 2016; Shohamy et al., 1996; Tsagari, 2009; Zhan & Andrews, 2014), and they emphasise the content of the test through "curriculum narrowing", thus reducing the scope and content of their teaching and lessons (Andrews, 2014; Madaus, 1988; Messick, 1996; Shohamy et al., 1996). Although various washback studies have highlighted anxiety and its debilitating (or facilitating) effects on teachers and students in teaching and learning, the extent to which test anxiety affects teaching is worth further investigation in relation to washback. In this study, the proposed conceptual model (see 3.10.2) highlights that the implementation of a new examination is often associated with other changes in the educational system, such as changes to the textbooks, teaching materials, classroom assessment objectives and practices, teaching workload, etc., all of which have an influence on the process of teaching.

Additionally, the educational context poses significant challenges. The interviewees highlighted two main issues: teacher training and communication with decision-makers from the MoE. The primary concern with teacher training was the absence of guidance on what was expected when teaching the new curriculum and preparing students for the new examination format. Some teachers criticised the method used to introduce changes to the examination, pointing out the lack of communication about training opportunities in their area. Communication problems arose due to the MoE decision-makers' lack of clarity regarding their decisions about the examination design and implementation, as well as the distribution of the exam specification handbook and mock examinations at a later stage. Teachers reported that they were informed of the introduction of the new examination shortly before its implementation and that no training was provided during the development process. Several teachers were nominated by their supervisors in each directorate to attend a series of meetings and seminars organised by the MoE decision-makers regarding the change and introduction of the new examination. These teachers were asked to cascade the discussions of the content and requirements of the new examination to their colleagues at their schools (see 5.2.7). Two participants in the MoE explained that one of the common features of these seminars was that the curriculum development team would provide full support for the

teachers regarding the changes made to the new elective diploma. During these seminars, the teachers were given a teacher's book, teaching plan, teaching resources, sample practice activities and tasks, and so on. The two MoE participants noted that the curriculum development team members are involved whenever there is a curriculum or assessment initiative as assessment changes follow changes in teaching content.

As Spolsky (1994) pointed out, the washback effect of a test is beyond its designers' control and unlikely to happen in the particular manner they intended, although the reasons for this can vary due to the complexity of the factors involved in different educational settings. Based on Bailey's (1996) washback model and the conceptual framework for this study, I would argue that it is essential for policymakers and decision-makers to communicate all the relevant information to teachers to ensure that the introduction of an examination yields the expected outcomes, as has also been suggested by the findings of previous studies (East, 2016; Fan et al., 2020; Winke, 2011; Zhang, 2021). This also reflects the study by Fan et al. (2020), which found that much of the discrepancy between teachers' perceptions about a new test (TEM4) and the intentions of the developer may have come about because the teachers were not given enough information to understand the intentions behind the reforms. Jin (2010) also suggested that policymakers and teachers should ensure there are clear communication channels to address discrepancies between teachers' perspectives and methods and encourage effective and equitable assessment practices.

The above points suggest that the teachers had mixed attitudes when it came to the newly introduced examination, but they were more critical than positive. Their predominantly negative attitudes can be ascribed to the challenges and difficulties arising from implementing the new curriculum in their teaching as the revised English elective curriculum was still new and the teachers had only a short time to get used to the many changes required. As part of this process, teachers encountered considerable difficulties, which Spratt (2005) referred to as "a tension between pedagogical and ethical decisions" (p. 24). They could teach as they believed appropriate based on their philosophy of what constituted genuine learning or follow the compulsion to teach to the test to help their students succeed in the examination. Although the teachers acknowledged that their teaching methodologies and strategies should change to suit the amendments to the curriculum and examination, as well as their students' needs, their teaching methods remained largely unchanged and few actually implemented the intended change in classroom teaching.

What is more, teachers in this context are not normally faced with having to modify their teaching practices in the classroom to embrace new content and strategies. Regardless of their attitudes towards the reform, it was only to be expected that they would have to change their previous way of teaching to prepare their students properly for the important end-of-year examination. The findings of this study are consistent with previous empirical washback studies (e.g. Gipps, 2011; Madaus, 1988; Shohamy, 1996; Smith, 1991; Vernon, 1956; Wiseman, 1961), which have shown that there is a high level of pressure associated with high-stakes tests – especially newly introduced ones – and consequently an influence on teachers’ attitudes and practices in that “they increase teachers’ stress and lower their morale” (Abu-Alhija, 2007, p. 57). Further discussion of the teachers’ instructional practices is presented in the section on classroom observation (see 6.2.3).

Teachers’ perceptions regarding their classroom teaching

Regarding the second theme concerning teachers' perceptions of the impact of the new examination on classroom instruction, the questionnaire responses indicated unanimous agreement on the adjustments teachers recognised as necessary in response to the revised examination. For example, most teachers agreed that they would like to pay more attention to productive skills (60%), although fewer agreed on the importance of emphasising receptive skills (40%). In practice, the teachers tended to focus on the skills tested in the examination and thus did not spend much time teaching speaking, as this was the one skill not tested. Indeed, the findings showed that teachers spent more time on reading and writing than oral skills, covering as many textbook activities and as much supplementary material as possible. This finding broadly supports the comment by Weir (2005) that “teachers may simply not teach certain important skills if they are not in the test” (p. 18). Similar findings were reported by Wall and Alderson (1993), Wall (2005) and Salehi et al. (2012), as evidence of a negative washback effect in the way teachers concentrated on the skills that were tested in examinations. This study showed a similar narrowing of the range of language skills taught to include only those that were tested in the new examination (reading, writing and listening), thus also indicating a negative washback effect (see 6.2.3). This further confirms the views of researchers who report that the washback from testing is more likely to be negative than positive (e.g. Davies, 1968; Fish, 1988; Vernon, 1959).

Another indication of curriculum narrowing was that the questionnaire responses revealed that most teachers (80%) felt they had to prioritise the skills most likely to be assessed in the examination. This was further elaborated upon in the interviews, in which three teachers reported that they were neglecting certain types of texts or activities that were less likely to

be tested – such as any writing text types not covered – and were focusing instead on exam-style reading questions. The teachers pointed out that their lesson preparation plans had to emphasise the exam specifications to a certain extent. This finding is compatible with other studies from different contexts (Abu-Alhija, 2007; Agrawal, 2004; Azadi & Gholami, 2013; Choi, 2008; Stecher et al., 2004), which suggested that narrowing of the curriculum can create an exclusive focus on test content rather than on the overall objectives of the curriculum. For example, Chen's (2006) study found that since the Basic Competence Test (BCT) failed to reflect the objectives of the curriculum, the intended washback of promoting a communicative approach in teaching was not achieved. Similarly, this study indicated evidence of curriculum narrowing, in which the introduction of the examination led teachers to focus only on the specific genres and texts being examined. The consequence of this is that negative washback is likely to take place because of teachers' tendency to spend time on "cramming content". This confirms the point made by Wall and Alderson (1993), that "A test will influence what teachers teach" (p. 120). Thus, this study's findings suggest that if the examination specifications were better aligned with curriculum outcomes, it would create beneficial washback and the tensions and negative washback on classroom teaching and learning would be reduced, as also suggested by Resnick and Resnick (1992). Moreover, the is in line with the findings of other studies conducted by Orafi and Borg (2009), Hoque (2011) and Rahman et al. (2021), who showed that a discrepancy between textbook objectives and test objectives could impair teachers' comprehension and knowledge of the curriculum objectives, thus resulting in negative washback.

When the interviewees were asked about their views on the impact of the newly implemented examination on their teaching methods, the responses showed that they were using the textbook in their classroom instruction in a manner that was different from what the MoE curriculum designers intended. The interviewees responded that they were not following the teachers' guidebook thoroughly because they believed that it was difficult to address all the content in the allotted teaching time and the responses in the questionnaire suggested that the teachers felt the teaching syllabus was not well prepared or structured to meet their students' needs. For instance, there was a lack of sufficient real-life tasks and authentic materials in the textbook and insufficient tasks reflecting students' language proficiency (see 5.2.5). The changes to teaching resulting from the reformed curriculum, including the textbook, teaching materials and testing specification, seemed to reflect positive washback. However, the changes to teaching practices resulting from the new examination appeared to be quite superficial. While the new examination may have changed the content of teaching

it did not have the same degree of impact on the way the teachers taught. This is why most of the teachers ranked the teaching activities geared towards the requirements of the new examination as the most common activities used in the classroom (see 5.9). These results agree with those of other empirical studies, which have found that public examinations have little to no effect on teachers' methodology (e.g. Alderson & Hamp-Lyons, 1996; Cheng, 2005; Qi, 2004; Shih, 2009; Wall & Alderson, 1993). Cheng (2005) confirmed that teachers' teaching methods in the classroom remained largely unchanged; for instance, reading aloud was changed to include roleplay and discussion, but these activities were still taught through drilling.

Teachers' perceptions regarding classroom assessment

This study found that the most important function of the mock examinations was to prepare students for the examination and enable them to practise exam-like questions. To enhance students' readiness for the examination, for example, one of the teachers interviewed stated that she discussed the requirements of each question and asked the students to practise the questions during lesson time. Some teachers also gave students extra examination practice for homework. The teachers gave their students information about examination questions and test-taking strategies and provided opportunities to practise under examination conditions. Three teachers interviewed reported that examination preparation occurred during regular lesson time, with one lesson typically devoted to each part of the examination. Some researchers have suggested that test preparation techniques are an effective way of boosting students' scores (Chamberlain et al., 2011; Mehrens, 1991), while others have argued that these practices may corrupt test scores and make them less dependable (Berry & Lewkowicz, 2000; Haladyna et al., 1991; Lumley & Stoneman, 2000).

According to Haladyna et al. (1991), ethical test preparation practices include training in test-wiseness skills and attempting to motivate students to perform well by discussing the questions and the importance of the examination, whereas over-emphasising the skills tested in the classroom and preparing teaching objectives which match the test are unethical test preparation practices (see Chapter 3). Moreover, other studies have found "overt backwash" (Prodromou, 1995), which tends to occur when students' over-reliance on test-related materials and test-focused notes influence their learning style and hinder their learning (e.g. Rahman et al., 2021; Shih, 2007; Zhan & Andrews, 2014). In this study, therefore, it was interesting to note that mock examinations were used to familiarise students with the format of the examination, but it was also important to investigate whether the learners had improved or learned more because they were better prepared for the examination.

When selecting techniques for assessing their students' learning, the teachers rated presentations as the method least utilised for evaluating students' oral abilities. This suggests that the introduction of the English elective examination caused teachers to focus less on evaluating students' speaking abilities since they were not assessed in the final examination. Indeed, the interviews revealed that the teachers emphasised reading, writing and listening more than speaking in their classroom assessments. This further emphasises the effect of the new examination on classroom assessment, particularly regarding the skills tested, assessment methods employed and selection of topics and tasks. These findings are in line with studies that have demonstrated the skills tested and the structure of the new test can have the unintended consequence of “narrowing the curriculum” (e.g. Airasian, 1988; Madaus, 1988; Popham, 1987), as well as confirming the findings of studies such as those conducted by Wall (2005), Shih (2007), Al Amin and Greenwood (2018), Hoque (2016), Sultana (2019), and Rahman et al. (2021).

Teachers' perceptions of the new examination

Reflecting on the results related to the three themes discussed above, this study confirms that aspects of the teachers' positive response to the new examination design and format did line up with the intentions of the MoE decision-makers. The teachers acknowledged that they had to alter their teaching to accommodate the new assessment goals. Nevertheless, the results indicated that the impact of the new examination on teachers' perceptions could have been somewhat superficial; that is, the examination may have impacted the material they taught but not the teaching methods they used or how they evaluated their students. The teachers may have held positive perceptions concerning the design and format of the new examination, but they still found it challenging to implement change in their classroom teaching. The data suggested that the intended change in teaching was altered by factors inherent in the English elective examination itself, as well as by the teaching context (see, e.g., Dawadi, 2021; Spratt, 2005; Wall, 2005).

An interesting set of findings in this regard can be found in Wall's (2005) study, which provided an overview of how washback works in the context of a new examination and how it is influenced by teachers' beliefs concerning the factors which may facilitate or impede the intended change in teaching that the test is promoting. Wall's (2005) analysis revealed that it is difficult for a new exam to exhibit much influence in classroom teaching as desired by the policymakers and that there are many factors that explain why a test may have a stronger influence on some aspects of teaching than on others. In this study, there were certain misconceptions among the teachers concerning the examination itself and the

materials on which it was based. While the teachers were familiar with the content and format of the examination, they did not seem to be aware of its underlying principles. Had the teachers known more about these principles, it is possible that they would have paid equal attention to all the skills introduced in the new curriculum and made full use of all the teaching strategies the examination change was designed to promote.

The confusion among the teachers and their lack of knowledge of the intended washback from the new exam makes it difficult to draw a definite conclusion about the nature and scope of the examination's influence. Based on the findings of this study and the proposed conceptual model, it appears that the individual teachers' beliefs and attitudes towards examinations, classroom teaching, assessment and contextual factors, such as the lack of teaching time, topics covered by the textbook, students' ability, (lack of) teacher training opportunities and the information available about the examination, greatly influenced the teaching environment and what that took place in their classes. Based on my own experience in this context, it could be argued that the lack of training and professional development were the most important factors in hindering take-up of the intended new teaching methodology, making it a real challenge for the teachers. This is because the new examination required a shift in teaching principles and objectives that had not been introduced previously in the lower grades (see 2.2.5), meaning that the teachers required specific training to develop their teaching practices and align them with the new curriculum objectives.

The importance of teachers' professional experience has also been emphasised by previous washback literature (e.g. Cheng, 2005; Shohamy et al., 1996; Wall, 2005; Watanabe, 2004) and has been deemed helpful for understanding why washback happens with some teachers but not others. Shohamy et al. (1996) found a noteworthy difference between experienced and novice teachers when examining the impact of testing; experienced teachers mainly relied on the test for teaching guidance and only used test-oriented materials, while novice teachers employed a variety of activities and materials in their teaching.

Overall, the findings indicate that the teachers generally regarded the new examination design favourably and felt that its features were appropriate. There is no sign, however, that the examination is affecting how teachers teach or that the teachers have understood or have yet been able to implement the proposed methodology introduced in the new textbook. This confirms Wall and Alderson's (1993) argument that washback is complex and elusive. There may not always be a direct or inevitable consequence of introducing a new examination in a particular context. What is clear is that the proposition that "good" tests are likely to create

positive washback with regard to teaching and that “poor” ones will probably generate negative washback is overly simplistic. The results of this study are consistent with previous research (Chen, 2006; Cheng, 2005; Imsa-ard, 2020; Rao & Haque, 2019; Tran, 2016; Yamashita, 2011), which found that while the introduction of a new test and textbooks was expected to have a beneficial effect on classroom teaching, teachers did not alter their teaching methods and only the classroom content was altered.

6.2.2 Intended washback from the new examination

RQ2 concerned the intended washback of the new English elective examination on teaching, as expressed directly by the MoE decision-makers. It has been suggested in the literature and the study framework that intended washback as perceived by the exam developers, or MoE decision-makers in this study, could aid understanding of how the new examination might influence teaching and it is necessary to establish the extent to which the intended washback has been achieved following the introduction of the new examination (see Alderson & Wall, 1993; Fan et al., 2020; Qi, 2005; Shohamy, 2021; Winke, 2011). These findings also provide the context for exploring teachers’ perceptions (RQ1) and their classroom practices (RQ3) in relation to the specific changes the examination was meant to bring about.

The interview findings showed that the MoE decision-makers shared a common view of the washback from the new examination, namely that they wanted it to: (i) enhance students’ language skills; (ii) foster continuous assessment for learning; (iii) improve student-centredness and communicative teaching practices. Each of these is addressed in turn.

Enhancing students’ language skills

The interviews revealed that the new examination was designed to provide a better means of assessing students’ language skills. This implies a need to focus on oral/aural skills rather than only on reading and writing. Three participants believed that the lack of aural assessment in the old examination hindered students’ communicative competence, as they would only learn these skills if they knew they would be tested. Additionally, the MoE reports showed that students were not prepared for IELTS, typically required for university admission and job opportunities. The new elective diploma was designed to fit with the IELTS question types and writing, reading and listening requirements; however, speaking was not tested (see 2.2.4).

In addition, there were several comments regarding the negative washback from the old examination design and content on teaching practices, particularly due to its inadequate coverage of reading strategies and its strong emphasis on writing skills over other core skills.

This relates further to how decisions about the old elective examination were made by the MoE decision-makers and others within the Omani educational system. Jin and Fan (2011), Sultana (2018) and Shohamy (2020) have all established that for assessment reform to be successful, policymakers must share pertinent information with other stakeholders, such as teachers. These scholars have proposed that understanding teachers' beliefs and perspectives is critical to accomplish and execute the desired objectives of a new curriculum.

This is also in line with the findings of Fan et al. (2020) in their washback study investigating teachers' perceptions regarding the recent reform of a high-stakes English proficiency test in China (TEM4). The study showed that the chances of TEM4 producing positive washback on teaching and learning were enhanced by the fact that the overall views of the teachers and the TEM experts converged with regard to the features included in the TEM4 and the purpose of adopting this examination reform. In Oman's centralised educational system, input from teachers is crucial as policies are typically created from the top down. Considering this issue in relation to the changes to the English elective examination, the implementation of the new design was supported by teachers as having a positive impact on teaching according to the responses of the MoE officials. Eliciting teachers' views will help MoE officials assess their perceptions of the revised version of the examination and evaluate the reform, thus clarifying whether the new examination is producing the intended washback on teaching.

The evidence also implies that the new examination was introduced to foster and refine the students' general and academic English language skills (reading, listening and writing), needed for higher education and the workplace. The new examination focuses on reading, writing and listening but not on speaking. As the interviewees suggested, this might render the new examination less effective in improving communication skills and preparing students for IELTS. Hence, the examination might not be an effective tool for achieving the desired pedagogical changes. The omission of any speaking assessment from the examination is likely to have had a considerable impact on the content of elective lessons and on how teachers design their classroom assessments, thus undermining the intended washback. As previous studies have suggested (e.g. Al Amin & Greenwood, 2018; Hoque, 2011; Sultana, 2019), the lack of testing of speaking and listening causes teachers to neglect these skills in the classroom, thus creating negative washback. Rahman et al. (2021) investigated the washback from the Secondary School Certificate (SSC) English examinations in Bangladesh. They found that some negative washback resulted from a misalignment between the design of the test and curriculum objectives in relation to speaking and listening skills. This was because teaching reading and writing was an effective way of

raising test scores; the teachers were anxious to cover the textbook content they felt was most closely related to the test and thus neglected skills not tested (speaking and listening). This confirms what Andrews (2004), Brown (2004), Johnson and Shaw (2018) and Au (2007) have identified, namely that the more test design is aligned with the target skills, the more positive the washback.

Providing continuous assessment for learning

The intended washback from introducing the new examination was to encourage continuous assessment strategies and procedures to help facilitate student learning. This approach was fostered through the different classroom assessment methods on offer to help teachers assess students' progress throughout the academic year (see 2.2.3). The MoE participants mentioned that continuous assessment applied only to the three main skills of reading, speaking and writing, whereas listening was only tested in the examination. This was attributed to the fact that the marks from continuous assessment aligned with the particular focus of the relevant textbook content. It could be argued, however, that even if the changes to the content of continuous assessment were geared towards enhancing students' learning, as mentioned by the participants, the divergence regarding the assessment of oral/aural skills was likely to affect teachers' ability to establish specific classroom teaching objectives and to conduct activities that would achieve these objectives (Muñoz & Álvarez, 2010). As suggested by Black and Wiliam (2003), for instance, teachers' formative work should "not be undermined by summative pressures" and "summative requirements might be better served by taking full advantage of improvements in teachers' assessment work" (pp. 623–624).

Another finding was that the preparation period for the mock examinations was unnecessary. The four participants noted that the mock examinations were mainly used to familiarise students with exam-style questions, including the question types and level of difficulty. Teachers did not have to write these examinations formally because they were deemed to be optional, depending on the students' needs. Instead, the teachers used past examination papers, textbook activities, published IELTS materials and other self-developed materials for examination preparation lessons. Although the mock examinations were not used as part of a formal assessment process, the teachers generally commented favourably on them and felt that they helped students prepare for the actual examination (see Table 5.10). This finding corresponds to those of other washback studies (e.g. Azadi & Gholami, 2013; Salehi et al., 2012; Waltman, 2008; Wisdom, 2018; Xie, 2015), which have confirmed the existence of a relationship between teachers' beliefs about examination preparation and their

classroom teaching. For example, Wisdom's (2018) study, investigating teachers' practices and attitudes involved in preparing students for high-stakes testing, found that they felt it was their responsibility to prepare their students for the tests and that they needed specific training to deal with examination preparation and to help students practise for the test.

Whether the recent changes to the examination content will achieve the intended positive washback on language teaching and assessment will depend mainly on how teachers respond to these changes. According to the findings (see 5.2.8), the opinions of the teachers and the MoE decision-makers needed to be in alignment concerning the content of continuous assessment. For example, the teachers said that they prioritised reading, writing and listening more than speaking in classroom assessments, even though listening was not covered in the continuous assessment materials. Moreover, the teachers argued that they relied on class tests to enable students to practise the examination requirements, rather than employing different assessment methods, as desired by the MoE decision-makers. This further suggests that the new examination had a strong impact on classroom assessment, specifically with regard to the skills tested and the methods used. For instance, the use of just one method of classroom assessment (class tests), rather than any of the other assessment strategies, is liable to have affected the reliability of the classroom assessment process, since the students were probably given fewer opportunities than intended to exhibit their competencies and knowledge (Black et al., 2003; Shohamy et al., 1996). This indicates that teachers' perceptions of classroom assessment may lead to undesirable implementation and consequences.

Improving student-centredness and teaching practices

Another significant reason for introducing the English elective examination was to promote students' independent learning. In this regard, the MoE decision-makers stated that the textbook had been supplemented with tasks and activities intended to allow students to take ownership of their learning and to enable teachers to emphasise student-centred instruction. In this regard, "student-centredness", as perceived by the MoE decision-makers, could stem from positive washback. For example, to address the students' speaking skills outlined in the curriculum, teachers were asked to adopt group discussions in which students had ample opportunities to express their views and discuss their feedback. In this way, the MoE decision-makers hoped to encourage teachers to develop teaching strategies that would promote independent learning among the students. In the context of a new test, however, the literature indicates that the interpretations of stakeholders such as teachers and students may differ from those of the test developers (see Fan et al., 2020; Malone, 2013; Qi, 2007; So, 2014). For example, Qi (2007) found that the teaching of writing was not practised as

intended by the policymakers because teachers were not familiar with the specific objectives or the role of the new examination reform. Similarly, the teachers in this study reported that there were many challenges that hindered them from adopting “student-centred” activities or making their lessons more interesting and shifting from a monotonous teaching style. In particular, the teachers tended to adopt traditional direct instruction because they regarded it as an effective approach to give their students practice in preparation for the examination.

Another central aspect mentioned in this regard was that the new English elective examination and the English Insights 3 textbook series were based on the CLT approach and designed to influence both what and how the teachers taught. Regarding the “what” element, the MoE representatives commented that they wanted to encourage teachers to focus on improving students’ learning rather than emphasising the examination content. Hence, the teaching content in the new textbook went beyond Omani students’ own familiar context, being designed to expand their understanding of the world and other aspects of the culture of the target language. In this regard, Alderson and Wall (1993) suggested there is a need to complement the data obtained from stakeholders with classroom observations to explore the complexity in classroom teaching practices arising from washback. The findings regarding the impact of the new examination on the content of teaching activities as intended by the MoE stakeholders are taken up later in the discussion of classroom observation.

Another issue discussed was that teachers were not meant to focus on the exam content in preference to the expected outcomes required at the end of grade 12. The MoE participants mentioned that there was “flexibility” in teachers’ selection of the teaching content. However, this flexibility meant that teachers might choose to focus on the skills included in the examination over and above those not tested. Thus, if a teacher chose not to address grammar or vocabulary because their students required further practice in the other main skills tested in the exam, would this mean that the materials were “flexible”, or that the teacher knew what the students needed to improve their attainment and increase their motivation? It is worth asking whether such changes were useful or whether they might have undermined the core goal of the intended washback.

Concerning the “how” element, the four MoE participants recounted that they had seen evidence that the new examination and textbook generally had a positive effect on teaching methods during their visits to schools. The use of new teaching techniques, teachers taking on the role of facilitators rather than dominators, the promotion of real-life and authentic communication tasks to encourage student-to-student interactions and the use of different

sources of materials were considered some of the positive impacts on teaching ushered in by the new examination. Although these features were all perceived as positive elements of teaching practice, the MoE decision-makers did not provide sufficient examples to support this view. It seems unclear how these practices were employed or manifested in the changes to the examination. The lack of shared policy according to the decision-makers' voices is indicated later in the discussion of classroom observation.

The intended washback of the new English elective examination

In view of the above findings, it seems that the intentions of the MoE decision-makers imposed constraints on achieving the intended washback on teaching and learning. East (2015) argued that introducing a new examination can be “a tricky business” (p. 101), because despite test designers having compelling reasons for reforming a test, they may not always succeed in achieving their objectives in terms of teaching language learners. Worse still, the revised examination may bring about unintended consequences or outcomes that the test designers fail to anticipate during the implementation stage (see also Bachman & Palmer, 1996). In this study, it was found that the intention to develop general and academic English language skills compelled teachers to focus on reading, writing and listening, because these were the elements the teachers believed were measured by the test. As the data suggest, much of the intended washback effect failed to occur.

It is also interesting to note the discrepancy between the MoE decision-makers' intentions and teachers' perceptions of their assessment practices in terms of the assessment of oral skills. The change in the content of the new examination here may have affected the teaching content in an undesirable way, allowing unintended washback to occur. The results of this study indicate that a test may narrow the content of a teaching syllabus, because teachers will teach to the test regardless of the textbook content. When a decision is taken to make teaching content “flexible” and teachers are given the freedom to choose of what to teach, there is the issue of who can afford not to teach or study for the test.

Several washback studies conducted in different contexts have shown that a lack of alignment between classroom teaching and high-stakes testing can result in negative washback (Cheng, 2005; Linn, 2000; Muñoz & Álvarez, 2010; Tan, 2008; Wall, 2005; Wall & Alderson, 1993). These researchers found that such a misalignment encouraged teachers to narrow the curriculum to cover only those skills that appeared in the examination. This meant that their students did not develop the full range of concepts and skills specified in the curriculum and thus showed how negative washback can significantly affect students'

academic performance (Popham, 1987). Decision-makers could potentially improve their language testing systems by ensuring alignment between the assessment format and the teaching-related objectives. In any case, the congruity between teaching and assessment tasks needs to be considered when designing a new test to promote and facilitate effective teaching practices, as suggested by Messick (1996), Morrow (1979), Qi (2005), Turner (2012), Muñoz and Álvarez (2010) and Rahman et al. (2021).

6.2.3 The perceived washback effect on teachers' classroom practices

This section discusses the impact of the introduction of the new English elective examination on classroom teaching in terms of teacher behaviours. According to Bailey (1996) and Taylor (2005), the impact of washback can be evaluated in light of the principles and practices of CLT. This impact is noticeable in how teachers adjust their teaching methods to align with the goals of the new curriculum. Several studies on washback have shown that although test creators aim to incorporate communicative practices in the classroom (thus resulting in positive washback), the new exam only causes a change in teaching content instead of a modification in teaching style (Andrews, 1994; Chen, 2006; Cheng, 2005). However, other washback studies have shown that tests can affect not only the content taught in lessons, but also how teachers teach (e.g. Alderson & Hamp-Lyons, 1996; Furaidah et al., 2015; Stecher et al., 2004). Furaidah et al. (2015), for instance, explored how teachers prepared students for a high-stakes test in Indonesia. The study revealed a positive washback effect due to the increased time spent preparing students for the test and honing their communicative skills, although the intensity of this effect varied between high- and low-achieving schools (see 3.7).

In this study, although English language teaching in Oman adopted the CLT approach about two decades ago (see 2.2.5), several studies have shown that teaching did not fulfil the goals of the old syllabus and curriculum (Al Balushi, 2001; Al Balushi & Griffiths, 2013; Al Shabibi & Silvennoinen, 2018). A series of classroom observations conducted by these researchers demonstrated a lack of use of student-centred approaches, with teachers showing a marked preference for a teacher-centred style. In similar terms, the interviews with the four MoE participants in this study revealed that the design of the old examination could have served to undermine the CLT approach that English language teaching in the Omani educational system was trying to promote (see 5.3).

As discussed in Chapter 4, 14 elective diploma lessons at three schools were examined in this study. It was decided to observe the lessons of several teachers to investigate whether

the influence of the new examination and curriculum varied between teachers and to gain a better understanding of the range of skills and methods employed in teaching. The results are divided into the following sections: participant organisation, lesson content, student modality and materials.

Participant organisation

This study sought to investigate teaching and communication between teachers and students by examining the organisation of participants in the classroom. In this study, most interactions were between teachers and students, with only a few opportunities for student–student interaction. This teacher-centred approach resulted from the teachers’ reliance on the textbook materials as the main source of instruction in the elective diploma classes, in large part because they were required to cover the textbook, which meant that they did not have time to employ alternative materials, as suggested in the questionnaire and interview responses (see 5.2.4 and 5.2.5). The adoption of these textbooks as the main source of learning would give students few opportunities to express their views, raise questions or even interact with either their teachers or other students in the class. Moreover, as noted in the interviews and open-ended questionnaire responses and identified in the classroom observations (see Chapter 5), the teachers’ main classroom activities focused on reading, writing and listening as the main skills tested in the new examination with less of a focus on conducting activities related to the communicative aspects of language learning as proposed by the MoE decision-makers (see student modality, 5.4.1).

Prior studies (e.g. Alderson & Hamp-Lyons, 1996; Barnes, 2016; Kılıçkaya, 2016) have also shown that teacher-centred teaching tends to have a detrimental washback effect, impeding students’ communicative ability, as in this study. The teacher initiation–learner response–teacher follow-up (IRF) pattern typical of teacher-centred teaching has been associated with negative consequences for language teaching and learning as it does not reflect genuine communication (Cheng, 2005; Saglam, 2018; Shohamy et al., 1996). Other research has demonstrated that pair/group work can enhance learning, suggesting that devoting more lesson time to student interactions and language use is advantageous in promoting CLT (Glover, 2006; Hayes, 2003; Taqizadeh & Birjandi, 2015; Wall & Horak, 2011). For instance, Hayes’s (2003) findings indicated that a teacher who devoted the most time to group work had a more student-oriented approach than others who employed a teacher-directed approach.

This study showed that Teacher B's and Teacher C's lessons focused primarily on teacher instruction and individual practice, whereas Teacher A allowed more opportunities for group work activities and student interactions, such as negotiating lesson content, discussing their work and using the target language in meaningful ways. Yet, although there were some opportunities for student–student interaction, few instances of student-led activities were observed in Teacher A's classes. This might suggest how individual teacher differences influence teaching and classroom activities. The teacher is considered another important variable in washback studies, shaping the style of classroom teaching (Cheng, 2005; Glover, 2006; Read & Hayes, 2004; Tsagari, 2011).

Lesson content

Through analysing classroom activities, it becomes apparent where the emphasis of teaching lies, whether it be on the meaning or structure of content, or a blend of the two. In 5.3, the findings indicated that the intention of introducing the new examination from a policy perspective was that teachers should focus more on a “meaning-oriented” form of language teaching rather than the traditional style of second language teaching that focuses primarily on “form”. The findings indicated an alignment between the MoE decision-makers' intentions and what teachers taught in the classroom. For example, teachers focused on communicating meaning rather than the functions of the language. Interestingly, other washback studies have mostly found that instruction focuses on form rather than language (Agrawal, 2004; Mohamad et al., 2018; Qi, 2005; Taqizadeh & Birjandi, 2015). Mohamad et al. (2018) found that form was the dominant focus in Syrian and Indian schools. The techniques employed included giving procedural instructions, translating sentences into the first language, constructing vocabulary lists and fixing pronunciation.

One possible explanation might be that the new topics in the teaching content primarily emphasised unknown vocabulary and did not refer to grammar or vocabulary-and-grammar elements. In addition, language components, such as grammar and vocabulary, were not assessed in the new examination, thus potentially leading teachers to focus on meaning in language use in their classroom activities. This supports the finding from the interviews with the teachers, who perceived the content of the grammar and vocabulary elements in the new *English Insights 3* syllabus to be challenging and reported that they would require more time to address them because many of the elements included were not covered in earlier grade levels. Why should teachers spend time on language elements not required in the examination, particularly when the teaching time they have is already limited (see 5.2.5)? This need to prioritise given limited class time implies that a test has the power to extend or

narrow the content of teaching, because teachers will teach to the test content regardless of the textbook in use. This practice undermines the MoE decision-makers' efforts to promote the intended washback.

Student modality

This category refers to the skills the students engaged in during lessons – listening, speaking, reading and writing – and other skills used in classroom activities. This aspect addressed the issue of whether the lessons observed covered a range of student modalities, as intended by the MoE decision-makers.

The findings showed that reading and listening were more intensively practised by the students, over and above the other skills. However, the time spent on these skills varied from teacher to teacher. This could be attributed to the different language abilities of the students and how each teacher delivered class activities. The three teachers who were observed conducted their classes in different ways according to their students' needs and abilities. For example, Teacher A lectured students on aspects of the language, explained how to carry out certain tasks and used various classroom assessment tasks to assess understanding, whereas Teacher B emphasised group discussions and presentations and provided individually differentiated materials in teaching classroom activities (see participant organisation, 5.4.1). Moreover, although speaking activities appeared in the lessons (18%) of all three teachers, speaking often involved a pre-lesson activity or was featured in combination with other skills (reading, listening, or writing) instead of receiving direct attention. This might be due to teachers' beliefs about the activities essential to meet the demands of the new examination format, which assessed both listening and reading skills (see 5.2.8) and supports Wall and Alderson's (1993) hypothesis that "A test will shape what teachers teach" (p. 120). It also confirms the findings from other research contexts (Abu-Alhija, 2007; Choi, 2008; Stecher, 2004), where it has been shown that teachers are liable to focus on the topics most likely to appear in an examination.

Another interesting finding was that less time was allocated to students' writing than to the other skills, as this represented only 8% of the total lesson time. However, if the examination were having an influence on the content of their teaching, as discussed above, the teachers could be expected to pay significant attention to writing skills. A possible explanation for this discrepancy that I posit here might be that the writing tasks were often given as homework rather than as practice in the classroom. From observing the writing lessons conducted by the three teachers, it was apparent that they discussed the style of the different

writing texts but did not allocate time for the students to write during the lesson. This, in turn, could explain why so much of the intended washback effect on the teaching of writing failed to occur, since English language teaching classrooms for second-language learners tend to include students who need more attention and practise to develop their writing skills (e.g. Azadi & Gholami, 2013; Estaji & Tajeddin, 2012; Qi, 2007; Yu, 2020). For example, the study conducted by Yu (2020) reported that the influence of writing tests on students' writing proficiency level was surprisingly limited. Hayes' (2003) study revealed that formal homework was used as a regular feature in teaching IELTS writing classes, because teachers were teaching the course from a textbook and IELTS preparation materials in a relatively short time, meaning that students had to cover a large number of topics on their own.

The data also suggest that the students in Teacher A's lessons employed the widest variety of skills, while the range of skills covered was more limited in Teacher B's and Teacher C's classes. This further indicates that tests can generate effects and change the approach to teaching to varying degrees, depending on the individual teacher. This finding also supports earlier research which has shown that washback can affect individual teachers' plans for teaching activities to different degrees, according to their different pedagogical approaches (e.g. Ferman, 2004; Lam, 1994; Read & Hayes, 2004; Shih, 2009; Wall & Alderson, 1993). However, while the student modality category provides useful information about the time devoted to particular skills and the combinations of skills in classroom activities, it does not directly address the degree to which the skills employed in the classroom teaching context have a communicative basis.

Materials

This section examines the essential characteristics of the classroom materials used during the lessons observed (see 5.4.1). The results revealed variations in the use of teaching materials among teachers. Teachers B and C relied primarily on the *English Insights 3* textbook and various extended written sources (standard materials for teaching English as a second language) in their lessons. This implies that the teachers' reliance on the existing textbook, composed of exam-style tasks, could have restricted their teaching approach. This is in line with previous studies that have shown that negative washback can appear when teachers use textbooks that are purposefully published nationwide to fit with the content of a newly reformed examination (Cheng, 2005; Hamp-Lyons, 1998; Kılıçkaya, 2016). It also echoes the findings of Andrews' (1995) study, which described the approach of teachers who spent a lot of lesson time on exam-related materials as "a limiting of focus for teachers and students rather than a broadening of horizons" (p. 80). One explanation for the use of the

textbook as the main teaching source in this context is that teachers are actually required to cover the textbook, which was published and prescribed for all schools by the MoE and this could mean that they did not have available teaching time to employ alternative materials, as was suggested in the questionnaire and interview responses (see 5.2.4 and 5.2.5). Furthermore, it might be expected that teachers' awareness of the requirements of the examination would mean they avoided alternative teaching resources, since what was tested in the examination was already available in the existing textbook.

In the case of Teacher A, however, the lesson used a wider variety of materials to integrate and practise different language skills and incorporated authentic materials (audio and video) from different sources. This further confirms that individual teachers contribute substantially to mediating positive washback in teaching practices.

Examination information and strategies in classroom teaching

Additional analysis from this study revealed that the three teachers invested most time in providing information about the examination and test-taking strategies (see 5.4.2). For example, Teacher A spoke about the writing text types in the examination, while some of the students took notes and others asked specific questions about the writing tasks. Other instances of this approach included giving tasks under examination conditions, linking the examination with the activities in the textbook, identifying examination skills and strategies and evaluating student performance in examination-like tasks. In some cases, the teachers used their first language to provide students with tips for the listening and writing questions in the examination, demonstrating an extra emphasis on providing students with examination-related information, rather than simply practising English.

Teacher C dedicated more time to discussing examination information than the other teachers. The increased emphasis in Teacher C's classes was to be expected, given that the classroom visits in this case took place closer to the time of the final examination. This could signify the seasonal nature of washback; teachers tend to concentrate mainly on examination components when the examination is close at hand (see Shohamy, 1996; Wall & Alderson, 1993). This finding corroborates Bailey's (1999) assertion that seasonality could be a "relevant concept in washback studies" (p. 40). Nevertheless, this discovery necessitates a more thorough examination to investigate, for instance, the impact of time on the implementation and utilisation of a test over a particular period.

The findings also revealed that the three teachers invested more time in teaching exam-taking strategies than providing information about the examination. This study identified 15 exam-

taking strategies used during the lessons that were observed, illustrated with examples (see 5.4.2: Tables 5.21 and 5.23). While some strategies were common to the lessons of all three teachers, the emphasis on teaching these strategies varied from one teacher to another. It is clear, however, that there was an “overt” washback effect from the examination on the classes of the three teachers. The examination may have exerted different effects at different levels, but the extent of its influence probably increased during the examination period. Of course, there were some classes in which few instances of the examination effect were observed, but the elements practised in the classrooms were limited to reading, writing and listening, these being the only elements tested.

Several washback studies have explored common elements of teacher–student interaction, including teacher discourse (Alderson & Hamp-Lyons, 1996; Burrows, 2004; Glover, 2014; Turner, 2001), student participation and teacher feedback on student performances (Choi, 2015), and teachers’ explanations of concepts to their students (Burrows, 2004; Glover, 2014; Stecher et al., 2004; Watanabe, 1996). However, guiding individual students in classroom activities is considered in this study to be a secondary form of interaction between teachers and students. Such secondary interaction, which includes monitoring students’ performance or responding to questions, has also been considered by previous washback studies examining differences in teaching methodologies and approaches (Burrows, 2004; Hayes & Read, 2004; Wesdorp, 1982). In this regard, Teacher A and Teacher B helped students during classroom activities in various ways, such as assisting in writing tasks, clarifying reading tasks and explaining language points. However, when Teacher C’s students completed a task, they mainly worked in groups and did not ask their teacher for explanations or assistance. This aspect of the classroom environment may be attributed to the teacher’s instructional style or other factors, including the lesson topic and plan, the class seating arrangement, the length of the lesson and the number of students in the class.

Washback on classroom teaching

This research and prior washback studies have demonstrated that implementing a new test can affect the material taught while having little or no effect on the teaching methods employed (see Chen, 2006; Cheng, 2005; Wall & Alderson, 1993). Although the new English elective examination significantly affected teaching content, there was little proof that it significantly influenced teaching practice. The teachers’ teaching methods were influenced by the examination and reflected in their implementation using the textbook (English Insights 3), which was designed to meet the objectives of the revised English elective curriculum. Hence, regardless of the broader content and objectives of the textbook, teachers tended to

put their beliefs about good teaching practices to one side and simply teach to the test. As a result, the teachers tended to adhere to a teacher-centred approach, the focus of which was on the content and skills included in the test, and thus they did not teach in a way that would promote learning and improve students' language abilities or best practice in communicative teaching.

Some classes, particularly those taught by Teachers B and C, experienced a negative washback effect. Nevertheless, determining whether an examination has a positive or negative washback effect is complicated due to the several variables that must be considered. Even if an examination has been created to foster CLT in the classroom, this does not guarantee that it will be implemented. As demonstrated above, the materials and sources used in the classroom significantly impact how teachers teach (Barnes, 2017; Cheng, 2005; Hamp-Lyons, 1998; Kılıçkaya, 2016; Read & Hayes, 2004; Tsagari, 2011). Given that the material in the English elective textbook attempted to reflect the format and content of the examination, the teachers spent much of their lesson time speaking to the whole class, teaching receptive skills rather than productive skills. Moreover, the teachers themselves spent more time in their lessons discussing exam-taking strategies rather than giving information about the examination. This in turn shows that the examination had the power to determine what skills would be taught and deemed important and conversely what skills would be treated as less important. In addition, this study suggested that teachers' characteristics were likely another important variable in washback as the teaching style in the classroom is shaped by those who teach in it. In this study, there were discernible differences between Teachers B and C and Teacher A in terms of their teaching experience, although they shared similar characteristics in terms of their academic qualifications and the time spent on teaching the English elective subject (section 5.4.1 highlights this particular issue).

Moreover, teachers' beliefs about the purpose and use of the elective diploma curriculum and what constitutes best practice in language teaching are likely to differ (Watanabe, 2004). Previous washback studies have found that intended washback is more likely to be achieved if a new test is compatible or consistent with what teachers believe to be good practice (see "Washback on teachers' perceptions", above). The closer the intentions underpinning the changes to the examination are to teachers' beliefs, the more receptive they will be. Conversely, the more the changes differ from teachers' beliefs, the more resistant they will be. Looking back to the findings pertaining to teachers' beliefs, it is clear that they had mixed views about the intended washback from the new examination. They were aware of the

intended goals and wanted to make positive changes in their teaching to adapt to the new examination, but there were other external factors that impeded their practice and gave rise to a more negative response. The teaching behaviours observed in this investigation showed that not all teachers reacted the same way. This again reflects the hypotheses of Wall and Alderson (1993), who stated that “tests will have washback effects for some learners and some teachers, but not for others” (p. 121). The findings here also seem to align with those of other studies (Cheng, 2005; Tsagari, 2009; Wall, 2005; Wall & Alderson, 1993; Watanabe, 2004), showing why the intended changes (washback) may vary in degree. In line with the theoretical framework of this study (i.e. Bailey, 1996; Hughes, 1994; Wall, 2005; Wall & Alderson, 1993), data from different sources – questionnaires, interviews and classroom observations – helped provide a complete, yet complex, picture of the washback effect on classroom teaching.

6.2.4 Extent to which teachers’ characteristics influence the intensity of perceived washback effects

As mentioned in the findings related to the classroom observations, Teacher A demonstrated a different overall approach from Teachers B and C. The latter employed a teacher-centred approach much more than Teacher A, which is something the MoE decision-makers had originally intended to discourage. When seeking to understand such issues, researchers have realised that other factors underlying the multifaceted and complex process of washback should not be overlooked (Green, 2007; Shohamy, 1993; Wall & Alderson, 1993; Watanabe, 1996). In terms of the occurrence, intensity and nature of the washback effect, Andrews et al. (2002) in particular, provide a reminder to consider individual differences between both teachers and students, emphasising that these can be responsible for the unpredictability of the washback phenomenon.

To better identify the washback of the examination on teaching, teachers’ perceptions of the new English elective curriculum were supplemented by an examination of a number of potentially influential factors relating to the teachers’ themselves and their personal characteristics and experience. In this study, the teacher factors examined comprised teaching experience, gender and age (Watanabe, 2000), training in teaching related to a specific examination (Andrews, 2001; Green, 2007; Wall & Alderson, 1993; Watanabe, 2000), the teachers’ academic qualifications (Watanabe, 2000), their current teaching situation (such as the grades currently taught), and the number of hours per week spent teaching the English elective (Watanabe, 2000).

The inferential statistics reported and discussed in Chapter 5 (see 5.5) were used to assess, in particular, the impact of various teacher-related factors on aspects of washback in the classroom. Specifically, the analysis revealed the effect of the independent variables on two dependent variables – teachers’ teaching practices (TQ3.3, TQ2.D and TQ4.3) and teachers’ assessment practices (TQ2.F and TQ3.2). Thus, if the exam produced a strong or weak effect only related to certain teachers, the effect would likely have been mediated by their specific characteristics.

Perceived washback on teachers’ teaching practices

There was only a strong, statistically significant relationship with gender, particularly with females who answered “No” with regard to skipping over certain activities in the textbook because they were less likely to be tested in the examination. This finding was expected, given that several research studies in Oman have revealed that gender seems to be an influential variable affecting the type of instruction teachers implement (Al Bulushi et al., 2018; Alkharusi, 2011; Rassekh, 2004). Although the influence of gender has been a focus of research in language education, particularly concerning effective teaching practices resulting from curriculum change (e.g. Atta et al., 2012; Hopf & Hatzichristou, 1996; Shazadi et al., 2011), it has rarely been a focus of washback research in relation to classroom instructional change. There is, therefore, a need to examine teachers’ gender as a relevant variable in future research.

However, while the findings concerning classroom teaching practices showed that there were some statistical relationships between certain practices and particular teacher characteristics, other relationships were statistically insignificant or there was no association between the categories of the independent variables and the dependent variable (see 5.5.1). As one example, there was only a small statistically significant relationship between female teachers aged 20–30 years and their teaching practice in terms of emphasising the skills more likely to be tested in the examination. Overall, this means that the washback from the new examination on teaching practices in this study was not particularly attributable to teachers’ characteristics. However, several washback studies investigating the relationships between high-stakes tests and second language teaching contexts (e.g. Shohamy et al., 1996; Spratt, 2005; Tavares & Hamp-Lyons, 2008; Turner, 2009; Urmston & Fang, 2008) have provided evidence that teacher-related factors, such as educational level and experience, play a significant role in whether and how washback occurs, and to what degree. In Chen's (2002) study, the teachers knew the objectives of the curriculum and test but lacked the knowledge

to adapt their teaching methods. This was likely caused by a lack of in-service teacher training.

One possible reason for the contradiction with the literature is that other variables need to be investigated to shed further light on the teaching context in this study, and above all to indicate whether the findings from Oman, which employs a centralised educational system, can be applied to other ESL contexts. As emphasised by Wall (2000), different issues should be investigated at the individual, school and societal levels. For example, the findings in this study showed that teacher-related factors did not account for the learning activities teachers deemed important for teaching the English elective classes. This could be due to the fact that teachers in Oman use the textbook, *English Insights 3*, which follows the same format as the new examination and is regarded as the main source for teaching the new syllabus. The activities that appear in the textbook are delivered according to the teaching scheme, planned by the MoE decision-makers, so there is generally no difference in the selection of which activity is deemed important for teaching.

Moreover, the finding that there were no significant associations between teachers' characteristics and their practices in improving student performance in the new examination could be due to the ability of the students who selected English as an optional subject. The students who chose this subject tended to exhibit the motivation and interest required to improve their language skills. The teachers may thus have believed that they did not need to give their students extra practice for the examination since they already had a high level of English language proficiency. Given the complexities underlying the washback phenomenon, the changes to the examination and teaching syllabus were not sufficient on their own to require the teachers to change their teaching practices. It appears that fundamental changes in teachers' practices need to be linked to the contexts within which a test is administered and take into account multiple aspects, including, for example, the ability of the students, as suggested by McNamara (2000) and Wall (2005).

Perceived washback on teachers' classroom assessment practices

Regarding teacher factors and the relation to classroom assessment practices reported in 5.5.2, the results indicated that the teachers' age and gender affected their use of class tests and written work as main tools in classroom teaching, whereas teacher-related factors did not affect the use of classroom observations, presentations, or quizzes during lessons. The most interesting finding was that the statistically significant differences in agreement levels were found in relation to the status of the teachers who were teaching the new curriculum

for the first time during the year of data collection (2020/2021) and their use of mock examinations in the classroom. This finding was expected and suggests that teachers who were teaching the new curriculum for the first time tended to pay more attention to mock examinations in their classroom teaching than those who had already taught the curriculum at an earlier stage. This could be something to do with the way the mock examinations had been designed by the MoE decision-makers when the examination was first implemented, since the design of the mock examinations became the responsibility of the teachers after the first year of implementation (see 5.3).

The perceived washback effect vis-à-vis teacher factors

From the foregoing results, it can be deduced that teacher characteristics cannot entirely be considered intervening variables in mediating washback and might not significantly influence the direction or intensity of the impact of the new examination. This is because there does not seem to be a strong linear relationship between teacher-related factors and their practices in the classroom vis-à-vis the washback effect of the new examination. Interestingly, however, the inferential statistics recorded differences in the impact levels of teacher-related factors with regard to their classroom practices.

The findings of the study are again consistent with those of Fan et al. (2020), who found that teachers' background, such as their experience and affiliation types (foreign language universities vs non-foreign language universities) did not have much influence on their perceptions towards the new English proficiency test (TEM4). Researchers have cautioned that studying teacher-related factors without considering the intricate and multifaceted nature of the washback phenomenon in teaching can lead to simplistic and inaccurate conclusions. McNamara and Roever (2006) contended that each test should be studied within its social and cultural context. This study only included the teacher factors deemed most pertinent to the inquiry, but other contextual factors at the micro and macro levels could have been included in the analysis. This could provide additional insights into the educational context of this study, as the questionnaire and interview results particularly highlighted the issues and difficulties associated with the learning environment, which should be examined in greater depth (see 5.2.4, 5.2.5 and 5.2.8). The more factors present when a new high-stakes test is implemented in an educational system, the more evident its effect on teaching and learning will likely be.

Chapter 7. Conclusion

7.1 Introduction

This chapter provides an overview of the research and a synthesis of all the major findings. In particular, it highlights the theoretical and methodological contributions. It also reviews the key findings related to the research questions in the context of the washback literature and existing knowledge within the context of the stud. It then discusses the limitations of the study. The chapter concludes with the potential implications for the field in relation to policy and practice and proposes recommendations for further research.

7.2 Summary of the Research

7.2.1 *Research problem*

It was not unexpected that the new English elective examination designed as part of the revised diploma (elective and core English), used as a gatekeeping mechanism by Omani universities and colleges to assess grade 12 graduates' language proficiency, would have a strong effect on the perceptions and behaviours of teachers and students, particularly given the English language requirements imposed by higher education institutions and promoted by the MoE. Despite the claims put of the MoE decision-makers that testing is used as a key component in educational plans to improve teaching and learning, evaluation studies implemented in this context have revealed concerns regarding the quality of teaching and students' proficiency in English (see Chapter 2). In an effort to address such concerns, the new English elective examination was introduced in the 2018–2019 school year, based on a newly developed curriculum. This research has explored whether or not the intended washback from the new English elective examination appeared to have been achieved. It has further investigated the complex relationship between the influence of the new examination, teachers' beliefs and their classroom teaching. The study aims to gain a better understanding of the washback mechanism, the extent to which the new examination has promoted change and how such an examination can lead to changes in teaching. For this purpose, four research questions were formulated, as follows:

- RQ1. What are teachers' perceptions of the new English elective diploma examination?
- RQ2. What is the intended washback from the new English elective diploma examination according to MoE decision-makers?
- RQ3. What are the nature and scope of the apparent washback effects resulting from the new English elective diploma examination on teachers' classroom practices?
- RQ4. How does the intensity of the apparent washback effects from the new English elective diploma examination differ according to teachers' personal characteristics?

Given that English language teachers must equip students with the skills that the new design of the examination seeks to promote, it would seem plausible that the more closely aligned the test is with the teaching that precedes it, the more likely it is the test will achieve the intended washback for both teachers and students, as well as other exam stakeholders. However, based on the review of the literature examining models of washback (see Chapter 3) and the findings of other empirical washback studies, this thesis suggests that the mechanisms of the washback process need to be clarified. Using Hughes' (1993) trichotomy – participants, processes and products – as the main framework to guide this study (see 3.10.2), it was posited that these three categories are interrelated and that they could successfully be employed together to investigate the mechanisms of washback. In Hughes' (1993) framework, the “processes” category refers to any actions taken by the participants that may contribute to learning, such as changes in teaching methods, syllabus design or teaching materials. However, because the process of washback here refers only to those elements that have changed, Hughes' (1993) framework does not address how and why these behaviours have occurred (or not occurred). The literature reviewed in this thesis suggests that the washback is a complex area of study; previous investigations have revealed that it is not easy to predict the washback effect of tests, or how the impacts of tests will manifest themselves outside the research boundary, especially when intervening variables (factors beyond the exam itself) in any educational setting may determine (or preclude) the nature and intensity of any washback impact (Shohamy, 2001).

Because of these complexities, as the literature has suggested, it is important to understand the factors that are likely to influence perceived washback effects on teaching practices and how these factors work. One of the main factors noted in this study was the effect of the attitudes and beliefs of the teachers in relation to the introduction of the new English elective examination in the Omani context. Previous washback studies have indicated that intended washback is more likely to be achieved if the intentions are compatible with what the teachers believe to be good teaching. The closer the purposes and objectives of the examination to the teachers' beliefs, the more receptive they will be to it. Similarly, the literature has shown that the specific expectations of a new examination and the changes it is intended to promote in the classroom are important in understanding the processes of washback. For this study, it was important to establish what aspects of classroom teaching would be facilitated by the new examination and why certain changes in classroom practice were expected to take place. Thus, it was necessary to examine the intentions underpinning the new examination according to MoE decision-makers and the extent to which these

intentions were achieved in the classroom. Teacher-related factors that might facilitate or impede the perceived impact on classroom teaching were also highlighted in the reviewed literature. These factors could affect teachers' perceptions of their capacity to implement change and could relate to the teachers themselves, for example age, gender, teaching experience, and so on, or to the wider teaching context, such as teacher training, workload, or the nature of the current teaching situation.

7.2.2 Research procedures

The philosophical assumptions and arguments underpinning this thesis required a certain procedural design which consisted of three main phases. The opening pilot phase aimed to gain a general overview of the phenomenon under study and to identify the issues that might arise in the main study with regard to the research methods. Phase two was particularly concerned with teachers' perceptions and attitudes towards the new examination in relation to their classroom teaching practices. The overall aim of this phase was to report how teachers found themselves to be affected by the introduction of the new examination and why this would have the washback that it did. Hence, the findings of this phase addressed the research question "What are teachers' perceptions of the new English elective diploma examination?" This phase was also concerned with the extent to which teacher factors might influence the degree of the washback intended from the exam.

Moreover, it addressed the development of the validated research instruments used in phase three (teacher interviews and classroom observations). Phase three aimed to build upon and expand on the results of the second phase and thus enhance understanding of the washback from the new examination. This phase explored the perceived washback on what teachers did in the classroom and whether or not the new examination affected teaching in the ways intended by the MoE decision-makers. Hence, this third phase of the research was concerned with answering the questions "What is the intended washback of the new English elective diploma examination according to MoE decision-makers?" and "What is the nature and scope of the apparent washback effects of the new English elective diploma examination on teachers' classroom practices?" In this phase, teachers' perceptions of their teaching practices were also scrutinised to examine the relationships (or lack thereof) between what the teachers believed, what they did inside the classroom and the reasons underpinning their practice.

7.2.3 Design and methodology

The study was conducted using a mixed-methods design involving both quantitative and qualitative approaches. The samples used in phases one, two and three comprised 215 teachers, selected on a nationwide basis (including 12 who participated in more than one of the research instruments in phase three), and four MoE decision-makers (representing the Assessment Measurement and Evaluation Centre, the Curriculum Development Directorate and the Supervision Directorate). The research tools developed for the study comprised a questionnaire, interview protocols and a classroom observation scheme. The study yielded 193 completed questionnaires and audio-recorded data from three sets of classroom observations and eight interviews (five with teachers and three with MoE decision-makers).

7.3 Key Findings

This thesis has shown that washback is neither simple nor direct, but a complicated and inconsistent process. This complexity is evident from the investigation of how all the different factors fit into the washback process. The evidence indicates that the effects of the examination were to some degree negative, in that the teachers narrowly focused their teaching on practising exam-related skills and tasks, rather than on developing English language skills in a broader sense. However, the results also reveal that while the new English elective examination had a superficial washback effect in terms of changing teachers' behaviour, it should not bear full responsibility for all the unintended effects. Indeed, the data show that the examination functioned neither as a facilitator nor as a constraint on the implementation of the changes in instructional behaviour.

The findings add further evidence to the existing literature showing that other factors related to the teaching context (such as teacher training, teacher involvement and teaching time) are involved in the washback phenomenon and they appear to be more influential than the teachers' own characteristics. The data obtained show further evidence of the factors that hinder the desired change in teaching processes. When a gap exists between the new examination design and format on the one hand and the implementation of the new curriculum on the other, a mismatch arises between the intended aims of the prescribed curriculum and how it is operationalised in the classroom.

Moreover, the study provides empirical evidence that teachers' beliefs and attitudes concerning the introduction of the new examination are key to how they make sense of behavioural changes in terms of instruction. The teachers' beliefs concerning the intended washback on their teaching practices made a difference to how washback on teaching

practices took shape. The study revealed not only how teachers interpreted the washback from the new examination but also how they changed their teaching as a result of the reform. These findings reinforce those seen in Andrews (2003) and Turner (2009), whose studies showed factors other than the examination to be an essential component driving teaching practices. It could be argued, therefore, that examining the washback phenomenon with little consideration of other factors, such as teachers' beliefs, issues relating to the teaching context and certain personal characteristics related to the teachers themselves, cannot fully explain critical washback issues.

Examining the results of this study in relation to other research in the EFL environment, two main recurring factors were found relating to testing reform and the educational system. The first of these had to do with the way the Omani EFL teachers perceived and reacted to the reform of assessment and ELT and can be seen as similar to findings in other EFL contexts. While the introduction of a test in such a context can take many different forms, it seems to make little difference to the way in which washback operates and the complexities involved in the process. The second factor is that the process of introducing and implementing the new examination in this case shared many common features with other sorts of educational systems, specifically those employing a centralised model. These features include: (i) the complexity involved in the changes exhibited in teachers' attitudes and instructional practice; (ii) the anxiety exhibited in teachers' behaviour; (iii) the differences seen in each individual teachers' behaviours; (iv) the discrepancies evident between what the teachers believed about what constitutes good teaching and their actual practice. A detailed discussion of these factors can be found in Chapter 5.

It is clear from the above that further research needs to be conducted to examine the nature of this educational phenomenon in relation to other factors and across different educational contexts. To develop a deeper understanding of the underlying levels of this phenomenon, research on washback must go beyond the discussion of test impact and address all the different issues relevant to the educational system. As explained above, it is hoped that the factors implicit in the washback processes concerning the new English elective examination will provide the basis for improvement in any further process of testing reform. It is also hoped that the study findings and insights highlighted in this thesis will help inform the design of reforms to consider washback on teaching practices in the future.

7.4 Study Limitations

Although the findings offer insights that are potentially instructive in evaluating the reform process, the scope of this thesis has imposed certain limitations influencing the design and conduct of the research. This thesis examines how teachers perceived the intended washback, as articulated by the MoE decision-makers, and their behaviour in relation to the new English elective diploma examination. The result of the interaction of teacher factors in the washback process is that the product is likely to be positive in relation to some aspects of teaching learning and negative for others. Thus, while the conceptual framework in this thesis addressed issues regarding the mechanism of washback within the Omani educational system (which could include washback in relation to learners and learning), the study focused principally on the washback to teachers and teaching. Although it was important to talk to the teachers, as they are responsible for implementing any new curriculum and their opinions must therefore be taken into account in the reform process, it is equally important to investigate washback to the product by exploring the views of other stakeholders in the educational system, such as learners, school principals and parents. However, this was beyond the scope of the study.

Another potential limitation is that this study sought to investigate the washback from the new English elective examination on grade 12 teachers' classroom teaching shortly after this examination was first introduced. One particular drawback was that there were no baseline data related to the previous examination. According to Wall (2005), such data provide a good understanding of the situation prior to the introduction of a new educational change and allow researchers to make comparisons and links with the data collected later. In this study, there was no statement outlining the previous examination before the introduction of the new examination in any official documentation from the MoE decision-makers. Thus, the only data available as a way of mitigating this weakness comprised the teachers' anecdotes and the statements from the MoE decision-makers about teachers' classroom practices before the introduction of the new English elective, along with the previous local studies reviewed in Chapter 3.

It should also be pointed out that classroom observation was used in this study as a way of triangulating the self-reported information with actual behaviours. Due to limitations related to time and accessibility resulting from the COVID-19 restrictions in schools during the data collection phase, it was only possible to conduct observations in 14 classes in 3 different schools. This was because schools in Oman at the time were using a blended learning approach, in which teachers' timetables allowed for only two classroom-based lessons per

week and it was not possible to arrange further lessons with the selected teachers as the end of the semester was approaching. As the sample of the classroom observations in this study included three teachers, caution is needed in attempting to generalise the findings to the wider population of teachers of the grade 12 English elective curriculum in Oman. However, as with many other qualitative studies, while not generalisable, the findings can still inform other research in meaningful ways and may be useful in other contexts with similar characteristics. Additionally, while the purpose of employing classroom observations in this study was to investigate teaching practices, a combination of tools would probably be required to provide sufficient evidence of a specific washback effect. In-depth follow-up interviews and a review of classroom materials might provide the evidence to support a more detailed account of classroom teaching.

This study is one of the few washback studies to have adopted a mixed-methods research design. This meant there were few tools available in the existing washback literature that could be drawn upon, especially in terms of classroom observation schemes for providing detailed descriptions of classroom interaction. Although the COLT scheme has been widely used in various washback studies (see 4.7.2), I feel that it could be refined further to provide more information on classroom teaching process. COLT only showed the types of classroom interaction; it did not provide any detailed information on lesson activities or student interactions. It would have been beneficial, in this study, if interviews had been conducted after the classroom observations, to explore how the teachers' understanding of the examination affected the way they taught and why the new test might not have resulted in positive washback effects in some situations and for certain teachers. For the particular focus of this study, it was necessary to capture and record certain aspects of the detailed process of classroom activities. Therefore, COLT had a specific weakness when used on its own as a research tool to search for evidence of washback.

7.5 Implications of the Study

Despite the above stated limitations, the findings of this study have important implications for future research and practice. The study provides several insights – theoretical, methodological and practical – into the washback effect, drawing on the investigation of the influence of the English elective diploma examination on teaching practices. From the theoretical perspective, this study adds to the existing literature on washback on the teaching process specifically in relation to teachers' perceptions of the test. The literature suggests that washback can influence teaching positively or negatively and research in this field is primarily concerned with identifying the process through which an examination can enhance

teaching practices, i.e. drive positive washback (see Bailey, 1996; Hughes, 1993). The framework for this thesis (see 3.10.2, Figure 3.5) proposed that the complexity of washback lies in the interactive and multi-directional nature of the process, which involves a complex interplay of factors, with different degrees of intensity, relating to the examination itself and to the educational setting (among other components). In this research, the English elective examination did not directly influence teachers' behaviour in the classroom in the way expected by the MoE decision-makers. The teachers' different levels of awareness of the intended washback from the new examination played a crucial role in influencing their teaching behaviour. The divergent attitudes and behaviours of the teachers in this study further support the idea that washback is context-dependent and complicated. Examining just one factor, without analysing other contextual factors, meant it was not possible to explain certain complex washback issues, such as the relationship between how the teachers would like to teach and how they considered the new examination required them to teach, or why and how the washback phenomenon influenced some of the teachers but not others. Indeed, as represented in the framework, this study supports Wall and Alderson's (1993) proposition that washback is neither straightforward nor avoidable but is rather a complicated phenomenon in terms of its manifestation and how various factors influence it. Further empirical research is needed to address factors other than the test itself to provide further insights into the role played by the nature of the teaching context.

Furthermore, the operationalisation of the conceptual framework adopted in this study shows the value of comparing teachers' views with those of the MoE decision-makers. This needs further exploration in future washback studies. The results of this study demonstrate an apparent discrepancy between the perspectives of these two groups of test stakeholders: the teachers seemed to focus on the potential effect of the new examination design on the students' scores rather than seeing it as an opportunity to enhance their teaching practice; in contrast, the MoE decision-makers prioritised effective assessment procedures and a focus on broader learning outcomes (aural and oral) as the desired impacts of this reform. An example of this relates to the assessment of oral skills. The teachers felt that the inclusion of speaking (rather than listening) in continuous assessment did not mirror students' needs for the examination (which assessed listening), and that its removal would allow them to provide students with further listening practice for the examination. The MoE decision-makers, in contrast, were more focused on promoting students' communicative ability through a combination of testing and classroom assessment. They contended that the motivation for this move was to address issues of practicality and reliability. When such a gap exists

between the objectives and purposes underpinning a new examination on the one hand and its application on the other, there will be a mismatch in the implementation of the reform.

From the practical perspective, Andrews (1994) suggested that washback should be viewed as a tool to be used to facilitate curriculum innovation and ensure that textbook writing, teacher training and assessment strategies are all working in harmony. Where these three components of curriculum innovation are not in harmony, however, the successful adoption and implementation of the proposed innovation is bound to be compromised in some way. To analyse this properly, detailed information is needed about the specific features of the context and any factors likely to promote or hinder the implementation of the curriculum change.

In this study, the washback from the new English elective examination was addressed in terms of the MoE decision-makers' intentions to bring about a positive change in teaching and learning. In this context, the washback from the new examination on teachers concerned the means of change, how the teachers felt about it and how they dealt with it. Some teachers felt unhappy that they were obliged to make changes to their teaching methods, whereas others treated this change as something new to learn about. The study has discussed how the teachers coped with the introduction of the new examination. It has not only provided evidence of how and why the teachers changed – or did not change – their teaching practices but has also made some suggestions concerning how the examination reform process might be improved. For instance, the teachers' perceptions highlighted that it is important for the right conditions to be in place in classrooms and schools so that teachers responding positively to change will be supported and those who are resistant are brought on board. It is important, then, that the team responsible for examination reform in the MoE includes representatives from all the key stakeholders to facilitate better communication and cooperation between them and ensure that such reforms and their implementation not only respect exam-writing criteria (such as reliability, practicality and validity), but are understood by teachers and are acceptable to other stakeholders, for example in higher education institutions. While the "cascading" approach in place for teacher training is good, it needs to be enhanced and developed. For instance, there should be a follow-up programme to evaluate the impact of the training on teachers' attitudes and their classroom practice and to give them the opportunity to comment on and make suggestions regarding the training.

As mentioned previously, this study is the first to investigate the influence of the new elective diploma examination on teaching practices in Oman. This means that the findings related to

the pedagogical practices of teachers may have important implications for the teaching context and examination reform process. Notably, the thesis has highlighted gaps between the intentions underpinning the new English elective examination as proposed by the MoE decision-makers and the teachers' classroom practice. The examination had only a superficial effect on how the teachers taught. The classroom observations showed that the teachers still spent a large amount of lesson time lecturing or simply leaving the students to work individually and thus their pedagogical approach did not reflect the new teaching strategies introduced in the textbook. Moreover, the teachers' responses in the questionnaire and interviews showed that their focus in their classroom activities was on aspects reflecting the requirements of the examination.

The teaching and learning context in this study, as in others, was complex because there was so much going on in the classroom. The washback from the new examination may have been influenced by mediating variables, such as those relating to the teacher and the context. As these factors have not been given much attention in the literature, this study focused on the impact of such factors in determining washback. However, there is no way of isolating all the factors that play a key role in the washback process. The most that can be suggested is that the more these variables are considered and aligned with the introduction of a new curriculum and/or examination, the more likely it is that intended washback will occur. The MoE decision-makers were thus mistaken in expecting sudden changes to occur in the manner they intended. Moreover, decision-makers should be aware that the introduction of a new examination will not result in positive change if the associated curriculum materials are not considered appropriate by the users.

Therefore, to bring about positive changes in teaching and learning as intended by the MoE decision-makers, the process of introducing a new examination should be evaluated with great care and should cover not only the examination specifications but also the other elements of teaching and learning (curriculum, materials, etc.) and the attitudes of the relevant stakeholders, especially the teachers, and the impact on their teaching. The results of the evaluation process should be disseminated to other key parties in the system, so that appropriate decisions can be made concerning the alignment between the textbook and the examination, the adequacy of communication, the planning of further teacher training and the allocation of teaching resources. Only when stakeholders within the educational system work together can fundamental changes in teaching and learning occur.

Furthermore, washback takes time to manifest and it may take different forms at different stages of the process. Thus, it is advisable not to be too ambitious when planning to introduce a new examination. It is reasonable to seek ways of inducing teachers to change the content of their teaching, but it is more difficult to attain a change in their methods and their beliefs about what is appropriate in pedagogical terms. It is important, therefore, to think about what an examination would need to look like to encourage teachers to change how they teach (e.g. to persuade them to employ effective strategies). Furthermore, there is the issue of a lack of training and limited language competence among teachers in this context, as highlighted by previous local studies (see 2.2.5) and as such, any change in the examination design would need to be accompanied by intensive professional training programmes to enhance teachers' ability to employ communicative teaching practices in the classroom and develop their testing methods. This could lead to the desired shift away from teacher-centred instruction to a more learner-centred form of English language learning.

It is also essential to draft examination specification documents that can be scrutinised by different parties in the educational system before being adopted as official blueprints for implementation in schools. The specification should include samples of examination items and tasks, along with marking criteria, so all users can be clear on the demands and requirements of the examination. The specification should provide information about the purpose of the examination and the underlying constructs and include a description of the intended levels of the candidates, and so forth. Moreover, it would be useful for stakeholders in the MoE to plan the examination materials and training programme in line with the examination design, so that issues such as authenticity and possible resistance from teachers due to misalignment with their attitudes and beliefs can be dealt with from the earliest stages of the reform and adjustments can be made throughout the process.

The results show that the teachers were much more inclined to focus on summative assessment strategies in their teaching rather than continuous assessment due to the pressures caused by the conflict between the design and content of the examination and the requirements of continuous assessment as intended by the MoE decision-makers. The study thus provides further evidence that the influence of high-stakes tests is considerably more powerful than classroom assessment. It will be necessary to gain a fuller understanding of why the purposes of classroom assessment are apparently not being achieved in terms of strategies and scope in Oman. This will provide insights into the application of continuous assessment, informing MoE stakeholders of the role such assessment plays in the teaching and learning system as a whole and what they can do to ensure classroom assessment is

implemented to support student attainment as required. Moreover, communication and cooperation between MoE stakeholders and teachers would help understand the demands placed on the latter by the new forms of continuous assessment and classroom activities aimed at improving students' learning rather than awarding scores. This might help teachers apply the criteria and strategies for continuous assessment in a manner more consistent with the intentions of the MoE. It could also induce the desired change in teachers' behaviours, allowing them to enhance their classroom teaching and assessment practices to improve students' learning. The MoE stakeholders would also benefit from this process as they could gain insights from teachers about ways of improving the examination and teaching activities. Direct input from teachers would be useful to assure the quality of examination design and specifications, conferring greater rigour and clarity. This would in turn increase the validity of the examination as an assessment of students' learning. In all, communication and collaboration are useful means of fostering a shared understanding of the principles underlying the curriculum and assessment among all stakeholders, thereby enhancing teaching, learning and assessment.

Regarding the methodological implications, the interviews with the teachers were particularly beneficial for this study in gaining a greater understanding of responses to the questionnaire. Although the number of participants in the interviews was small (five teachers), the data collected revealed many facets of the influence of the examination and were therefore helpful in mapping the scope of washback. The interview data also helped in identifying and clarifying the existence of external factors beyond the examination itself that drove or hindered change, such as those relating to features of the educational setting, which seemed to have a strong influence on the English elective classes. Hence, washback studies should not employ a single method of data collection but rather a combination that can provide complementary evidence of washback (see Hayes, 2003; Saglam, 2018). Moreover, the timing and sequencing is potentially important. For example, if additional interviews had been conducted after the lesson observations for the three teachers, it could have helped understand why some of the MoE decision-makers' intentions were addressed in the classes of Teacher A, but others were not at all, as in Teacher B's and Teacher C's classes.

Additionally, this appears to be one of the few washback studies to have utilised interviews with stakeholders at the front end of reform, in this case the MoE decision-makers. The study suggests that asking policymakers and test designers about their intentions in amending the examination process is crucially important, especially if there are no policy statements/documents available addressing the goals of introducing a new examination or

test, as was the case in this research context. Indeed, as stated in Chapter 4 on methodology (see 4.7.3), gathering data from a range of participants is certainly helpful for exploring the washback effect on teachers' behaviours from a variety of angles, including both the perspectives of the "influencer" (the decision-makers behind the examination design) and "the influenced" (the users of the test, including teachers). This study found it valuable to use a pre-defined interview protocol and coding scheme for MoE decision-makers and recommends this as a methodological tool for investigating washback with those responsible for test design or policy decisions, keeping in mind the research problem and context, since this might enable the results to be compared across different research contexts.

Furthermore, this study employed a teacher questionnaire that can be drawn upon in future washback studies, particularly those conducted in other Middle Eastern countries, given their similar exam-oriented contexts. The questionnaire was adapted from Cheng's (2005) washback study in Hong Kong (see 4.7.1). Based on its utility, this study argues the value of administering a single uniform questionnaire to survey teachers' perceptions and attitudes about language testing reform in the field of washback, taking into consideration the features of the examination/test under investigation and the characteristics of the educational system, the aims of the study, the research problem. As proposed by Bailey (1999), "the systematic development of a widely usable questionnaire for teachers and another for students would be a valuable contribution to the available methodological tools for investigating washback" (p. 38) and Cheng's (2005) teacher questionnaire potentially provides the opportunity for washback studies to yield results that can be compared and contrasted.

7.6 Recommendations for Further Research

This study investigated the introduction of a new English elective examination (the change agent) associated with a new curriculum in terms of the washback on teaching in grade 12 Omani public schools. It looked at the intentions of the MoE decision-makers in introducing the changes and teachers' attitudes and behaviours in response to the new English elective examination. The combination of quantitative and qualitative approaches achieved an overall understanding of the washback phenomenon. Thus, one recommendation for further washback research would be to consider the balance between the breadth and the specificity of the object of enquiry. Taking this study as an example, one recommendation would be to study the washback effect among teachers with different teaching styles. This study has shown that the washback effect varies from one teacher to another and that washback was obvious for some teachers but not for others. To understand more fully how and why teachers

teach in the way that they do, careful analysis is needed as the answers are not immediately clear from classroom observations.

Another direction for future research could be to investigate washback on teaching and learning by examining the textbooks and teaching materials used in the classroom. As mentioned earlier in this thesis, it is evident in the data from the interviews and the classroom observations that washback related to the textbook was one area that showed intensity in this study. The teachers tended to use the textbook as the main source for their classroom teaching; their scheme of work was structured based on the textbook design and they would arrange their teaching according to this scheme of work, thus affecting the content of the lessons. This likely implies that teachers' reliance on the existing textbook, which is composed of test-like tasks, has the potential to constrain teaching approaches. The extent to which the new examination has a washback effect may largely be determined by how the authors of the textbook interpret the examination. A detailed analysis of the new textbooks could therefore be key in studying washback. However, this was beyond the scope of the thesis, which focused on teachers' perceptions and behaviours.

The findings of this study illustrate several challenges regarding the teaching of the new English elective that would be worthy of further research. These include issues relating to training and the teachers' experience in adopting changes in their classroom practice. What is needed is not only opportunities to practise the methods emphasised in the new curriculum, but a sound approach to enhancing English language proficiency level of students in this context. This study has shown that changing the curriculum and examination would mainly affect what the teachers teach but not how they teach. Therefore, there is a need to study other factors in the educational system and find a better means of promoting communicative teaching practices in this context. Factors that are thought to impact the process of teaching include teachers' beliefs about the new examination and their teaching practices, teachers' characteristics and inadequate teacher training opportunities, and a lack of communication between relevant stakeholders (see 3.8). Investigating these factors and their interactions would provide insights into the washback process and suggest ways of promoting positive washback.

Although this study focused on the perceptions of teachers as they are considered the most important key stakeholders in washback studies, it would be beneficial to carry out similar studies focusing on students' perceptions. More washback studies should incorporate the perceptions of students as they are directly affected by test results (see Cheng, 2008; Pan,

2014; Spratt, 2005). Therefore, further studies should focus on the relationship between what students think and know about the test, how these perceptions influence their reactions to the test and how washback operates for students, as these areas have often been neglected or not explicitly addressed in the washback literature.

Another key issue is that previous studies have shown that washback can develop over time (see Shohamy et al., 1996; Wall & Horak, 2011), which suggests that a longitudinal approach would be the best way of observing and monitoring washback and its development. Follow-up studies will be indispensable for investigating the long-term washback from the new English elective examination. Since the MoE decision-makers considered the pass rates to be a key indicator of performance in higher education, it is believed the impacts will increase over time.

7.7 Personal Reflections

Before concluding this thesis, the following paragraphs describe how undertaking this PhD has contributed to my learning and development. The different stages of this research undertaken over four years of study have greatly contributed to my learning. I found that having an interest in and clear awareness of the topic and context from the start helped me continue over the long years with confidence, passion and determination. My personal enthusiasm for the field of language testing, as well as my expectations of the impact that the reform of the curriculum and assessment might bring to teaching practices in Oman and the contribution of the research topic to the existing literature, helped me maintain my drive towards the end of this study.

Through this process, I also learned many strategies and skills, such as critical thinking, working with groups and taking decisions. Additionally, I learned the importance of being systematic and organised. Moreover, my commitment and improved time management skills have helped me reduce my feelings of anger, stress and anxiety. Furthermore, this experience has helped me to develop my research skills and learn new things as a novice researcher, starting from research methodology and design to the writing of the thesis. I have become more analytical with regard to my workplace. I am now looking at my environment through the lens of scientific research enquiry. I question policies and am more aware of the challenges and needs in the context of the reform process.

Since this research has been sponsored by the Ministry of Higher Education and Innovation in Oman, its implications, recommendations and suggestions will hopefully be useful for the ongoing reforms to the educational system and English language teaching, both in Oman

and in other countries with similar educational contexts. This study has contributions to make to the development of policies in terms of proposing appropriate strategies for examination reform and ensuring the effectiveness of the educational system. I have realised that any reform employing a new initiative in the educational system requires a deep and comprehensive understanding of the innovation and its potential impact. Appropriate implementation strategies would likely boost learners' motivation to learn English and improve their English proficiency and performance in examinations. I would recommend that decisions regarding educational change be structured following a bottom-up rather than a top-down strategy to give teachers greater opportunities to become involved and participate in building future plans rather than having them imposed from on high.

7.8 Concluding Remarks

As mentioned in Chapter 1, I chose this research topic because of my expectations and personal curiosity about the potential changes that examination reform would bring to the teaching process. The introduction of reform provokes anxiety among teachers, students and other stakeholders. In particular, the role of teachers in addressing the demands and requirements of examination reform and adopting the change is crucial, since teaching affects students' learning. In this case, three years after the introduction of the new high-stakes examination, the participating grade 12 English language teachers appeared not to have adapted to the change. Thus, this study suggests tests or examinations should be aligned with the teaching objectives and aims and appropriate teacher training and development should accompany them. This study confirmed my belief that it is possible to improve students' English through enhanced teaching strategies and methods, but that for this to happen, it is necessary to take into account and align the intentions of MoE stakeholders with the needs of teachers and their students.

References

- Abu-Alhija, F. N. (2007). Large-scale testing: Benefits and pitfalls. *Studies in Educational Evaluation, 33*(1), 50–68.
- Agrawal, M. (2004). Curricular reform in schools: The importance of evaluation. *Journal of Curriculum Studies, 36*(3), 361–379.
- Aiken, L. (1997). *Questionnaires and inventories: Surveying opinions and assessing personality*. John Wiley.
- Airasian, P. W. (1988). Measurement driven instruction: A closer look. *Educational Measurement: Issues and Practice, 7*(4), 6–11.
- Al Amin, M., & Greenwood, J. (2018). The examination system in Bangladesh and its impact: On curriculum, students, teachers and society. *Language Testing in Asia, 8*(1), 1–18.
- Al Balushi, K. (2017). “...they feel that they have a voice and their voice is heard”: *Towards participatory forms of teachers’ CPD in Oman* [Unpublished doctoral dissertation]. University of Exeter, UK.
- Al Balushi, S., & Griffiths, D. (2013). The school education system in the Sultanate of Oman. G. Donn & Y. Al Manthri (Eds.), *Education in the broader Middle East: Borrowing a baroque arsenal* (pp. 107–126). Symposium Books
- Al- Hinai, M. K., & Al Jardani, K. S. (2021). Washback in language testing: An exploration with a focus on a specific EFL context in Oman. *International Journal of English Linguistics, 11*(1), 68–74.
- Al- Kharusi, H. (2011). Teachers' classroom assessment skills: Influence of gender, subject area, grade level, teaching experience and in-service assessment training. *Journal of Turkish Science Education, 8*(2), 39–48.
- Al Mamari, M., Al-Mekhlaif, A., & Al-Barwani, T. (2018). Is communicative language teaching being tested communicatively? An analysis of English tests in Oman. *English Language Teaching, 11*(11), 83–91.
- Al- Nabhani, M. (2007). *Developing the education system in the Sultanate of Oman through implementing total quality management: The Ministry of Education central headquarters a case study* [Unpublished doctoral dissertation]. University of Glasgow, UK.
- Al Sawafi, O. S. (2014). *Investigating English teachers’ beliefs and practices in relation to the continuous assessment reform in the Sultanate of Oman* [Unpublished doctoral dissertation]. University of Leeds, UK.

- Al-Balushi, R. (2001). ELT in the Sultanate of Oman. *RELO Newsletter*, 5, 5–6.
- Al-Bulushi, M., Al-Said, S., & Lietzén, O. (2018). Teacher competencies in Oman: How gender and teaching experience are associated with self-assessment. *Education*, 5(3), 261–282.
- Alderson, J. C. (1992). Guidelines for the evaluation of language education. In J. C. Alderson & A. Beretta (Eds.), *Evaluating second language education* (pp. 274–304). Cambridge University Press.
- Alderson, J. C. (2004). Foreword. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.) *Washback in language testing: Research contexts and methods* (pp. ix–xii). Lawrence Erlbaum Associates.
- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13(1), 280–297.
- Alderson, J. C., & Wall, D. (1993) Does washback exist? *Applied Linguistics*, 14(2), 115–129.
- Al-Dhafiry, A. (2003). Unemployment in the Gulf countries: Reasons and remedies. *Applied Econometrics and International Development*, 3(3), 61–82.
- AL-Hammami, H. (1999). *Education for the 21st century general education reform in the Sultanate of Oman: Motives, nature and strategies of implementation* [Unpublished doctoral dissertation]. University of Birmingham, UK.
- Al-Issa, A. (2007). An ideological discussion of the implications of implementing a “flexible” syllabus for ESL policy in Sultanate of Oman (August, 2007). *RELC (Regional Language Centre) Journal*, 38(2), 199–215.
- Al-Issa, A. S. (2002). *An ideological and discursive analysis of English language teaching in the Sultanate of Oman* [Unpublished doctoral dissertation]. University of Queensland, Australia.
- Al-Issa, A. S. (2010). ELT transformations in Oman: A critical perspective. Paper presented at *TESOL Arabia Conference, Zayed University, Dubai, United Arab Emirates*.
- Al-Issa, A., & Al-Bulushi, A. (2010). Training English language student teachers to become reflective teachers. *The Australian Journal of Teacher Education*, 35(4), 41–64.
- Al-Issa, A., & Al-Bulushi, A. (2011). English language teaching reform in Sultanate of Oman: The case of theory and practice disparity. *Educational Research for Policy and Practice*, (11), 141–176.
- Al Jabri, M. J. H., Silvennoinen, H., & Griffiths, D. (2018). Teachers' professional

- development in Oman: Challenges, efforts and solutions. *International Journal of Learning, Teaching and Educational Research*, 17(5), 82–103.
- Al-Kharusi, B., & Atweh, B. (2008). *Civic participation and educational reform in the Sultanate of Oman*.
https://oneworldripples.com/uploads/3/5/1/5/35158413/alkarusi_atweh_2008_aare.pdf.
- Al-Lamki, N. (2009). *The beliefs and practices related to continuous professional development of teachers of English in Oman* [Unpublished doctoral dissertation]. University of Leeds, UK.
- Allen, D. (2016). Investigating washback to the learner from the IELTS test in the Japanese tertiary context. *Language Testing in Asia*, 6(1), 1–20.
- Allen, P., Fröhlich, M., & Spada, N. (1984). The Communicative Orientation of Language Teaching: An observation scheme. In D. Handscombe, R. Orem, & B. Taylor (Eds.), *On TESOL 1983* (pp. 231–252). TESOL.
- Allwright, D., & Bailey, K. (1991). *Focus on the language classroom*. Cambridge University Press.
- Al-Mahrooqi, R. (2012). English communication skills: How are they taught at schools and universities in Oman? *English Language Teaching*, 5(4), 124–130.
- Al-Mahrooqi, R., Abrar-Ul-Hassan, S., & Asante, C. C. (2012). Analyzing the use of motivational strategies by EFL teachers in Oman. *Malaysian Journal of ELT Research*, 8(1), 36.
- Al-Mamari, A. (2012). General foundation program in higher education institutions in Oman national standards: Implementation & challenges. Paper presented at the *Oman Quality Network Regional Conference, Muscat, Oman*.
- Al-Mashikhi, E., Al-Mahrooqi, R. & Denman, C. J. (2014). Investigating college of science student attitudes towards using English as a medium of instruction. *The 2014 WEI International Academic Conference Proceedings New Orleans, USA*.
<https://www.researchgate.net/publication/283122177>.
- Al-Najar, N. (2016). View of education development in Oman. *International Journal of Academic Research in Education and Review*, 4(1), 10–18.
- Al-Seyabi, F., & Tuzlukova, V. (2014). Writing problems and strategies: An investigative study in the Omani school and university context. *Asian Journal of Social Sciences & Humanities*, 3(4), 37–48.
- Al-Shabibi, A. S., & Silvennoinen, H. (2018). Challenges in education system affecting teacher professional development in Oman. *Athens Journal of Education*, 5(3),

- Al-Shaqsi, H. (2012). Unemployment in Oman. *Muscat, Oman: Daily*.
<http://www.muscatdaily.com>.
- Al-Sheraiqi, K. R. (2010). *Washback effect of the CEPA English test on teaching in an Educational Zone in the UAE* [Unpublished doctoral dissertation]. American University of Sharjah.
- Al-Toubi, S. S. (1998). *A perspective on change in the Omani ELT curriculum: Structural to communicative* [Unpublished Master's thesis]. University of Bristol, UK.
- Amengual-Pizarro, M. (2009). Does the English test in the Spanish university entrance examination influence the teaching of English? *English Studies*, 90(5), 582–598.
- Amer, A. A. (1993). Teaching EFL students to use a test-taking strategy. *Language Testing*, 10(1), 71–77.
- Anderson, N. (1998). The people make the paradigm. *Journal of Organizational Behavior*, 19, 323–328.
- Andrews, S. (1994). The washback effect of examinations: Its impact upon curriculum innovation in English language teaching. *Curriculum Forum*, 4(1), 44–58.
- Andrews, S. (1995). Washback or washout? The relationship between exam reform and curriculum innovation. In D. Nunan, R. Berry, & V. Berry (Eds.), *Bringing about change in language education. Hong Kong: Proceedings of the International Language in Education Conference, Department of Curriculum Studies, University of Hong Kong*.
- Andrews, S. J. (2001). Reflecting on washback: High-stakes tests and curriculum innovation. In *International Language in Education Conference*. Hong Kong Institute of Education.
- Andrews, S. (2004) Washback and innovation. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing* (pp. 37–50). Lawrence Erlbaum Association.
- Andrews, S., Fullilove, J., & Wong, Y. (2002). Targeting washback – A case study. *System*, 30(2), 207–223.
- Angrosino, M. (2007). *Doing ethnographic and observational research*. Sage.
- Ary, D., Jacobs, L. C., & Sorensen, C. (2010). *Introduction to research in education* (8th ed.). Thomson Wadsworth.
- Atkins, L., & Wallace, S. (2015). Interviewing in education research. In L. Atkins & S. Wallace (Eds.), *Qualitative research in education* (pp. 85–106). Sage Publications.
- Atta, N., Ahmed, N., Ahmed, M., & Ali, Z. (2012). Role of gender and teaching experience on teachers' self-efficacy. *Language in India*, 12(9), 251–262.

- Au, W. (2007). High stakes testing and curricular control: A qualitative meta synthesis. *Educational Researcher*, 36(5), 258–267.
- Azadi, G., & Gholami, R. (2013). Feedback on washback of EFL tests on ELT in L2 classroom. *Theory and Practice in Language Studies*, 3(8), 1335.
- Babbie, E. (2004). *The practice of social research*. Wadsworth Publishing.
- Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing*, 17(1), 1–42.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.
- Bachman, L. F., & Palmer, A. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.
- Bailey, K. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257–279.
- Bailey, K. (1999). *Washback in language testing*. Princeton, New Jersey: ETS.
- Bailey, K. D. (1994). *Methods of social research* (4th ed.). The Free Press.
- Barnes, M. (2017). Washback: Exploring what constitutes “good” teaching practices. *Journal of English for Academic Purposes*, 30, 1–12.
- Belkbir, R. (2019). Investigating the impact of English for Specific Purposes (ESP) curriculum on Moroccan graduates’ career. *International Journal of Language and Literary Studies*, 1(1).
- Berg, B. L., & Lune, H. (2012). *Qualitative research methods for the social sciences* (8th ed.). Pearson Education.
- Bergman, M. M. (2010). On concepts and paradigms in mixed methods research. *Journal of Mixed Methods Research*, 4(3), 171–175.
- Berry, V., & Lewkowicz, J. (2000). Exit-tests: Is there an alternative? *Hong Kong Journal of Applied Linguistics*, 5(1), 19–49.
- Biesta, G. J. J. (2010). Pragmatism and the philosophical foundations of mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social and behavioral research* (pp. 95–118). Sage.
- Biggs, J. B. (1995). Assumptions underlying new approaches to educational assessment: Implications for Hong Kong. *Curriculum Forum*, 4(2), 1–22.
- Birjand P. & Taqizadeh M. (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),

- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
- Black, P., & Wiliam, D. (2003). In praise of educational research: Formative assessment. *British Educational Research Journal*, 29(5), 623–637.
- Black, P., Harrison, C., & Lee, C. (2003). *Assessment for learning: Putting it into practice*. McGrawHill Education.
- Blewchamp, P. (1994). *Washback in TOEFL classrooms: An exploratory investigation into the influence of the TOEFL test on teaching content and methodology* [Unpublished Master's thesis]. University of Lancaster.
- Bogdan, R. C., & Biklin S. K. (1998). *Qualitative research for education: An introduction to theory and methods* (3rd ed.). Allyn and Bacon.
- Booth, D. K. (2012). *Exploring the washback of the TOEIC in South Korea: A sociocultural perspective on student test activity*. The University of Auckland.
- Borg, S. (2005). Teacher cognition in language teaching. In K. Johnson (Ed.), *Expertise in second language learning and teaching* (pp.190–209). Palgrave Macmillan, London.
- Bowen, A. G. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27–40.
- Brannen, J. (1992). *Mixed methods: Qualitative and quantitative research*. Ashgate Publishing.
- Brannen, J. (2005). Mixed methods: The entry of qualitative and quantitative approaches into the research process. *International Journal of Social Research Methodology*, 8(3), 173–184.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. Sage.
- Breakwell, G. M. (2012). Interviewing. In G. M. Breakwell, J. A. Smith, & D. B. Wright (Eds.), *Research methods in psychology* (4th ed., pp. 439–460). Sage.
- Brown, D. H. (2004). *Language assessment principles and classroom practices*. Pearson Education.
- Brown, J. D. (1997). Designing surveys for language programs. In D. Griffiee & D. Nunan (Eds.), *Classroom teachers and classroom research* (pp. 109–122). The Japan Association for Language Teaching.

- Brown, J. D. (1997). The washback effect of language tests. *University of Hawaii Working Papers in ESL*, 16(1), 27–45.
- Bryant, C. G. (1985). *Positivism in social theory and research*. St. Martin's.
- Bryman, A. (1988). *Quantity and quality in social research*. Unwin Hyman.
- Bryman, A. (2004). *Social research methods* (2nd ed.). Oxford University Press.
- Bryman, A. (2007). Barriers to integrating quantitative and qualitative research. *Journal of Mixed Methods Research*, 1(1), 8–22.
- Bryman, A. (2012). *Social research methods* (4th ed.). Oxford University Press.
- Buck, G. (1988). Testing listening comprehension in Japanese university entrance examinations. *JALT Journal*, (10), 12–42.
- Burchell, B., & Marsh, C. (1992). The effect of questionnaire length on survey response. *Quality and Quantity*, 26(3), 233–244.
- Burrows, C. (1998). *Searching for washback: An investigation of the impact on teachers of the implementation into the Adult Migrant English Program of the assessment of the Certificate in Spoken and Written English* [Unpublished doctoral dissertation]. University of New South Wales.
- Burrows, C. (2004). Washback in classroom-based assessment: A study of the washback effect in the Australian adult migrant English program. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 113–128). Lawrence Erlbaum Associates.
- Cavan, S. (1977). Review of J. D. Douglas's (1976) "Investigative social review: Individual and team field research." *The American Journal of Sociology*, 83(3), 809–811.
- Chalhoub-Deville, M., & Deville, C. (2008). Utilizing psychometric methods in assessment. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed., Vol. 7, pp. 211–224). Springer Science + Business Media LLC.
- Chamberlain, S., Daly, A. L., & Spalding, V. (2011). The fear factor: Students' experiences of test anxiety when taking A-level examinations. *Pastoral Care in Education*, 29(3), 193–205.
- Chapman, D. W., & Snyder Jr, C. W. (2000). Can high stakes national testing improve instruction: Re-examining conventional wisdom. *International Journal of Educational Development*, 20(6), 457–474.

- Chappell, P., Bodis, A., & Jackson, H. (2015). The impact of teacher cognition and classroom practices on IELTS test preparation courses in the Australian ELICOS sector. *IELTS Research Reports Online Series*, (6), 1–61.
- Chen, L. M. D. (2002). *Taiwanese junior high school English teachers' perceptions of the washback effect of the basic competence test in English* [Unpublished doctoral dissertation]. Ohio State University.
- Chen, L. M. (2006). Washback effects on curriculum innovation. *Academic Exchange Quarterly*, 10(2), 206–210.
- Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, 11(1), 38–54.
- 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. (2000). *Washback or backwash: A review of the impact of testing on teaching and learning*. Opinion Papers, ERIC ED442280.
- Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. Cambridge University Press.
- Cheng, L. (2008). Washback, impact and consequences. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopaedia of language and education, Vol. 7: Language testing and assessment* (2nd ed, pp. 349–364). Springer.
- Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 3–17). Lawrence Erlbaum.
- Cherryholmes, C. H. (1992). Notes on pragmatism and scientific realism. *Educational Researcher*, 21(6), 13–17.
- Choi, E., & Lee, J. (2017). EFL teachers' self-efficacy and teaching practices. *ELT Journal*, 72(2), 175–186.
- Choi, I. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing*, 25(1), 39–62.
- Choi, S. Y. (2015). *Washback of TOEFL iBT in Korean commercial test-preparation institutes* [Unpublished doctoral dissertation]. University of Bristol, UK.
- Cholis, H. W. N., & Rizqi, F. (2018). Senior High School English Teachers' Perceptions on a High-Stakes Test (SBMPTN): A washback study. *International Journal of Education and Literacy Studies*, 6(3), 47–52.

- Cinkara, E., & Tosun, Ö. Ö. (2017). Face validity study of a small-scale test in a tertiary-level intensive EFL program. *Bartın Üniversitesi Eğitim Fakültesi Dergisi*, 6(2), 395–410.
- Clark, V. L. P., & Creswell, J. W. (2008). *The mixed methods reader*. Sage.
- Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education* (5th ed.). Routledge Falmer.
- Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). Routledge.
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.
- Collis, J., & Hussey, R. (2003). *Business research* (2nd ed.). Palgrave Macmillan.
- Corbin, J., & Strauss, A. (2008). Strategies for qualitative data analysis. In J. Corbin & A. Strauss (Eds.), *Basics of qualitative research. Techniques and procedures for developing grounded theory* (3rd ed., pp. 158–327). Sage.
- Creswell, J. W. (2009). *Research design: Qualitative and mixed methods approaches*. Sage.
- Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Sage.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Sage.
- Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). Sage.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438–481.
- Curtis, A. (2000). A problem-solving approach to the management of change in language education. *Korea TESOL Journal*, 3(1), 1–12.
- Dammak, H., Khatibi, A., & Azam, S. F. (2022). Tunisian English language teachers' perceptions and practices: Washback of the English Baccalaureate Exam. *Journal of Language Testing & Assessment*, 5(1), 22–43.
- Darlington, Y., & Scott, D. (2002). *Qualitative research in practice: Stories from the field*. Allen and Unwin.
- Davies, A. (1968). Language testing symposium: A psycholinguistic approach. *Language and Language Learning [Series]*, No. 21.
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of language testing* (Vol. 7). Cambridge University Press.

- Davison, C. (2008). Using summative assessments for formative purposes: The ultimate justification for learners and teachers. Presented at *30th Annual Language Testing Colloquium, (LTRC 08), Hangzhou*.
- Dawadi, S. (2021). Factors affecting washback of a high-stakes English as a foreign language test. *TESL-EJ: The Electronic Journal for English as a Second Language*, 25(3).
- Denman, C. J., & Al-Mahrooqi, R. (2019). Quality assurance in English education in Omani schools: A critical look. In R. Al-Mahrooqi & C. J. Denman (Eds.), *English Education in Oman: Current Scenarios and Future Trajectories*. Singapore: Springer. https://doi:10.1007/978-981-13-0265-7_4.
- Denscombe, M. (2003). *The good research guide*. Open University Press.
- Denscombe, M. (2008). Communities of practice: A research paradigm for the mixed methods approach. *Journal of Mixed Methods Research*, 2(3), 270–283.
- Denscombe, M. (2014). *The good research guide: For small-scale social research projects*. McGraw-Hill Education.
- Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods* (2nd ed.). McGraw Hill.
- Denzin, N. K. (2009). *The research act: A theoretical introduction to sociological methods* (3rd ed.). Prentice Hall.
- Denzin, N. K., & Lincoln, Y. S. (1994). Paradigms and perspectives in transition. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 245–252). Sage.
- Denzin, N. K., & Lincoln, Y. S. (2013). *The landscape of qualitative research* (4th ed.). Sage.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). John Wiley & Sons.
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford University Press.
- Duff, P. A. (2006). Beyond generalizability: Contextualization, complexity, and credibility in applied linguistics research. In M. Chalhoub-Deville, C. A. Chapelle, & P. Duff (Eds.), *Inference and generalizability in applied linguistics: Multiple perspectives* (pp.65-95). John Benjamins.
- Dunkley, D. (2010). Washback in language testing. *Foreign Languages & Literature*, 35(1), 115–128.

- East, M. (2015). Coming to terms with innovative high-stakes assessment practice: Teachers' viewpoints on assessment reform. *Language Testing*, 32(1), 101–120.
- East, M. (2016). *Assessing foreign language students' spoken proficiency: Stakeholder perspectives on assessment innovation* (Vol. 26). Springer.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford University Press.
- Estaji, M., & Tajeddin, Z. (2012). The learner factor in washback context: An empirical study investigating the washback of the IELTS academic writing test. *Language Testing in Asia*, 2, 1–21.
- Falvey, P. (1996). The education of teachers of English in Hong Kong: A case for special treatment. *Proceedings of ITEC' 95*, International Teacher Education Conference, Hong Kong, 5–7 June 1995, pp. 107–113.
- Fan, J., Frost, K., & Liu, B. (2020). Teachers' involvement in high-stakes language assessment reforms: The case of Test for English Majors (TEM) in China. *Studies in Educational Evaluation*, 66, 1–10.
- Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods*, 5, 1–11.
- Ferman, I. (2004). The washback of an EFL national oral matriculation test to teaching and learning. In L. Cheng & Y. Watanabe (Eds.), *Washback in language testing: Research contexts and methods* (pp. 213–232). Routledge.
- Fish, J. (1988). *Responses to mandated standardized testing*. Unpublished PhD thesis, University of California.
- Flick, U. (2011). *Introducing research methodology: A beginner's guide to doing a research project*. Sage.
- Flick, U. (2014). *An introduction to qualitative research*. Sage.
- Fournier-Kowaleski, L. A. (2005). *Depicting washback in the intermediate Spanish language classroom: A descriptive study of teacher's instructional behaviors as they relate to tests* [Unpublished doctoral dissertation]. The Ohio State University.
- Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, 19(9), 27–32.
- Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist*, 39(3), 193–202.
- Freeman, J. (1986). Data quality and the development of organizational social science: An editorial essay. *Administrative Science Quarterly*, 31(2), 298–303.
- Fulcher, G., & Davidson, F. (2012) *The Routledge handbook of language testing*.

Routledge.

- Fullan, M. (2001). *The new meaning of educational change* (4th ed.). Teachers College Press.
- Fullan, M. (2015). *The new meaning of educational change* (5th ed.). Teachers College Press.
- Fullan, M. G., & Stiegelbauer, S. (1991). *The new meaning of educational change* (2nd ed.). Cassell.
- Furaidah, F., Saukah, A., & Widiati, U. (2015). Washback of English national examination in the Indonesian context. *TEFLIN Journal*, 26(1), 36–58.
- Gates, S. (1995). Exploiting washback from standardized tests. In J. D. Brown & S. Yamashita (Eds.), *JALT applied materials: Language testing in Japan* (pp. 101–106). Japan Association for Language Teaching.
- Gay, L. R., & Airasain, P. (2003). Characteristics of qualitative research. In L. R. Gay & P. Airasain (Eds.), *Educational research: Competencies for analysis and applications* (pp. 163–172). Merrill Prentice Hall.
- Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment*, 6(4), 304.
- Genesee, F. & Upshur, J. (1996). *Classroom-based evaluation in second language education*. Cambridge University Press.
- Gillham, B. (2005). *Research interviewing: The range of techniques: A practical guide*. McGraw-Hill Education.
- Gillham, B. (2007). *Developing a questionnaire* (2nd ed.). Continuum.
- Gipps, C. (1999). Chapter 10: Socio-cultural aspects of assessment. *Review of Research in Education*, 24(1), 355–392.
- Gipps, C. (2002). *Beyond testing: Towards a theory of educational assessment*. Routledge.
- Gipps, C. (2011). *Beyond testing (Classic edition): Towards a theory of educational assessment*. Routledge.
- Glover, P. G. A. (2006). *Examination influences on how teachers teach: A study of teacher talk* [Unpublished doctoral dissertation]. Lancaster University.
- Glover, P. (2014). Do language examinations influence how teachers teach. *International Online Journal of Education and Teaching (IOJET)*, 1(3), 197–214.
- Gray, D. E. (2013). *Doing research in the real world*. Sage.
- Gray, D. E. (2017). *Doing research in the real world* (4th ed.). Sage.
- Green, A. (2007). IELTS washback in context: Preparation for academic writing in higher

- education. *Studies in Language Testing* 25. UCLES/Cambridge University Press.
- Greene, J. C., & Caracelli, V. J. (1997). Defining and describing the paradigm issue in mixed-method evaluation. *New Directions for Evaluation*, 74, 5–17.
- Grix, J. (2001). *Demystifying postgraduate research*. University of Birmingham Press.
- Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. Jossey-Bass Publishers.
- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105–117). Sage.
- Gunn, J., Al-Bataineh, A., & Abu Al-Rub, M. (2016). Teachers' perceptions of high stakes testing. *International Journal of Teaching and Education*, 4(2), 49–62.
- Haddadin, A., Dweik, B., & Sheir, A. (2008). Teachers' and students' perceptions of the effect of public examinations on English instruction at the secondary stage in Jordan. *Jordanian Journal of Applied Sciences*, 11(2), 331–344.
- Haladyna, T. M., Nolen, S. B., & Haas, N. S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20(5), 2–7.
- Hammersley, M., & Atkinson, P. (2007). *Ethnography: Principles in practice* (3rd ed.). Routledge.
- 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: Principles, Policy & Practice*, 10(2), 169–207.
- Hawkey, R. (2006). *Impact theory and practice: Studies of the IELTS test and Progetto Lingue 2000*. UCLES/Cambridge University Press.
- Hayes, B. M. (2003). *IELTS preparation in New Zealand: An investigation into the nature of the courses and evidence of washback* [Unpublished doctoral dissertation]. Victoria University of Wellington.
- Hayes, B., & Read, J. (2004). IELTS test preparation in New Zealand: Preparing students for the IELTS academic module. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 119–134). Routledge.
- Heaton, J. B. (1990). *Classroom testing*. Longman.

- Henn, M., Weinstein, M. & Foard, N. (2006). *A short introduction to social research*. Sage.
- Herman, J., & Golan, S. (1991). *Effects of standardized testing on teachers and learning: Another look*. National Center for Research on Evaluation, Standards, and Student Testing, UCLA Graduate School of Education.
- Herman, J. L., & Golan, S. (1993). The effects of standardized testing on teaching and schools. *Educational measurement: Issues and practice*, 12(4), 20–25.
- Heron, J., & Reason, P. (1997). A participatory inquiry paradigm. *Qualitative Inquiry*, 3(3), 274–294.
- Heyneman, S. P. (1987). Use of examinations in developing countries: Selection research and education sector management. *International Journal of Education Development* 7, 251–263.
- Hoang, D. T. P. (2020). Washback of English proficiency test in classroom activities at national university of arts education. *VNU Journal of Science: Education Research*, 36(1).
- Hopf, D., & Hatzichriston, C. (1999). Teacher gender-related influences in Greek schools. *British Journal of Educational Psychological Society*, 69, 1–18.
- 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* [Unpublished doctoral dissertation]. Jahangirnagar University, Dhaka, Bangladesh.
- Hsu, H.-F. (2010). *The impact of implementing English proficiency tests as a graduation requirement at Taiwanese universities of technology* [Unpublished doctoral dissertation]. University of York, UK. <https://etheses.whiterose.ac.uk/576/>.
- Hughes, A. (1988). Introducing a needs-based test of English language proficiency into an English medium university in Turkey. In A. Hughes (Ed.), *Testing English for University Study*. *ELT Documents 127* (pp. 134–153). Modern English Publications.
- Hughes, A. (1993). *Backwash and TOEFL*. Unpublished manuscript, University of Reading.
- Hughes, A. (2003). *Testing for language teachers* (2nd ed). Cambridge University Press.
- Hughes, M., & Greenhough, P. (2004). Learning from homework: A case study. In L. Poulson & M. Wallace (Eds.), *Learning to read critically in teaching and learning* (pp. 85–109). Sage.
- Hwang, H.-J. (2003). The impact of high-stakes exams on teachers and students: A

- washback study of the university entrance exam at the secondary school level in South Korea [Unpublished Master's thesis]. McGill University (Canada).
- Ibrahim, M. (2019). The washback effect of the Thanaweya Amma English test: Drawbacks and solutions. *English Language Teaching Research in the Middle East and North Africa: Multiple Perspectives*, 73–91.
- Imsa-ard, P. (2020). Voices from Thai EFL teachers: Perceptions and beliefs towards the English test in the national examination in Thailand. *LEARN Journal: Language Education and Acquisition Research Network*, 13(2), 269–287.
- Ismail, M. A. (2011). *Language planning in Oman: Evaluating linguistic and sociolinguistic fallacies* [Unpublished doctoral dissertation]. Newcastle University, UK.
- Issan, S., & Gomaa, N. (2010). Post Basic Education reforms in Oman: A case study. *Literacy Information and Computer Education Journal (LICEJ)*, 1(1).
- Jaeger, R. M. (1988). *Complementary methods for research in education*. American Educational Research Association.
- Jin, Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing*, 27(4), 555–584.
- Jin, Y., & Fan, J. (2011). Test for English majors (TEM) in China. *Language Testing*, 28(4), 589–596.
- Johnson, K. E. (1994). The emerging beliefs and instructional practices of preservice English as a second language teachers. *Teaching and Teacher Education*, 10(4), 439–452.
- Johnson, M., & Shaw, S. (2018). What is computer-based testing washback, how can it be evaluated and how can this support practitioner research? *Journal of Further and Higher Education*, 43(9), 1255–1270.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14–26.
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112–133.
- 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), 23–29.
- Khaniya, T. R. (1990). *Examinations as instruments for educational change: Investigating the washback effect of the Nepalese English exams* [Unpublished doctoral

- dissertation]. University of Edinburgh, UK.
- Khodabakhshzadeh, H., Zardkanloo, R., & Alipoor, I. (2017). The effect of mock tests on Iranian EFL learners' test scores. *International Journal of Education & Literacy Studies*, 5(3), 47–51.
- Kılıçkaya, 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.
- Kim, H., & Isaacs, T. (2018). Teachers' voices in the decision to discontinue a public examination reform: Washback effects and implications for utilizing tests as levers for change. In D. Xerri & P. V. Briffa (Eds.). *Teacher involvement in high-stakes language testing* (pp. 263–282). Springer.
- Kim, Y. (2011). The pilot study in qualitative inquiry: Identifying issues and learning lessons for culturally competent research. *Qualitative Social Work*, 10(2), 190–206.
- Kirkland, M. C. (1971). The effects of tests on students and schools. *Review of Educational Research*, 41(4), 303–350.
- Kuhn, T. (1970). *The structure of scientific revolutions* (2nd ed.). University of Chicago Press.
- Kvale, S. (1996). *InterViews: An introduction to qualitative research*. Sage.
- Lai, E. R., & Waltman, K. (2008). Test preparation: Examining teacher perceptions and practices. *Educational Measurement: Issues and Practice*, 27(2), 28–45.
- Lam, H. P. (1993). *Washback – Can it be quantified? A study on the impact of English examinations in Hong Kong* [Unpublished Master's thesis]. University of Leeds, UK.
- Lam, H. (1995). Methodology washback – an insider's view. In D. Nunan, R. Berry, & V. Berry (Eds.), *Bringing about change in language education* (pp. 83–102). University of Hong Kong.
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169–197.
<http://dx.doi.org/10.1177/0265532214554321>.
- Leech, N. L., & Onwuegbuzie, A. J. (2009). A typology of mixed methods research designs. *Quality & Quantity*, 43(2), 265–275.
- Li, X. (1990). How powerful can a language test be? The MET in China. *Journal of Multilingual and Multicultural Development*, 11(5), 393–404.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.

- Lincoln, Y. S., and E. G. Guba. (2003). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), *The landscape of qualitative research: Theories and issues* (pp. 253–291). Sage.
- Lincoln, Y. S., Lynham, S. A., & Guba, E. G. (2011). Paradigmatic controversies, contradictions, and emerging confluences, revisited. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (pp. 97–128). Sage.
- Linn, R. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4–16.
- Lodhi, M. A., Robab, I., Mukhtar, S., Farman, H., & Farrukh, S. (2018). Impact of washback on ESL students' performance at secondary level. *International Journal of English Linguistics*, 8(6), 227–239.
- Lumley, T., & Stoneman, B. (2000). Conflicting perspectives on the role of test preparation in relation to learning? *Hong Kong Journal of Applied Linguistics*, 5(1), 50–80.
- Luong-Phan, N. H., & Effeney, G. (2015). TOEFL iBT and language learning motivation: An investigation into teaching styles and influential factors for Vietnamese adolescents. *International Journal of Research Studies in Language Learning*, 4(3), 3–18.
- Mackey, A., & Gass, S. (2005). *Second language research: Methodology and design*. Lawrence Erlbaum Associates, Inc.
- Madaus, G. F. (1988). The influence of testing on the curriculum. In L. N. Tanner (Ed.), *Critical issues in curriculum, 87th Yearbook of the National Society for the Study of Education* (pp. 83–121). University of Chicago Press.
- Madaus, G. F., & Clarke, M. (2001). The adverse impact of high stakes testing on minority students: Evidence from 100 years of test data.
https://archive.org/details/ERIC_ED450183.
- Madaus, G. F., & Kellaghan, T. (1992). Curriculum evaluation and assessment. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 119–154). Macmillan.
- Mahmoudi, L., & Bakar, K. A. (2013). Iranian pre-university English teachers' perceptions and attitudes towards the Iranian National University Entrance Exam: A washback study. *International Journal of Education and Literacy Studies*, 1(2), 47–53.
- Mahmud, N. (2018). *Investigating the washback effect of the MUET as a university entry test on students in Malaysia* [Unpublished doctoral dissertation]. University of York, UK.
- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329–344.

- Mark, S. H. (2011). Tension between conflicting beliefs of an EFL teacher in teaching practices. *RELC Journal*, 42(1), 53–67.
- Markee, N. (1997). *Managing curricular innovation* (Vol. 198). Cambridge University Press.
- Marshall, C., & Rossman, G. B. (2010). *Designing qualitative research* (5th ed.). Sage.
- Mathews, J. C. (1985). *Examinations: a commentary*. George Allen and Unwin.
- Mathison, S. M. (1987). *The perceived effects of standardized testing on teaching and curricula*. University of Illinois at Urbana-Champaign.
- Maxcy, S. (2003). Pragmatic threads in mixed methods research in the social sciences. The search for multiple modes of inquiry and the end of the philosophy of formalism. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp.51-89), Sage.
- Maxwell, J. A. (2005). *Qualitative research design: An interactive approach* (2nd ed.). Sage.
- McCarthy, S. J. (2008). The impact of No Child Left Behind on teachers' writing instruction. *Written Communication*, 25(4), 462–505.
- McDonough, J., & McDonough, S. (1997). *Research methods of English language teachers*. Arnold.
- McNamara, T. & Roever, C. (2006). *Language testing: The social dimension*. Blackwell Publishing.
- McNamara, T. (2000). *Language testing*. Oxford University Press.
- Mehrens, W. A. (1991). *Defensible/indefensible instructional preparation for high stakes achievement tests: An exploratory dialogue*. <https://eric.ed.gov/?id=ED334202>.
- Merriam-Webster. (2000). *Merriam-Webster's collegiate dictionary*. Merriam-Webster.
- Mertens, D. M. (2012). What comes first? The paradigm or the approach? *Journal of Mixed Methods Research*, 6(4), 255–257.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241–256.
- Miles, M. B. (1979). Qualitative data as an attractive nuisance: The problem of analysis, *Administrative Science Quarterly*, 24(4), 590–601.
- Miles, M. B., & Huberman, A. M. (1984). *Qualitative data analysis: A sourcebook of new methods*. Sage.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Sage
- Mills, J., & Birks, M. (2014). *Qualitative methodology: A practical guide*. Sage.

- MoE. (1997). *Guidelines and information for Regional Inspectors of English in the Sultanate of Oman*. Ministry of Education, Sultanate of Oman.
- MoE. (1999). *English language curriculum framework*. Ministry of Education, Sultanate of Oman.
- MoE. (2003). *Report from the consultancy study on the report reform of years 11 and 12*. Zawaia.
- MoE. (2004). *National report on quality education in Oman*. Ministry of Education, Sultanate of Oman.
http://www.ibe.unesco.org/International/ICE47/English/Natreps/reports/oman_part_1.pdf.
- MoE. (2005). *The development of education in the Sultanate of Oman*. Retrieved from: www.moe.gov.om.
- MoE. (2008). *Post-basic education grades 11 and 12*. Ministry of Education, Sultanate of Oman.
- MoE. (2009). *The philosophy and objectives of education in the Sultanate of Oman*. Ministry of Education, Sultanate of Oman.
- MoE. (2010). *Educational revolution in the Sultanate of Oman in numbers 1970–2010*. Ministry of Education, Sultanate of Oman.
- MoE. (2016). *The Annual Educational Statistics Book (Issue Forty-Six, 2015–2016)*. Ministry of Education, Sultanate of Oman.
- MoE. (2017a). *Suggestions for improving the Student Assessment Handbook for English (Comments from Teachers, SETS, RS, SS)*. March. Ministry of Education, Sultanate of Oman.
- MoE. (2017b). *General learning objectives of the Elective Course (English Insights)*. Ministry of Education, Sultanate of Oman.
- MoE. (2018). *Suggestions for improving the Student Assessment Handbook for English (Comments from Teachers, SETS, RS, SS)*. February. Ministry of Education, Sultanate of Oman.
- MoE. (2022). *Students' Assessment Handbook for English grades 11 & 12, 2022*. Ministry of Education, Sultanate of Oman.
- Ministry of Education, Oman & The New Zealand Education Consortium. (2013). *Evaluation of school system in Oman*. Ministry of Education, Sultanate of Oman.
- Ministry of Education, Oman & World Bank. (2012). *Education in Oman: The drive for quality*. Ministry of Education, Sultanate of Oman.
<http://documents.worldbank.org/curated/en/246211468291645003/pdf/757190ESW>

- Mohamad, M., Sarma, M. M., & Mohapatra, D. (2018). Teaching English in ESL and EFL contexts: A comparative study in Syrian and Indian public schools. *Journal of English Language and Literature*, 5(1).
- Mohamed, M. T. (2019). *An exploration of why learning English for twelve years in Omani public schools is inadequate preparation for Omani students entering higher education* [Unpublished doctoral dissertation]. The University of Liverpool, UK.
- Monfils, L. F., Firestone, W. A., Hicks, J. E., Martinez, M. C., Schorr, R. Y., & Camilli, G. (2004). Teaching to the test. In W. A. Firestone, R. Y. Schorr, & L. F. Monfils (Eds.), *The ambiguity of teaching to the test* (pp. 49–74). Routledge.
- Moody, J. (2009). A neglected aspect of ELT in the Arabian Gulf: Who is communication between? In Zhang, L. J., R. Rubdy, & Alsagoff, L. (Eds.), *Englishes and literatures-in English in a globalized world: Proceedings of the 13th International Conference on English in Southeast Asia* (pp. 99–119). National Institute of Education, Nanyang Technological University, Singapore.
- Morgan, G., & Smircich, L. (1980). The case for qualitative research. *Academy of Management Review*, 5(4), 491–500.
- Morris, P. (1988). Teachers' attitudes towards a curriculum innovation: An East Asian study. *Research in Education*, 40, 75–87.
- Morrow, K. (1979). Communicative language testing: revolution or evolution? In C. J. Brumfit & K. Johnson (Eds.), *The communicative approach to language teaching* (pp. 143–157). Oxford University Press.
- Moss, P. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practices*, 22(4), 13–25.
- Muñoz, A. P., & Álvarez, M. E. (2010). Washback of an oral assessment system in the EFL classroom. *Language Testing*, 27(1), 33–49.
- Mutereko, S. (2017). The washback effect of the National Senior Certificate Examinations: Perceptions of teachers in Umgungundlovu district, South Africa. *International Journal of Social Sciences and Humanity Studies*, 9(2), 126–142.
- NCSI. (2023). <http://www.ncsi.gov.om>.
- Noah, H. J., & Eckstein, M. A. (1992). The two faces of examinations: A comparative and international perspective. M. A. Eckstein & H. J. Noah (Eds.), *Examinations: comparative and international studies* (pp. 147–170). Pergamon.
- Noble, A. J., & Smith, M. L. (1994). Old and new beliefs about measurement-driven

- reform: “The more things change, the more they stay the same.” *CSE Technical Report 373*. National Center for Research on Evaluation, Standards, and Student Testing.
- Norris, J. M. (2008). *Validity evaluation in language assessment*. Peter Lang.
- Nunan, D. (1992). *Research methods in language teaching*. Cambridge University Press.
- O, K.-M. (2001). *Washback on high school classrooms of the English tests within Korean university entrance exams*. Columbia University Teachers College, New York.
- Onaiba, A. M. E. M., & Mustafa, A. (2014). *Investigating the washback effect of a revised EFL public examination on teachers' instructional practices, materials and curriculum* [Unpublished doctoral dissertation]. University of Leicester, UK.
- Oppenheim, A. N. (1992). *Questionnaire design, interviewing, and attitude measurement*. Pinter Publishers Ltd.
- Orafi, S. M. S., & Borg, S. (2009). Intentions and realities in implementing communicative curriculum reform. *System*, 37(2), 243–253.
- OECD (2005). *Teachers matter: Attracting, developing and retaining effective teachers*. Organisation for Economic Co-operation and Development.
- Oxford Business Group (2013). *Oman: Budget boost for education*.
<https://oxfordbusinessgroup.com/news/oman-budget-boost-education>.
- Oxford University Press (2000). *Shorter Oxford English Dictionary*. Oxford University Press.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307–332.
- Pan, Y., & Newfields, T. (2011). Teacher and student washback on test preparation evidenced from Taiwan's English certification exit requirements. *International Journal of Pedagogies and Learning*, 6(3), 260–272.
<https://doi.org/10.5172/ijpl.2011.6.3.260>.
- Patton, M. (1987). *How to use qualitative methods in evaluation*. Sage.
- Payne, G., & Payne, J. (2004). *Key concepts in social research*. Sage.
- Pearson, I. (1988). Tests as levers for change. *ESP in the classroom: Practice and evaluation*, 12898–107.
- Popham, W. J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappa*, 68, 679–682.
- Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., & Singer, E. (2004). Methods for testing and evaluating survey questions. *Public Opinion Quarterly*, 68(1), 109–130.

- Pring, R. (2004). Conclusion: Evidence-based policy and practice. In G. Thomas & R. Pring (Eds.), *Evidence-based practice in education* (pp. 201-213). Open University Press.
- Prior, J., & Lister, R. (2004). The backwash effect on SQL skills grading. *ACM SIGCSE Bulletin*, 36(3), 32–36.
- Prodromou, L. (1995). The backwash effect: from testing to teaching. *ELT Journal*, 49, 13–25.
- Punch, K. F., & Oancea, A. (2014). *Introduction to research methods in education*. Sage.
- Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing*, 22(142), 142–173.
- Qi, L. (2007). Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes English test in China. *Assessment in Education*, 14(1), 51–74.
- Quesada Inces, R. (2001). *Washback overrides the curriculum: An exploratory study on the washback effect of a high-stakes standardized test in the Costa Rican EFL high school context*. The Pennsylvania State University.
- Rabe, M. (2003). Revisiting insiders and outsiders as social researchers. *African Sociology Review*, 7(2), 149–161.
- Rahman, K. A., Seraj, P. M. I., Hasan, M. K., Namaziandost, E., & Tilwani, S. A. (2021). Washback of assessment on English teaching-learning practice at secondary schools. *Language Testing in Asia*, 11(1), 12.
- Rahman, K., Hasan, M., Namaziandost, E., & Ibna Seraj, P. M. (2021). Implementing a formative assessment model at the secondary schools: attitudes and challenges. *Language Testing in Asia*, 11(1), 1–18.
- Rao, M. M., & Haque, M. I. (2019). A study on impact of testing on English as a foreign language teaching in a Saudi Arabian University. *Humanities & Social Sciences Reviews*, 7(2), 58–71.
- Rassekh, S. (2004). *Education as a motor for development: Recent education reforms in Oman with particular reference to the status of women and girls*. International Bureau of Education.
- Read, J., & Hayes, B. (2003). The impact of IELTS on preparation for academic study in New Zealand. *International English Language Testing System (IELTS) Research Reports 2003: Volume 4* (pp. 153–191). IELTS Australia.
- Rea-Dickins, P. (1997). So, why do we need relationships with stakeholders in language testing? A view from the UK. *Language Testing*, 14(3), 304–314.

- Rea-Dickins, P., & Scott, C. (2007). Washback from language tests on teaching, learning and policy: Evidence from diverse settings. *Assessment in Education: Principles, Policy & Practice*, 14(1), 1–7.
- Ren, Y. (2011). A study of the washback effects of the College English Test (band 4) on teaching and learning English at tertiary level in China. *International Journal of Pedagogies and Learning*, 6(3), 243–259.
- Resnick, L. B., Resnick, D.P. (1992), Assessing the thinking curriculum: New tools for educational reform. In B. R. G. M. C. O’Connor (Ed.), *Changing assessments: Alternative views of aptitude, achievement and instruction* (pp. 37–75). Kluwer Academic Publishers.
- Richards, J., Platt, J., Weber, H., Inman, P., & Inman, P. (1986). Longman dictionary of applied linguistics. *RELC Journal*, 17(2), 105–110.
- Richards, J. C., & Nunan, D. (1990) *Second language teacher education*. Cambridge University Press.
- Richards, K. (2003). *Qualitative inquiry in TESOL*. Palgrave Macmillan.
- Richards, L. & Richards, T. (1994). Using computers in qualitative research. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (pp. 445–462). Sage.
- Ritchie J., & Lewis, J. (2003). *Qualitative research practice*. Sage.
- Ritchie, J., Lewis, J., & Elam, G. (2003). Designing and selecting samples. In J. Ritchie & J. Lewis (Eds.), *Qualitative research practice: A guide for social science students and researchers* (pp. 77–108). Sage.
- Ritter, S., & Idol-Maestas, L. (1986). Teaching middle school students to use a test-taking strategy. *The Journal of Educational Research*, 79(6), 350–357.
- Robson, C. (2002). *Real world research: A resource for social scientists and practitioner-researchers* (2nd ed.). Blackwell.
- Rolstad, S., Adler, J., & Rydén, A. (2011). Response burden and questionnaire length: Is shorter better? A review and meta-analysis. *Value in Health*, 14(8), 1101–1108.
- Rubin, H. J., & Rubin, I. S. (2012). *Qualitative interview: The art of hearing data* (3rd ed.). Sage.
- Rudman, H. E., Kelly, J. L., Wanous, D. S., Mehrens, W. A., Clark, C. M., & Porter, A. C. (1980). *Integrating assessment with instruction: A review (1922–1980)* (Research Series No. 75). Michigan State University, College of Education.
- Ryan, J. (2011). Teaching and learning for international students: Towards transcultural approach. *Journal of Teachers and Teaching: Theory and Practice*, 17(6), 631–648.

- Saglam, A. L. G. (2018). Can exams change how and what teachers teach? Investigating the washback effect of a university English language proficiency test in the Turkish Context. *Eurasian Journal of Applied Linguistics*, 4(2), 155–176.
- Saif, S. (1999). *Theoretical and empirical considerations in investigating washback: A study of ESL/EFL learners* [Unpublished doctoral dissertation]. University of Victoria.
- Saif, S. (2006). Aiming for positive washback: A case study of international teaching assistants. *Language Testing*, 23(1), 1–34.
- Saldaña, J. (2009). *The coding manual for qualitative researchers*. Sage.
- Saldaña, J. (2021). *The coding manual for qualitative researchers*. Sage.
- Salehi, H., & Yunus, M. (2012). The washback effect of the Iranian universities entrance exam: Teachers' insights. *Journal of Language Studies*, 12(2).
- Salehi, H., Yunus, M. M., & Salehi, Z. (2012). Teachers' perceptions of high-stakes tests: A washback study. *International Journal of Social Science and Humanity*, 2(1), 70–74.
- Sallee, M. W., & Flood, J. T. (2012). Using qualitative research to bridge research, policy and practice. *Theory into Practice*, 51(2), 137–144.
- Sandelowski, M. (2003). Tables or tableaux? The challenges of writing and reading mixed methods studies. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 321–350). Sage.
- Saville, N. D. (2009). *Developing a model for investigating the impact of language assessment within educational contexts by a public examination provider* [Unpublished doctoral dissertation]. University of Bedfordshire.
- Scaramucci, M. V. R., & Kobayashi, E. (2013). Washback effect of Cambridge English: Key for schools in São Paulo, Brazil: A case study. *Research Notes*, (54), 13–20.
- Scott, C. (2007). Stakeholder perceptions of test impact. *Assessment in Education*, 14(1), 27–49.
- Scott, J. (1990). *A matter of record: Documentary sources in social research*. Polity Press.
- SQA. (1996). *Report on visit to Sultanate of Oman*. Scottish Qualifications Authority.
- Seaton, B. (1982). *A handbook of English language teaching: terms and practice*. Macmillan.
- Sevimli, S. E. (2007). *The washback effects of foreign language component of the university entrance examination on the teaching and learning context of English language groups in secondary education: A case study*. [Unpublished Master's thesis]. Institute of Social Sciences, Gaziantep University.

- Shannon-Baker, P. (2016). Making paradigms meaningful in mixed methods research. *Journal of Mixed Methods Research, 10*(4), 319–334.
- Shazadi, T., Khatoon, S., Aziz, S., & Hassan, H. (2011). Determining factors affecting teachers' self-efficacy at secondary schools level. *Language in India, 11*(10), 385–395.
- Shepard, L. A., Flexer, R. J., Hiebert, E. H., Marion, S. F., Mayfield, V., & Weston, T. J. (1996). Effects of introducing classroom performance assessments on student learning. *Educational Measurement: Issues and Practice, 15*(3), 7–18.
- Shih, C. M. (2007). A new washback model of students' learning. *The Canadian Modern Language Review, 64*(1), 135–162.
- Shih, C. M. (2009). How tests change teaching: A model for reference. *English Teaching: Practice and Critique, 8*(2), 188–206.
- Shih, P.-C. (2013). *The English Benchmark Policy for Graduation: An investigation of perception, motivation, and approaches to learning at a University of Technology in Central Taiwan* [Unpublished doctoral dissertation]. Durham University, UK.
- Shohamy, E. (1992). Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning. *The Modern Language Journal, 76*(4), 513–521.
- Shohamy, E. (1993). *The power of tests: The impact of language tests on teaching and learning*. NFLC Occasional Papers.
- Shohamy, E. (1997). Testing methods, testing consequences: Are they ethical? Are they fair? *Language Testing, 14*(3), 340–349.
- Shohamy, E. (2001). *The power of tests: A critical view of the uses of language tests*. Pearson Longman.
- Shohamy, E. (2007). Language tests as language policy tools. *Assessment in Education, 14*(1), 117–130.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing, 13*(3), 298.
- Silverman, D. (Ed.) (2010). *Qualitative research*. Sage.
- Silverman, D. (2005). *Doing qualitative research: A practical handbook* (2nd ed.). Sage.
- Simpson, M., & Tuson, J. (2003). *Using observations in small-scale research: A beginner's guide*. SCORE.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.
- Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher, 20*, 8–11.

- Smith, M. L., Edelsky, C., Draper, K., Rottenberg, C. & Cherland, M. (1991). *The role of testing in elementary schools*. Arizona State University.
- Smith, M. L., Noble, A. J., Cabay, M., Heinecke, M., Junker, M.S. & Saffron, Y. (1994). *What happens when the test mandate changes? Results of a multiple case study. CRESST Report 380*. Arizona State University.
- So, Y. (2014). Are teacher perspectives useful? Incorporating EFL teacher feedback in the development of a large-scale international English test. *Language Assessment Quarterly*, 11(3), 283–303.
- Solomon, P. G. (2002). *The assessment bridge: Positive ways to link tests to learning, standards, and curriculum improvement*. Corwin Press.
- Spada, N., & Frohlich, M. (1995). *COLT observation scheme: Communicative orientation of language teaching observation scheme coding conventions & applications*. NCELTR.
- Spada, N., & Lyster, R. (1997). Macroscopic and microscopic views of L2 classrooms. *TESOL Quarterly*, 31(4), 787–795.
- Spolsky, B. (1994). The examination-classroom backwash cycle: Some historical cases. *International Language in Education Conference (ILEC) Hong Kong*, University of Hong Kong, pp. 55–66.
- Spolsky, B. (1989). *Conditions for second language learning: Introduction to a general theory*. Oxford University Press.
- Spradley, J. (1979). *The ethnographic interview*. Holt, Rinehart & Winston.
- Spratt, M. (2005). Washback and the classroom: the implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, 9(1), 529.
- Stake, R. E., & Trumbull, D. J. (1982). Naturalistic generalizations. *Review Journal of Philosophy and Social Science*, 7(1–2), 1–12.
- Stecher, B., Chun, T., & Barron, S. (2004). The effects of assessment-driven reform on the teaching of writing in Washington State. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.). *Washback in language testing: Research contexts and methods* (pp. 3–18). Lawrence Erlbaum Associates.
- Stecher, M. B. (2002). Consequences of large-scale, high-stakes testing on school and classroom practice. In L. Hamilton, M. B. Stecher, & S.P. Klien (Eds.), *Making sense of test-based accountability in education* (pp. 79–100). RAND.
- Stobart, G. (2003) The impact of assessment: intended and unintended consequences, *Assessment in Education*, 10(2), 139–140.

- Sukyadi, D., & Mardiani, R. (2011). The washback effect of the English national examination (ENE) on English teachers' classroom teaching and students' learning. *k@ ta lama*, 13(1), 96–111.
- Sultana, N. (2018). A brief review of washback studies in the south Asian countries. *The Educational Review*, 2(9), 468–474.
- Sultana, N. (2019). *Exploring the alignment of the Secondary School Certificate English Examination with curriculum and classroom instruction: A washback study in Bangladesh* [Unpublished doctoral dissertation]. Queen's University, Canada.
- Tan, M. (2008). Bilingual high-stakes mathematics and science exams in Malaysia: Pedagogical and linguistic issues. *Paper presented at the Language Testing and Research Colloquium*, Hangzhou, China.
- Tan, M., & Turner, C. E. (2015). The impact of communication and collaboration between test developers and teachers on a high-stakes ESL exam: Aligning external assessment and classroom practices. *Language Assessment Quarterly*, 12(1), 29–49.
- 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.
- Tashakkori, A., & Creswell, J. W. (2007). The new era of mixed methods. *Journal of Mixed Methods Research*, 1(1), 3–7.
- Tavares, N., & Hamp-Lyons, L. (2008). ESL teachers learning to assess interactively: The challenges and teacher development issues. *Presented at 30th Annual Language Testing Colloquium*, (LTRC 08). Hangzhou.
- Tayeb, Y. A., Aziz, M. S. A., Ismail, K., & Khan, A. B. M. A. (2014). The washback effect of the General Secondary English Examination (GSEE) on teaching and learning. *GEMA Online Journal of Language Studies*, 14(3).
- Taylor, L. (2005). Washback and impact. *ELT Journal*, 59(2), 154–155.
- Taylor, L. B., & Weir, C.J. (2009). Language testing matters: Investigating the wider social and educational impact of assessment. *Proceedings of the ALTE Cambridge Conference*, April 2008. Cambridge University Press.
- Teddle, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Sage.

- Terry, G., & Hayfield, N. (2021). *Essentials of thematic analysis*. American Psychological Association.
- Thomson, S. B. (2011). Qualitative research: Validity. *JOAAG*, 6(1), 77–82.
- Thorne, S. (2000). Data analysis in qualitative research. *Evidence-based nursing*, 3(3), 68–70.
- Tran, T. D. (2016). Impact of using TOEIC as an exit requirement at a public university in Vietnam. In *4th British Council New Directions in English Language Assessment: Standardised Testing and Proficiency Scales*.
- Tsagari, D. (2006). *Investigating the washback effect of a high-stakes EFL exam in the Greek context: Participants' perceptions, material design and classroom applications* [Unpublished doctoral dissertation]. Lancaster University.
- Tsagari, D. (2009). *The complexity of test washback: An empirical study*. Peter Lang.
- Tsagari, D. (2011). Washback of a high-stakes English exam on teachers' perceptions and practices. *Selected Papers from the 19th ISTAL* (1996), 431–445.
- Turner, C. E. (2001). *Developing an empirically based rating scale for evaluating speaking ability at the secondary 4 and 5 levels*. Ministry of Education, Quebec.
- Turner, C. E. (2009). Examining washback in second language education contexts: A high stakes provincial exam and the teacher factor in classroom practice in Quebec secondary schools. *International Journal on Pedagogies and Learning*, 5(1), 103–123.
- Turner, C. E. (2012). Classroom assessment. In G. Fulcher & F. Davidson (Eds.), *Routledge handbook of language testing* (pp. 65–78). Routledge.
- UNESCO. (2010). *World data on education*. United Nations Educational, Scientific and Cultural Organisation.
http://www.ibe.unesco.org/fileadmin/user_upload/Publications/WDE/2010/pdfversions/Oman.pdf.
- Urmston, A. & Fang, F. (2008). Stakeholder involvement in the language assessment evaluation process: A Hong Kong case study. *Presented at 30th Annual Language Testing Colloquium*, (LTRC 08). Hangzhou.
- van Lier, L. (1988). *The classroom and the language learner: Ethnography and second-language classroom research*. Longman.
- Van Teijlingen, E. & Hundley, V. (2002). Getting your paper to the right journal: A case study of an academic paper. *Journal of Advanced Nursing*, 37(6), 506–511.
- Vernon, P. E. (1956). *The measurement of abilities*. University of London Press.

- 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. (1999). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory* [Unpublished doctoral dissertation]. Lancaster University, 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. (2005). *The impact of high-stakes testing on classroom teaching: A case study using insights from testing and innovation theory*. *Studies in Language Testing* 22. UCLES/Cambridge University Press.
- Wall, D. (2012). Washback. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 79–92). Routledge.
- Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10(1), 41–69.
- Wall, D., & Horak, 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. Report Number: RR-06-18, TOEFL-MS-34. Educational Testing Service.
- Wang, C., Yan, J. & Liu, B. (2014). An empirical study on washback effects of the internet-based college English test Band 4 in China. *English Language Teaching*, 7(6), 26–53.
- Wang, J. (2010). *A study of the role of the “teacher factor” in washback* [Unpublished doctoral dissertation]. McGill University.
- Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, 13(3), 318–333.
- Watanabe, Y. (1997). *The washback effects of the Japanese university entrance examinations of English: Classroom based research* [Unpublished doctoral dissertation]. Lancaster University.
- Watanabe, Y. (2000). Washback effects of the English section of the Japanese university entrance examinations on instruction in pre-college level EFL. *Language Testing Update*, 27, 42–47.
- Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* [pp. 19–36]. Lawrence Erlbaum Associates.

- Weir, C. J. (2005). *Language testing and validation*. Palgrave MacMillan.
- Wesdorp, H. (1982). *Backwash effects of language testing in primary and secondary education*. Stichting Centrum voor onderwijsonderzoek van de Universiteit van Amsterdam.
- Williams, B. R., Woodby, L. L., Bailey, A. F. & Burgio, K. L. (2008). Identifying and responding to ethical and methodological issues in after-death interviews with next-of-kin. *Death Studies*, 32(3), 197–236.
- Wilson, M., & Sapsford, R. (2006). *Data collection and analysis*. Sage.
- Winke, P. (2011). Evaluating the validity of a high-stakes ESL test: Why teachers' perceptions matter. *TESOL Quarterly*, 45(4), 628–660.
- Wisdom, S. (2018). *Teachers' perceptions about the influence of high-stakes testing on students* [Doctor of Education project study]. Walden University, Minneapolis.
- Wiseman, S. (1961). *Examinations and English education*. Manchester University Press.
- Wolcott, H. F. (1994). *Transforming qualitative data: Description, analysis, and interpretation*. Sage.
- Woods, D. (1996). *Teacher cognition in language teaching: Beliefs, decision-making, and classroom practice*. Cambridge University Press.
- Xerri, D., & Briffa, P. V. (2016). Teacher-designed, high-stakes English language testing: Washback and impact. In P. C. L. Ng & E. F. Boucher-Yip (Eds.), *Teacher agency and policy response in English language teaching* (pp. 86–100). Routledge.
- Xie, Q. (2015). Do component weighting and testing method affect time management and approaches to test preparation? A study on the washback mechanism. *System*, 50, 56–68.
- Yamashita, M. Y. (2011). *How does high stakes testing influence teachers' classroom instruction? Institutional pressure and classroom instruction* [Unpublished doctoral dissertation]. University of Pittsburgh.
- Yıldırım, Ö. (2010). Washback effects of a high-stakes university entrance exam: Effects of the English section of the university entrance exam on future English language teachers in Turkey. *Asian EFL Journal*, 2(12).
- Yin, M. (2005). *A progressively focused qualitative study of teacher thinking in English for Academic Purposes (EAP) classroom language assessment* [Unpublished doctoral dissertation]. University of Bristol, UK.
- Yin, R. K. (2006). Mixed methods research: Are the methods genuinely integrated or merely parallel. *Research in Schools*, 13(1), 41–47.

- Young, R., & Lee, S. (1984). *EFL curriculum innovation and teachers' attitudes*.
https://archive.org/details/ERIC_ED274179.
- Yu, C. (2018). *Reliability of self-report data*.
<http://www.creativewisdom.com/teaching/WBI/memory.shtml>.
- Yu, C. (2020). The washback of the new writing tasks in China's National Matriculation English Test. *English Language Teaching*, 13(1), 99–111.
- Zhan, Y. (2009). *Washback and possible selves: Chinese non-English-major undergraduates' English learning experiences* [Unpublished doctoral dissertation]. The University of Hong Kong.
- 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.
- Zhang, H., & Bournot-Trites, M. (2021). The long-term washback effects of the National Matriculation English Test on college English learning in China: Tertiary student perspectives. *Studies in Educational Evaluation*, 68, 1–21.
- Zhang, X. (2021). Stakeholders' test perceptions on test reform. *Studies in Educational Evaluation*, 70, 101064.

Appendix A. Examination Specifications and Sample Items

EXAM SPECIFICATIONS — GRADE 12 (DIPLOMA) — ENGLISH “ELECTIVE”

SEMESTER ONE

3 hours 70 Marks	LISTENING	READING	WRITING
	Multiple Choice AND T/F (Dialogue)	Wh-Qs AND Gap-fill (Interactive Text)	Topic/Title/Question (Opinion)
	Note-taking AND Wh-Qs (Informative Text)	Multiple Choice AND Matching (Narrative Text)	Description of Data (Graph/Table)

SEMESTER TWO

3hours 70 Marks	LISTENING	READING	WRITING
	Multiple Choice AND Note- Taking (interview)	Wh-Qs AND T/F (Evaluative text))	Informative text (report)
	Wh-Qs AND Gap-fill (Narrative text)	Multiple Choice AND Matching (Informative Text)	Task Instructions (Formal Letter/e- mail)

Source: MoE, Student Assessment Document, 2022/2023

Listening 2 (Items 11 – 20)

(10 marks)

You're going to hear a volunteer talking about a charity called 'Houses for Life'.

Task One: Listen, and make notes in the table below. (*not more than TWO WORDS*).

'Houses for Life' Projects	
- the number of projects the presenter has worked on	⁽¹¹⁾ _____
- construct stronger houses after natural	⁽¹²⁾ _____.
- join teams of around	⁽¹³⁾ _____
- to avoid flooding, new homes are constructed above	⁽¹⁴⁾ _____ level.
- volunteers need to be fit and healthy, but no	⁽¹⁵⁾ _____ experience is required.

Task Two: For each item, write a short answer (*not more than THREE WORDS*).

16. Other than constructing houses, what else do volunteers do? List One

_____.

17. How old must you be to register as a volunteer?

_____.

18. Where do the volunteers stay?

_____.

19. Who meets the volunteers once they arrive in the host country?

_____.

20. What do volunteers have to pay for? List One

_____.

Do not write in this space

Reading 2 (continued)

Task two: Match the five Paragraphs on the left with the sub-headings in the box. (There are three extra texts in the box).

Paragraphs	Sub-headings
41. Paragraph 3. _____	A. Major contributions to saving turtles
42. Paragraph 4. _____	B. The cause of his death
43. Paragraph 5. _____	C. His time in the navy
44. Paragraph 6. _____	D. His qualifications and career
45. Paragraph 7. _____	E. His famous books and awards
	F. His early life and love of nature
	G. How he became interested in turtles
	H. His family

Do not write in this space

Do not write in this space

Appendix B. Skills Covered in English Insights 3

Reading

- Reading skills include understanding the organisation of a text, summarising a text, reading for gist, analysing graphs and statistics.
- Students are given a reason to read with the pre-reading question which brings students' own knowledge and experience into the classroom. Suggestions for warm-up activities are given in the Unit notes.
- The texts are about real people and the real world. They give students the chance to see the grammar and vocabulary in context.
-

Writing

- Writing skills include writing a review, various types of essay, a formal email, instructions, narrative texts, a report, comparisons, applications and descriptions of a chart.
- The topics covered in the writing section revise basic text types such as emails and letters and also teach the skills required to write extended text types (description, review, "for and against" essay, descriptive essay and short essay).
- Full models of the target text types are always given and are topic-related to each unit.
- The **Writing mechanics** box focuses on an element of language or punctuation, aiming to improve the quality of students' writing.
- The Writing skill information box highlights a specific writing skill and encourages students to notice these features in the model text before producing their own work.

Listening, speaking and pronunciation

- These skills include talking about television, people, tastes, discussing habits, the future of technology, a problem.
- The listening sections cover a variety of activity types and common exam questions, such as multiple-choice, true and false, and open questions.
- The listening text is thematically related to the reading text, thus providing additional pre-listening support. The graded tasks practise many kinds of listening activity types that allow the students to develop all the skills necessary to manage in English.
- There are two speaking activity types. One is fluency-based, open-ended pair work. The other type gives students more personalised practice of the target language of the lesson.
- The **Practice** tasks move from controlled to personalised exercises that give the students the chance to use the language in a guided and supported activity.
- Full coverage of pronunciation is given for students to master the language.

Grammar

- Grammar points include present tenses, future tenses, past tenses, narrative tenses, passive, conditionals and relative clauses.
- The key grammar is presented in highlighted boxes. You can support this with the full explanations of form and use in the Grammar GPS.
- Grammar practice tasks are extended in the Workbook
- Grammar sections are cross-referenced to the Grammar GPS section at the back of the Student's Book, which gives further examples.

Studying & exam skills

Topics covered include making notes, mind maps, dictionary skills, using visuals and summarising information.

Exercises in English Insights which specifically address P21's Framework for 21st Century Learning are labelled with an icon, signposting students to the activities that go beyond language learning, and practise skills that will support success in all areas.

These sections aim to build students' study and exam skills steadily and consistently throughout the year. They draw attention to the development and practice of exam skills, reinforcing them in the *Test your skills* at the end of each unit.

(MoE, English Language Curriculum Development Section, 2017/2018)

Appendix C. English Elective Teachers at Grade 12 in 2018–2021

Table C1. Number of teachers overall teaching the English Elective at grade 12 according to gender and governorates, 2018–2021

Governorate	No. of schools	Male	Female	Total
Muscat	29	18	28	46
Al Batinah North	24	11	14	25
Al Batinah South	13	4	9	13
Al Dakhiliyah	36	20	18	38
Al Sharqiyah South	10	3	7	10
Al Sharqiyah North	20	7	13	20
Al Dhahirah	34	19	15	34
Dhofar	14	5	10	15
Al Wusta	1	--	1	1
Musandam	4	1	3	4
Al Buraimi	3	1	2	3
Total	188	89	120	209

Source: Directorate General of Planning and Quality Assurance Database, MoE, 2021

بيان بمعلمي مادة مهارات اللغة الانجليزية للصف الثاني عشر

بيان بمعلمي مادة مهارات اللغة الانجليزية للصف الثاني عشر وعدد المدارس المطبقة للمادة حسب المحافظة و الجنس عينة المعلمين في عام 2021				
عدد المدارس	الجملة	اناث	ذكور	المحافظة
29	46	28	18	محافظة مسقط
24	25	14	11	محافظة شمال الباطنة
13	13	9	4	محافظة جنوب الباطنة
36	38	18	20	محافظة الداخلية
10	10	7	3	محافظة جنوب الشرقية
20	20	13	7	محافظة شمال الشرقية
34	34	15	19	محافظة الظاهرة
14	15	10	5	محافظة ظفار
1	1	1	--	محافظة الوسطى
4	4	3	1	محافظة مسندم
3	3	2	1	محافظة البريمي
188	209	120	89	الجملة

دائرة التخطيط التربوي بوزارة التربية والتعليم – المصدر: 2021

Appendix D. Teachers' Questionnaire (Main Study)

Research Project: Examining the Washback from the English Elective Examination on Classroom Teaching in Grade 12 Omani Schools

Dear Teacher,

I am currently carrying out a research project to investigate attitudinal and behavioural changes among teachers concerning the new English Elective Diploma Exam introduced in the school year 2018/2019. Specifically, this study seeks to gather information about how teachers feel about the English Elective Diploma Exam changes and the preparation they are carrying out in the classroom to cope with the changes in this exam. The study also identifies the existence of factors that seem to influence teaching beyond the exam itself. I would like to invite you to take part in this project.

You will be asked to complete an online questionnaire. It contains approximately 20 questions which will ask for your opinions on the English Elective Diploma Exam changes. The questionnaire takes approximately 15–20 minutes to complete.

Participation is optional. If you do decide to take part, you will be emailed a copy of this information sheet for your records. If you change your mind at any point during data collection, you will be able to withdraw your participation without having to provide a reason.

The data collected through the questionnaire will be anonymous and will be treated confidentially and shared on a need-to-know basis only. I will not ask for your name or any other identifying information. The data that you provide will be stored by code number. Any information that identifies you will be stored separately from the data.

Data will be stored in a password protected electronic file. This will only be accessible to the researcher involved in the project. The anonymous data may be used in presentations, online, in research reports, in project summaries or similar. In addition, the anonymous data may be used for further analysis. The data that you provide will not be traced back to identify you as all data will be anonymous at the time it is collected. If information is gathered that raises concerns about your safety or the safety of others, or about other

concerns as perceived by the researcher, the researcher may pass on this information to another person.

If you agree to complete the questionnaire you are free to leave any questions unanswered or to stop completing the questionnaire altogether at any point. Once the questionnaire is submitted online, the data cannot be withdrawn as it is anonymous so there will be no way to identify your data. The data will be kept in a repository indefinitely for future use by me or any other researcher interested in this area of research.

This research has been approved by the Dept of Education, University of York Ethics Committee. If you have any questions or complaints about this research please contact Dr Khaled el Ebyary (khaled.elebyary@york.ac.uk), or Chair of the Ethics Committee (education-research-administrator@york.ac.uk). If you are still dissatisfied, please contact the University's Data Protection Officer at dataprotection@york.ac.uk.

Many thanks for your help with this research.

Please continue if you agree with the above information and proceed by clicking on the "Next" button.

Safa Al Hinai (Researcher), Department of Education, University of York.

Email: siah501@york.ac.uk

Questionnaire

Part One: Please tick the appropriate answer.

- 1) Gender: F M
- 2) Age group: 20–30 31–40 41–50 above 50
- 3) Highest academic qualification: Bachelor's Master's PhD or equivalent
- 4) Number of years you have been teaching English:
 0–2 3–6 7–10 11–15 more than 15
- 5) Number of years you have been teaching the grade 12 English Elective:
 0–2 3–6 7–10 11–14
- 6) Are you currently teaching the grade 12 English Elective:
 Yes No
- If yes, number of hours you teach the grade 12 English Elective per week:
 4 8 12
- 7) Have you been involved in any training relating to the newly introduced elective diploma exam 2018–2019?
 Yes No

If yes, please provide details of the training you undertook that most benefited you (e.g. training content/delivered by/link):

Part Two: Please place a tick mark (✓) in the box next to each item or provide written answers.

A	To what extent do you agree that each of the following could be a reason for the introduction of the elective diploma exam in 2018–2019?	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
1	To meet the demands of higher education					
2	To improve students' communicative ability in English					
3	To cope with the present decline in English standards					
4	To prepare students for working in an English-speaking academic environment					
5	To use a communicative approach to language learning					
6	If you think there are other reasons, please add them here:					

B	To what extent do you agree with the following statements about the changes to the elective diploma exam specifications introduced in 2018–2019 compared to the previous exam?	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
1	More related to the objectives of the elective grade 12 curriculum English Insights 3.					
2	More emphasis on reading skills					
3	More emphasis on grammatical usage					
4	More emphasis on essay writing					
5	More emphasis on listening skills					
6	More use of authentic tasks					
7	Other key changes, please specify here:					

C	To what extent do you agree that each of the following has put pressure on you in your teaching as a result of the introduction of the elective exam 2018–2019?	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
1	Teaching a new syllabus: English Insights 3					
2	Adopting new teaching strategies					
3	Providing more supplementary teaching materials					
4	Preparing more authentic tasks					
5	Arranging more assessment-related activities (e.g. short quizzes, presentations, written work, etc.)					
6	Other opinions:					

D	Which significant changes are you likely to make in your teaching in response to the introduction of the elective diploma exam?	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
1	Teach according to the new test specifications					
2	Adopt new teaching methods					
3	Put more emphasis on communication practices (e.g. group work, forum, role play, etc.)					
4	Put more emphasis on receptive skills (e.g. reading and listening)					
5	Put more emphasis on productive skills (e.g. writing, speaking)					
6	Employ more authentic language tasks					
7	Other opinions:					

E	To what extent do you agree that each of the following is the most challenging aspect of teaching towards the English Insights 3?	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
1	Students' current English level					
2	Class size					
3	The lack of teaching and learning resources					
4	Heavy workload (e.g. too many classes per week, arranging online materials, etc.)					
5	Inadequate lesson time					
6	Volume of the textbooks					
7	Learning environment					
8	Other opinions:					

F	To what extent do you agree that each of the following is the basic function of mock tests in schools?	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree

1	To give feedback to students					
2	To motivate students					
3	To prepare students for the real exam					
4	To identify areas of re-teaching					
5	To get students pay attention in class					
6	To encourage students to study regularly					
7	Other opinions:					

G	To what extent do you agree that the following factors influence how you teach English elective in grade 12?	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
1	Professional training					
2	Teaching experience					
3	Teaching beliefs and attitudes					
4	Textbooks					
5	The need to obtain satisfaction in teaching					
6	Final exam results					
7	Learners' expectations					
8	Parents' expectations					
9	Supervisor expectations					
10	Other opinions:					

Part Three: Please tick (✓) all that apply.

1) What is your current reaction to the Elective Diploma Exam 2018–2019?

- It needs further development
- I endorse the change

- I am sceptical about the change
- It influences my decision about which language skill is more important to be taught

Others, please specify: _____.

2) What methods have you been using to assess students' learning since the introduction of the Elective Diploma Exam 2018–2019?

- Day-to-day observation
- Presentations
- Written work
- Quizzes
- Class tests
- Other, please specify: _____.

3) What do you do to improve students' learning in the classes of Elective Diploma?

- I emphasise the skills which are more likely to be tested on the exam
- I skip over certain activities in the textbook because they are less likely to be tested on the exam
- I do more exam practices
- I put more emphasis on the integration of skills
- Other, please specify: _____.

4) Who made the primary decisions of the new specifications of elective diploma examination 2018–2019?

- Grade 12 English teachers together
- English language technical members from the Ministry of Education (e.g. curriculum developers, exam writers, etc.)
- Members from higher education institutions
- Decision-making directors from the Ministry of Education (e.g. director of examination and test administration)
- Other, please specify: _____.

Part Four: Please answer the following questions (1–2).

1) Does the new elective diploma exam influence the way in which you teach your English class? (If yes, what are these ways, and how are they different?)

2) Do you think the new elective diploma exam assesses your students' English abilities appropriately? (If yes, how? If no, why?)

3) What are the learning activities that are important in teaching English elective diploma classes? (Tick all that apply to you)

- Tell the students the aims of each lesson
- Demonstrate how to do particular language activities
- Do activities similar to mock exams
- Discuss textbook exercises
- Organise group work or discussion
- Organise real-life language activities (e.g. mock interview, sketches, etc.)
- Read newspapers, magazines or books written in English
- Watch movies or news in English
- Practise skimming and scanning skills
- Memorise model samples for writing
- Memorise vocabulary lists
- Practise notetaking
- Practise model samples for speaking tasks
- Practise pronunciation
- Memorise useful expressions for writing text in the final exam
- Study grammar rules
- Other (please specify)

➤ Opportunity for further involvement:

If you would like to take part in a 25–30 minute interview (conducted in Arabic or English, depending on your preference) to discuss the previously mentioned issues in more depth,

please fill in your details below. I will get in touch with you, and we can meet at a time and place convenient to you.

Name: _____ Phone Number: _____

—— End of Questionnaire ——

Thank you very much for your help!

Appendix E. Observation Instruments

Communicative Orientation of Language Teaching (COLT) Observation Scheme

Time	Activities & Episodes	Participant Organization							Content						Content Control			Student Modality					Materials								
		Whole class			Group		Individual		Manage		Lang			Other topics		Teacher/Text	Teacher/Text/Student	Student	Listening	Speaking	Reading	Writing	Other	Type			Source				
		S ↔ S/C	T ↔ S/C	Choral	Different tasks	Same task	Different tasks	Same task	Procedure	Discipline	Form	Function	Discourse	Narrow	Broad									Written	Audio	Visual	L2-NNS	L2-NS	L2-NSA	Student-made	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	

(Source: adapted from COLT, Spada & Frohlich, 1995)

Observation Note-Taking Sheet

Name of school:

Grade:

Teacher:

No. of visit:

Textbook(s) unit/lesson:

No. of students in class:

Start:

End:

Time	Test-taking strategies	Exam-related materials	Teacher	Student	On Board	Comment/Question

Appendix F. Sample Observation Scheme and Field Notes – Teacher A

Communicative Orientation of Language Teaching (COLT) Observation Scheme

Time	Activities & Episodes	Participant Organization							Content						Content Control			Student Modality					Materials								
		Whole class			Group		Individual		Manage		Lang			Other topics		Teacher/Text	Teacher/Text/Student	Student	Listening	Speaking	Reading	Writing	Other	Type			Source				
Start		S ↔ S/C	T ↔ S/C	Choral	Different tasks	Same task	Different tasks	Same task	Procedure	Discipline	Form	Function	Discourse	Narrow	Broad									Written	Audio	Visual	L2-NNS	L2-NS	L2-NSA	Student -made	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	
8:40	A1: Revision for last lesson (5 min)		✓						✓					✓		✓				✓							✓				
8:45	A2: Writing notes about timeline (10 min)		✓					✓	✓	✓			✓				✓	✓	✓		✓			✓		✓					
9:00	A3: Listen & discuss (20 min)		✓					✓	✓						✓				✓	✓						✓	✓				
9:20	A4: Listen & discuss (20 min)		✓					✓	✓						✓				✓	✓					✓		✓				

Observation Note-Taking Sheet

Name of school: School A
Textbook(s) unit/lesson: 5/A

Grade: 12
No. of students in class: 12

Teacher: A
Start: 8:40 am

No. of visit: 1st
End: 9:40 am

Time	Test-taking strategies	Exam-related materials	Teacher	Student	On board	Comment/Question
55 min	Listening text was repeated twice as in the exam Listening text specified the same as the exam listening Same listening type/topic in Ss book which will be similar in terms of language and vocab T did not use extra listening worksheet	Listening tasks included in the Ss book	–	Ss work in group and pairs Good interaction between T and Ss	T wrote the answers of the Ss book	Most tasks in this lesson were listening and some occasions there were speaking Ss feel difficult to hear the recording once Good interaction between Ss themselves and T with Ss

Appendix G. Sample Observation Scheme and Field Notes – Teacher B

Communicative Orientation of Language Teaching (COLT) Observation Scheme

Time	Activities & Episodes	Participant Organization							Content						Content Control			Student Modality					Materials								
		Whole class			Group		Individual		Manage		Lang			Other topics		Teacher/Text	Teacher/Text/Student	Student	Listening	Speaking	Reading	Writing	Other	Type			Source				
		S ↔ S/C	T ↔ S/C	Choral	Different tasks	Same task	Different tasks	Same task	Procedure	Discipline	Form	Function	Discourse	Narrow	Broad									Written	Audio	Visual	L2-NNS	L2-NS	L2-NSA	Student-made	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	
7:45	A1: Introducing formal letter writing <u>structure</u> (5 min)		✓						✓				✓			✓			✓	✓	✓			✓		✓					
7:50	A2: Discussing overall formal letter structure (25 min) A3: Discussing grammar <u>usage</u> (5 min) Overall (30 min)		✓						✓		✓		✓			✓			✓	✓	✓			✓		✓					
8:20	A3: Read & Answer (10 min)		✓					✓			✓		✓	✓			✓		✓		✓			✓		✓					

Observation Note-Taking Sheet

Name of school: School B

Grade: 12

Teacher: B

No. of visit: 2nd

Textbook(s) unit/lesson: 6 & 7

No. of students in class: 16

Start: 7:45 am

End: 8:40 am

Time	Test-taking strategies	Exam-related materials	Teacher	Student	On board	Comment/Question
45 min	Memorize vocabulary and formal letter language and phrases T highlighted sentences & structures of formal letter T asked Ss to memorize verb tenses in each sentence	Formal letter layout	TTT used in all activities	No group or pair work	Key words & phrases Layout of any formal letter T asked & elicited answers in discussion of formal letter	This lesson was theoretical of another practical part T used traditional way of teaching writing This lesson was seen with another class (1 st visit) and the teacher used same process and methods (consistency in teaching methods for different Ss)

Appendix H. Interview Schedule for Teachers (Main Study)

A: Arabic Version

مقابلات المعلمين:
سيقوم الباحث بتحديد وقت المقابلة بحوالي 40 دقيقة، حيث سيبدأ المقابلة بتقديم الشكر للمشاركة والتذكير بأهداف الدراسة وطريقة جمع البيانات وتحليلها، إضافة إلى حق المشارك في الانسحاب من الدراسة:

م	المحور	السؤال
1	التمهيد	<p>بداية لدي بعض الأسئلة حول خبرتك في مجال التدريس:</p> <p>1.1 ما هي مؤهلاتك الأكاديمية؟</p> <p>2.1 كم عدد سنوات خبرتك في التدريس؟</p> <p>3.1 كم عدد سنوات خبرتك في تدريس الصف 12؟</p>
2	آراء ومعتقدات المعلمين حول امتحان اللغة الإنجليزية المطور للمادة الاختيارية لطلبة دبلوم التعليم العام.	<p>1.2 ما مدى الاختلاف أو التشابه بين امتحان اللغة الإنجليزية للمادة الاختيارية المطور عام 2018-2019 عن السابق؟</p> <p>2.2 من هم متخذي القرار الذين قاموا باتخاذ قرار تغيير وتطوير امتحان اللغة الإنجليزية للمادة الاختيارية الذي بدأ تطبيقه في العام الدراسي ٢٠١٨ / ٢٠١٩ ؟</p> <p>- وهل لديك علم بمشاركة بعض معلمي المادة في مناقشات اتخاذ ذلك القرار؟</p> <p>3.2 من الممكن أن يتم وصف التغيير كالتالي: "التغيير قد يكون عميق جداً، بحيث يؤثر على عمق المهارات المكتسبة والمعتقدات ومفاهيم التعليم" وقد يختلف من حيث أهدافه، غايته، والمفهوم الذاتي له" (Fullan & Stiegelbuer, 1991 P. 45) وبناء على هذا التعريف:</p> <p>- من وجهة نظرك، إلى أي مدى يعكس هذا التعريف تغيير امتحان مادة المهارات الجديدة في الصف ١٢ في السلطنة، ولماذا؟</p> <p>- من وجهة نظرك هل تعتقد أن استبدال مواصفات الاختبار المنهج القديم بالمنهج الجديد لمادة المهارات امرًا ضروريًا، ولماذا؟</p> <p>4.2 ما هي الأهداف المرجوة من تطوير امتحان مادة المهارات لطلبة دبلوم التعليم العام؟</p> <p>5.2 الأسئلة القادمة ستركز على رأيك كمعلم في تأثير تطبيق الامتحان المطور لمادة مهارات اللغة الإنجليزية للصف 12 على أساليب التدريس:</p> <p>- هل هناك أي تغيير في خطة التدريس الناتجة ناتج عن تطبيق الامتحان المطور مقارنة بالاختبار القديم؟ فسر ذلك؟</p> <p>- هل أثر تطبيق الامتحان المطور على استخدامك لطرق تدريس مختلفة مقارنة بالامتحان السابق؟</p> <p>(إذا كانت إجابتك نعم، ما هي هذه الأساليب؟ وما مدى اختلافها؟ ومن وجهة نظرك ما أسباب وجود هذا الاختلاف؟</p> <p>(إذا كانت إجابتك بلا، لماذا في اعتقادك أن أساليب التدريس لم تتغير أو تتأثر؟</p> <p>- هل لاحظت وجود أي نوع من أنواع التغيير في حصتك الصفية ناتج عن تطبيق الامتحان المطور.</p> <p>(إن كان كذلك هل من الممكن وصف هذا التغييرات، وتوضيح أسباب حدوثها أثناء تدريسيك/بحسب اعتقادك).</p> <p>6.2 إن كانت إجابتك بنعم لأحد الأسئلة السابقة، فهل تعتقد أن هذا التغيير سببه المباشر هو تطبيق الامتحان المطور؟</p> <p>7.2 كيف يمكن أن تصف آراء ومعتقدات المجموعات الآتية حول الامتحان المطور؟</p> <p>- مدراء المدارس .</p> <p>- المشرفين.</p> <p>- زملائك أو المعلمين الآخرين.</p> <p>- الطلبة.</p> <p>- أولياء الأمور.</p>
3	ممارسات المعلمين داخل الحصة الصفية	<p>1.3 ما هو رأيك حول كتاب المعلم لمادة المهارات المطور للصف الثاني عشر؟ بشكل عام؟</p> <p>- أثناء تحضيرك للحصة الدراسية هل تقوم باتباع ما جاء بالدليل حرفياً، أم تقوم بالتعديل أو التغيير في طرق وأساليب التدريس أو الوسائل التعليمية المحددة في هذا الكتاب؟ إذا كانت إجابتك بنعم؟ ما هي نسبة أو عدد مرات التغيير، أو التعديل؟ وما أسباب قيامك بالتغيير وعدم اتباع ما جاء في الدليل حرفياً من وجهة نظرك؟</p> <p>- كيف تقوم بتدريب الطلبة على أنشطة مهارات الكتابة التي قد يتعرض لها الطلبة في الامتحان؟</p> <p>- كيف تقوم بتدريب الطلبة على أسئلة الاختبار من متعدد التي قد يتعرض لها الطلبة في الامتحان؟</p>

<ul style="list-style-type: none"> - كيف تقوم بتدريب الطلبة على أنشطة مهارة القراءة والتي قد يتعرض لها الطلبة في الامتحان؟- - كيف تقوم بتدريب الطلبة على الاستعداد لانشطة واسئلة مهارة الانصات والتي قد يتعرض لها الطلبة في الامتحان؟ - ما هو الدعم الذي يجب عليك تحقيقه للطلبة المتدني تحصيلهم في امتحان مادة المهارات المطورة في الصف 12؟ 2.3 هل تقوم بتخصيص بعض الحصص الدراسية لإعداد الطلبة للامتحان النهائي؟ إذا كان أجابتك بلا، فالرجاء تفسير ذلك. وإن كانت إجابتك بنعم فالرجاء الإجابة على الأسئلة الآتية: - متى تقوم بتنفيذ حصص الاستعداد للامتحان النهائي؟ - كم عدد المرات التي تقوم فيها بتقديم هذه الحصص خلال الفصل او العام الدراسي؟ - ما هي المصادر التي تعتمد عليها في تحضير حصص لأعداد الطلبة للامتحان في المادة المطورة الاختيارية؟ - كيف تستخدم هذه المصادر داخل الحصة الصفية؟ اذكر مثال على ذلك؟ - هل تعتقد بأن الوقت الذي تخصصه لأعداد الطلبة للامتحان في المادة المطورة الاختيارية مناسب من حيث الفترة الزمنية؟ولماذا؟ 3.3 صف استخدامك لوثيقة مواصفات الامتحان في حصتك الصفية ؟ على سبيل المثال: - هل تستخدمها في كل حصة؟ - ما هي الأساليب/طرق استخدامك لها. 4.3 صف استخدامك لإنشطة التقييم المستمر ، وبشكل عام هل تستخدم هذه الأنشطة في كل حصة دراسية؟ - من يقوم بتصميم هذه الأنشطة؟ - هل يحتوي دليل المعلم على مقترحات أو نماذج لهذه الأنشطة؟ - من وجهة نظرك ما هي وظيفة التقييم المستمر؟ بمعنى آخر ما الهدف من تطبيقه؟ وما مدى علاقة أنشطة التقييم المستمر بالامتحان المطور للصف الثاني عشر؟ 5.3 في رأيك ، ما هو الهدف الأساسي من تطبيق الامتحانات التجريبية (Mock Exams) في المدارس. - من الذي يتخذ قرار تطبيق هذه الامتحانات؟ - هل يعد تطبيق هذه الاختبارات الزاميا في المدارس؟ هل يشترط على الطلبة حضورها؟ - من هي الجهة المسؤولة عن إعداد وتصحيح هذه الاختبارات ؟ - ما هي آلية التعامل مع نتائجها؟ هل يتم مشاركة النتائج مع الطلبة؟ أولياء الأمور؟ أم الوزارة؟ - ما هي إجراءات تطبيق الامتحانات التجريبية في الحصص الصفية؟ - ما هي اعتقادات، اتجاهات، آراء (الطلبة، أولياء الأمور، مديري المدارس) حول هذه الامتحانات. 6.3 ما مدى الاختلاف والتشابه في مهام عملك منذ البدء في تطبيق الامتحان المطور عام 2018-2019 عن ما كان عليه سابقا ؟ 7.3 ما هي التحديات التي تواجهها منذ البدء في تطبيق مواصفات الاحتمار النهائي المطور لمادة المهارات الصف 12 بجانب مهامك المكلف بها؟ 		
<ul style="list-style-type: none"> - هل يوجد لديك أية نقاط أو مواضيع (أو مقترحات) تود المشاركة بها في الحديث عن تطوير امتحان مادة المهارات الصف 12 أو التعليق عن ما تم الحديث عنه سابقا؟ في نهاية اللقاء أشكر المعلم على مشاركته في التعبير عن آرائه مع الباحثة . 	النهائية	4

B: English Version

Teacher Interview Schedule

Name of participant: _____

Date & venue of interview: _____

Time started: _____ Time ended: _____

Materials: Two past elective diploma exam papers, one in the first-semester version and one in the second-semester version, are at hand for reference.

Introductory statement: Thank you for agreeing to participate in the study. The purpose of this project is to investigate attitudinal and behavioural changes among teachers concerning the new English Elective Diploma Exam introduced in the school year 2018/2019. Specifically, this study seeks to gather information about how teachers feel about the English Elective Diploma Exam changes and the preparation they are carrying out in the classroom to cope with the changes in this exam. The study also identifies the existence of factors that seem to influence teaching beyond the exam itself.

The data you provide (notes or audio recordings of the interview) will be confidential. A pseudonym will be used, and data will be anonymised and stored by a code number, making personal information untraceable.

SECTION 1: Introductory questions

1. What are your teaching qualifications?
2. How long have you been teaching?
3. How long have you been teaching grade 12?

SECTION 2: Teachers' perceptions of the new elective exam

1. How is the current diploma elective exam introduced in 2018–2019 similar to or different from the previous one?
2. Who participated in the decision to change the elective exam?
 - a) Are you aware of teacher involvement in the decision-making process?
3. Change has sometimes been described like this: “Change can be very deep, striking at the core of learned skills and beliefs and conceptions of education, and creating doubts about purposes, sense of competence, and self-concept.” (Fullan & Stiegelbauer, 1991, p. 45). To what extent do you think this describes the introduction of the new elective subject in grade 12? Why?
4. Do you think the elective exam change was necessary? Why?

5. What are the primary goals to be achieved by introducing a new exam?
6. Now I will ask you whether teachers feel their teaching has changed since the implementation of the new elective diploma exam.
 - a) Do you think the plan of your classroom teaching has changed? Explain.
 - b) Has the implementation of the new exam led you to use different teaching methods? If yes, what are these methods, and how are they different? Why do you think this change happened? If no, why do you think your teaching methods did not change?
 - c) Have you observed any other changes in your classroom that you feel are a result of the implementation of the new exam? If so, could you please describe what they are and explain why they have occurred?
 - d) If you answered yes to any of the above, do you think the changes are the direct result of the implementation of the new exam?
7. How do you perceive the attitudes of each of the following groups towards the elective examination changes?
 - a) School principals
 - b) Supervisors
 - c) Your peers or other teachers
 - d) Students
 - e) Parents

SECTION 3: Teachers' practices in the classroom

8. What do you think about the teachers' guidebook? Do you generally modify the teaching material, methods and practices outlined in that guide? If yes, how often does this happen? Why do you modify them?
 - a) How do you train your students for the writing tasks?
 - b) How do you train your students for the MCQ test items?
 - c) How do you train your students for the reading test items?
 - d) How do you train your students for the listening test items?
 - e) What sort of support are you expected to provide to students who achieve low marks on the elective exam?
9. Do you provide students with exam preparation lessons? If no, please explain why. If yes, please answer the following questions.
 - a) When do teachers use exam preparation for the elective exam?

- b) How often do you do exam preparation in a semester?
 - c) What resources do you use for exam preparation?
 - d) How do you use these resources in the classroom? Can you give me an example?
 - e) Do you think that the time devoted to exam preparation should be minimised or maximised? Why is that?
10. Could you please describe your utilisation of the new exam specifications in your classroom? For example,
- a) How often do you use it?
 - b) In what ways do you use it?
11. Do you generally use continuous assessment tasks? If yes, in what way?
- a) Who designs these tasks?
 - b) Does the teachers' guide include any suggestions or examples of continuous assessment tasks?
 - c) What do you think is the function of your continuous assessment? (Why do you do it?) How is this related to the new test?
12. What do you think are the basic functions of mock exams implemented in schools?
- a) Who made the decision to administer mock exams?
 - b) Is it a requirement that all schools carry out mock exams? Do students have to attend the mock exam?
 - c) Who designs or marks the mock exam?
 - d) How are the results handled? Are they shared with students? Parents? The Ministry?
 - e) What is the procedure for the mock exams' implementation in the classroom?
 - f) How do students/parents/school principals feel about having to do a mock?
13. How different or similar are your responsibilities since the change in test format?
14. What challenges to fulfilling your teaching responsibilities have you faced with the new curriculum?
15. Is there anything you want to share about the new elective exam that has not been covered in previous questions?

Closing remarks: Thank you so much for sharing your views with me. It's been great talking to you.

Appendix I. MoE Decision-Maker Interview Schedule (Main Study)

A: Arabic version

مقابلات متخذي القرار في وزارة التربية والتعليم
(الأعضاء الفنيين في تطوير المناهج، والأعضاء الفنيين واضعي الاختبارات)
سيقوم الباحث بتحديد وقت المقابلة بحوالي 40 دقيقة، حيث سيبدأ المقابلة بتقديم الشكر للمشاركة والتذكير بأهداف الدراسة وطريقة جمع البيانات وتحليلها، إضافة إلى حق المشارك في الانسحاب من الدراسة:

م	المحور	السؤال
1	التمهيد/الافتتاحية	سأبدأ بالسؤال عن وظيفتك الحالية: 1.1 ما هي أهم مهام ومسؤوليات وظيفتك الحالية؟ 2.1 كم عدد سنوات خبرتك في وظيفتك الحالية؟
2	آراء ومعتقدات متخذي القرار حول الامتحان المطور لمادة مهارات اللغة الإنجليزية للصف 12.	1.2 ما الجهة التي قامت باتخاذ القرارات الخاصة بوضع واعتماد مواصفات تغيير الورقة الامتحانية للمنهج المطور لمادة اللغة الإنجليزية الاختيارية للعام 2018-2019؟ (الرجاء توضيح إجراءات التغيير والتطوير التي تم اتخاذها؟) 2.2 ما هي الأهداف الأساسية المرجوه من تطبيق الامتحان المطور للمادة؟ (هل هي معتمدة؟) 3.2 صف آراء ومعتقدات الفئات الآتية حول تغيير امتحان المهارات للصف 12؟ - مدراء المدارس. - المعلمين. - الطلبة. - أولياء الأمور. 4.2 وضح ما هي الإجراءات التنفيذية التي إتخذتها الوزارة لمشاركة المعلمين حول تطوير إختبار مادة المهارات للصف 12؟ (إذا كانت إجابتك بنعم، وضح ذلك، وإذا كانت إجابتك بلا، فسر إجابتك) 5.3 براك، هل يؤثر تغيير مواصفات الورقة الامتحانية على تحضير واعداد المعلم للحصة الصفية؟ (إذا كانت إجابتك بنعم، وضح ذلك، وإذا كانت إجابتك بلا، فسر إجابتك) 6.3 ما هي الاهداف الأساسية للاختبارات التجريبية (Moke Exam) لمادة اللغة الإنجليزية للمادة الاختيارية للصف الثاني عشر والمطبقة في المدارس؟ - من هي الجهة المسؤولة التي افرت تطبيق هذه الاختبارات؟ - هل يعتبر تطبيق الاختبارات التجريبية في المدارس مطلب إلزامي لجميع الصفوف التي تدرس المادة الاختيارية؟ - من يقوم بإعداد وتصحيح هذه الاختبارات؟ - ما هي الإجراءات المتبعة في تطبيق هذه الاختبارات في الحصة الصفية؟ 7.3 هل يوجد اختلاف بين إستراتيجيات التقويم المستمر المتبعة في مادة المهارات المطورة للصف 12 مقارنة بالاستراتيجيات المطبقة في المنهج السابق؟ (إذا كانت إجابتك بنعم، وضح ذلك، وإذا كانت إجابتك بلا، فسر إجابتك). 7.3 هل قامت وزارة التربية والتعليم بتدريب وإعداد المعلمين على مواصفات اختبار المهارات المطور؟ (إذا كانت إجابتك بنعم فسر ذلك؟ وإن كانت إجابتك بلا فذكر السبب؟) 8.3 ما هو نوع الدعم المتوقع من المعلمين تقديمه للطلبة ذوي التحصيل الدراسي المتدني في إمتحان مادة المهارات المطورة في الصف 12؟

<p>9.3 ما مدى اختلاف أو تشابه أدوار ومهام المعلمين نتيجة لتطبيق المواصفات الجديدة للورقة الامتحانية للمنهج الجديد مقارنة بالمنهج السابق؟</p> <p>10.3 ما مدى الادراك بالتحديات التي يواجهها المعلمين في تدريس مادة المهارات المطورة للصف 12 داخل الغرفة الصفية من وجهة نظرك كمتخذ قرار؟</p> <p>11.3 هل لديكم مؤشرات عن عدد الطلبة الذين يختارون المادة المطورة منذ العام الدراسي ٢٠١٨ / ٢٠١٩ مقارنة باخر ٣ أعوام من تطبيق المنهج السابق؟ - هل هناك ارتفاع في عدد الطلبة ام العكس؟ - ما أسباب ذلك من وجهة نظرك؟</p> <p>12.3 هل تعتقد أن نتائج الطلبة في مادة المهارات المطورة للصف 12 خلال السنتين الماضيتين من التطبيق أفضل من نتائج مادة المهارات السابقة في اخر عامين؟ وضح ذلك؟</p>		
<p>- هل يوجد لديك أية مواضيع أو نقاط تود إضافتها حول اختبار مادة المهارات المطورة، والتي لم يتم التطرق إليها فيما سبق؟</p> <p>في نهاية اللقاء أشكر المشاركين على مشاركتهم في هذا اللقاء مع الباحثة .</p>	النهاية	3

B: English Version

MoE (curriculum developers and exam writers)

Name of participant: _____ Email: _____

Date & venue of interview: _____ Job title: _____

Time started: _____ Time ended: _____

Materials: Two past elective diploma exam papers, one in the first-semester version and one in the second-semester version, are at hand for reference.

Introductory statement: Thank you for agreeing to participate in the study. The purpose of this project is to investigate attitudinal and behavioural changes among teachers concerning the new English Elective Diploma Exam introduced in the school year 2018/2019. Specifically, this study seeks to gather information about how teachers feel about the English Elective Diploma Exam changes and the preparation they are carrying out in the classroom to cope with the changes in this exam. The study also identifies the existence of factors that seem to influence teaching beyond the exam itself.

The data that you provide (notes or audio recordings of the interview) will be confidential. A pseudonym will be used, and data will be anonymised and stored by a code number, making personal information untraceable.

Section 1: Participant's Background

1. What are your roles or responsibilities at work?
2. How long have you been in your current role?

Section 2: Perceptions in Relation to the New Elective Exam

1. Who made the major decisions regarding the specifications of the elective diploma examination 2018–2019?

Please explain the decision-making process.

2. What are the primary goals of introducing a new exam?
3. How would you describe the attitudes of each of the following groups towards the elective examination change?
 - a) School principals
 - b) Teachers
 - c) Students
 - d) Parents

4. Did the Ministry of Education involve teachers in decisions about the new test? If so, please describe the procedures. If no, why do you think that teachers were not involved in this change?
5. How do you think the examination changes affect teacher preparation for the classroom?
6. What do you think are the basic functions of the mock exams implemented in schools?
 - a) Who made the decision to administer mock exams?
 - b) Is it a requirement that all schools administer mock exams?
 - c) Who designs and marks the mock exams?
 - d) What is the procedure for implementing the mock exams in the classroom?
7. Is there a difference between the assessment strategies used in the new elective curriculum and the ones used in the previous curriculum? (If yes, please describe. If no, why not?)
8. Did the Ministry of Education prepare teachers for the specifications of the new test? If yes, how? If no, why not?
9. What sort of support are teachers expected to provide to students who score low marks on the elective exam?
10. How different or similar are teacher responsibilities due to the new exam format compared to the old format?
11. Are you aware of any specific challenges teachers may encounter when teaching the new elective subject in the classroom?
12. Has the number of students who are selecting Elective English increased with the new examination system? Please explain.
13. Do you think students' results in Elective English of the last two years of its implementation are better than the results of the old exam? Explain?
14. Is there anything else you want to mention about the new Elective exam that has not been covered in previous questions?

Closing remarks: Thank you so much for sharing your views with me. It has been great talking to you.

Appendix J. Sample Observation Transcript – Teacher A

Information

Lesson One: 1 st visit				
Day & Date	Lesson time	Lesson duration	No. of students	Recording folder
Tuesday (6/4/2021)	Second (8:40)	60 min	12	LS - P4(D) recorders + phone
Topic of the lesson	Coming of Age: speaking and listening lesson (Unit 5/A)	Resources	Big TV screen + white board + worksheets	

No	Time	Activity
1	8:40-8:45	<p>Speaking</p> <p>T: What did we have last lesson/what did we do? How did we start the previous lesson? What was the previous lesson about and what did you learn?</p> <p>Ss: (described picture of Burj Khalifa; talk about Shakespeare poem; discussed general Qs about events, quotations and seventh stages.</p>
2	8:45-9:00	<p>Speaking & listening & Writing</p> <p>T: We are going to talk about your lifetime line (task 3/p. 70): I will give you a timeline showing three periods of your life: 0-6; 7-15; 16 until now. For example: you can write about: something about your old school, parents, old memories, etc. Each student is going to write about herself. You must write notes only not sentences in 3 minutes (students work individually). Teachers counted from 1-10 to let students stop writing.</p> <p>All students in class read their notes aloud.</p> <p>All students read their notes aloud (about their lifetime main events)</p> <p>T comments on some of Ss answers by asking why and where, or explain what happened.</p>
3	9:00-9:20	<p>Speaking & Reading</p> <p>T: Do you celebrate when you reach 15 years old? Is it a ritual in your family? Do you have a special party when you reach special age?</p> <p>Ss: When I get to 18, I celebrate with my family and friends.</p> <p>T: Today we are going to listen to two different societies talking about coming of age. I want you to open your book page 70 /activity 5.1. Who is going to read the activity?</p> <p>A student read the rubric.</p> <p>T: So, what we are going to do?</p> <p>Ss listened and completed about the two speakers.</p> <p>T: What are the information you will listen about?</p> <p>Ss: Celebrations of two different societies.</p> <p>T: Write the name of the festivals and complete the notes in the table. Are you ready to listen? Did you read the notes? Please read first the notes then I will play the task.</p> <p>Ss listened to two interviews and completed the missing notes (Ss asked teacher to play the listening twice as in the final exam). You will listen to the first festival and then the second.</p> <p>Ss listened and completed the notes individually.</p> <p>T: Did you get the answers? Would you like to listen again? (T repeated the listening task again).</p> <p>T elicited answers about the listening tasks of the first and second speakers: (what each festival is about, what did the people do there, purpose of each festival, how old are the speakers, are they a good age, what characters they present, what about the food, what they learned about, how girls celebrate, what they choose, girls dance with whom, what the name of the dance, what do you think of this event, which festival was more interesting?).</p>

		T played a short video about these celebrations (no discussion was done after this).
4	9:20-9:40	<p>Speaking & Reading</p> <p>T: Let's listen to speakers talking about becoming adults. So, now go to activity 8 page 71. Who is going to read the rubric?</p> <p>A student read it aloud.</p> <p>T: So, what is it about?</p> <p>Ss: When did the speakers feel they had entered into adulthood?</p> <p>Ss listened silently.</p> <p>Ts: What the speakers had when they were adult?</p> <p>Ss: Go university and left their hometown (speaker 1), celebrating her high school graduation (speaker 2), working a mechanic after high school and set up a business (speaker 3).</p> <p>T: Read and answer activity 9 /page 71: complete the sentences with the correct word in the box and then listen and check your answers. Listen again the same listening task about the three speakers and complete the notes in your textbook. Read your answers and check with your friends.</p> <p>After listening, T checked students' answers orally.</p> <p>T asked Ss to prepare for the next lesson: When did you feel you become an adult?</p> <p>End of lesson</p>

Appendix K. Sample Observation Transcript – Teacher B

Information

Lesson One: 1 st visit				
Day & Date	Lesson time	Lesson duration	No. of students	Recording folder
Sunday (4/4/2021)	Third (10:30)	60 min	12	LS - P4(D) recorders + phone
Topic of the lesson	Job application: writing lesson on formal letter (Units 6 & 7)		Resources	Big TV screen + white board + worksheets

No.	Time	Activities/episodes
1	10:05-10:25	<p>Speaking</p> <p>T: From the lesson title written on the board: what is the lesson about? and what are we going to do?</p> <p>Ss: Write a job application letter.</p> <p>T: Why do we use this letter? Where can we find this type of letter?</p> <p>Ss: While applying for a job.</p> <p>T: Yes, to prepare you for next step and your future. Today we are going to have this type of letter and it is preparation for future. It is a formal way of writing and this form of writing is included in Unit 6 & 7 of the coursebook ...</p> <p>No need talking please. We will focus on this type today.</p>
2	10:25-10:45	<p>Reading & Speaking</p> <p>T distributed a reading passage worksheet.</p> <p>T: The piece of reading... Do you see it? Read the passage silently and then analyse it step by step. Spend about 2 min. (T wrote key words on the board while Ss read). Have you done girls? So, before we start remind me: Any letter we learned how many points/elements it may consist?</p> <p>Ss: 3</p> <p>T: Who said this thing to you!? No, 5 points. Let's go one by one.</p> <p>T: What is the first point/element about?</p> <p>Ss: Opening</p> <p>T asked for part two</p> <p>Ss: Greeting</p> <p>T: What do we call greetings?</p> <p>Ss: Introduction</p> <p>T: We will go then to third part; which we call it?</p> <p>Ss: The body parts.</p> <p>T: That the body paragraphs may consist of 1, 2, or 3 parts as it depends.</p> <p>T: What is number 4: Since we introduction, we have ...</p> <p>Ss: Conclusion.</p> <p>T: Since we have the opening; there is the last part what is it?</p> <p>Ss: Closing.</p> <p>T: Good... you remember these points!</p> <p>T: In any letter formal or informal: it is supposed to consist of 5 parts: opening, introduction, body parts, conclusion & closing</p> <p>T: Go to your reading passage (worksheets)</p> <p>T asked them how many points do we have?</p> <p>Ss: 5 parts.</p>
3	10:45-11:05	<p>Reading & Speaking</p> <p>T: Look at the opening and tell me what it is in front of you?</p> <p>Ss: The sentence in the opening starts with dear.</p> <p>Ts: Can I start with Hello or Hi?</p> <p>Ss: No.</p> <p>T: It is supposed to be always dear. So, this the first part.</p> <p>T: Give me the second part:</p> <p>Ss: Introduction.</p> <p>T: Now, focus on the introduction: what do we have in the intro?</p> <p>Ss: The purpose of writing the letter and where they found the advert.</p>

T: Good, the general aim of the letter. Later we go further on this.
T: Then, we will move now to the main body: how many bodies here?
Ss: Three.
T: Can you guess what are they about? Or give me the main idea of the first one?
Ss: The first one is about introduction about the candidate himself; second about the experience, third about skills and qualities.
T: This is generally and then we will go deeply later on.
T: We are done from the body paragraph, we have then conclusion: what is about, is there anything special? What have you notice?
Ss: Closing remarks.
T: The last part is a closing.

4 11:05-11:10

Reading & Writing

T: Reads from the TV screen: what the outcomes for the lesson: identify letter layouts/parts; language used in each part; and diversity of phrases for letter writing.
T: The thing that is in front of you is what? Which type of job the letter in your book is about?
Ss: Tour guide.
T: Read the intro part loud?
A student read aloud.
T: What is about?
T: What the layout of the letter and then what each part consists. Let's move part by part now ... Let's go to the introduction.
Ss read the intro.
T: What do you have in the part of opening?
Ss: Dear.
T: We cannot use "hi" or "hello". The only formal and associated word here is "Dear". After that what have you noticed? What do we have after dear?
Ss: Mrs.
T: What else do you notice?
Ss: A comma after the name.
T: What you notice from the name?
Ss: There is the first letter in the name is capital.
Ss: He is not using his first name but his surname.
T: Whenever you write a formal letter entitled the name of the person Mr, Ms, Mrs, etc. and then we have to put the name. Ok ... after that, what is after the name?
Ss: Comma.
T: Ss what if I the letter doesn't mention a name, to who I am going to write, which words can I use (can you remember we took this last year):
Ss: Madam or sir.
T: Yes, if don't know to whom I am sending, I will write Sir/or Madam. And then comma after.
T: Any question before I leave the opening part?
Ss: No.
T: What have you notice in the intro part?
Ss: Position she wants and when he saw the advert.
T: The first things we need to mention is the reason, purpose of writing: why are you writing this letter & when have you seen it, & where you saw the letter. So, the main information are: Why, when and/or where you seen or heard this letter. Now, focus on the language use here: what the type of the language?
Ss: Present.
T: Since he is applying, he is using present. What about the second part "which was advertised"; what type of past tense: is it past simple?
Ss: (did not recognize answer): most answered past continuous.
T: Since we have "ing" it will be what: continuous. So, where is the "ing" here! "The window was broken": after was what do we have:
Ss: Past participle.
T: So when we have verb to be and past participle: what is it:
Ss: Past perfect or past simple.

T: It is a passive voice. Have you heard that before? When you do not want to mention who is doing the action ... you use passive voice. Why we use passive here because it is a formal letter and most formal language use formal voice or passive. So, can I see: I am writing a letter which my friend told me about?

Ss: No, because it is not formal.

T: You got it? Shall we move on?

Ss: Yes.

T: How many paragraphs we are having?

Ss: Three.

A student read the paragraph aloud.

T: This part is about?

Ss: Personal information qualification.

T: A degree candidate have.

A student read the second and third body paragraphs aloud.

T: What they about (personal profile, experience & skills, & personal qualities). Let's go to each paragraph: what the writer wrote in each sentence (1st para: current information about the writer (age, status, major, specific specialization which attach to the job that I am applying for). I will give you example about my profession "teaching English" and I want to "imitate" the same writing style using my profession: how can I manage it or say? How can I write about myself. So, if you are a students in a university and you are applying for a job? I can you write it down and say it to me? You can just imitate the writer words or way of writing? Have you done?

A student answered: "I am a twenty-year old student and I am studying at Sultan Qaboos University. I want to become a teacher at the college of Education"; "I study English art at Sultan Qaboos University".

T: So in this part you should write about current situation, study, graduate where from.

T: What do we have at the end related to the writer language?

Ss: "qualities you have beside your major".

T: So he is well qualified in language. This is part of the qualification a person have. So, current situation and personal information we have to cover. What type of language we are using here:

Ss: Present.

T: Why is it present?

Ss: Because he talks about something current or at present.

T: So, we use present language. If the writer graduate next summer: I will use present continuous or future. Any question so far?

Ss: No.

T: The second body paragraph: we said before it is about?

Ss: Past experience.

T: Read the paragraph silently for few seconds.

Ss read the text silently.

T: He starting his paragraph by telling what?

Ss: Where did he work and when.

T: What about the second sentence?

Ss: What did he do while he was working.

T: The third sentence?

Ss: What did he learn.

T: So we have to mention three Qs: What did you work? Where and when. So, what the next sentence:

Ss: What did you do.

T: So: what were your duties at that job? Which type of work did you do? What did you do? What did you gain from this experience. (T wrote these Qs on the board). You have to cover three main questions in this paragraph: what the job, where, what your duties and what have you gain. What type of language in this paragraph:

Ss: Past simples

T: We can add to it what? Look at the last sentence?

Ss: Passive voice.

T: Please read loud the last sentence?

Ss: Present perfect.

T: When do I use the perfect: I am not teaching grammar but recall it for you. (No answer) ... it happened in the past and still the result happening. Then explain why the writer used this tense in the last sentence in this text. (T wrote the present perfect tense on the board). Any question before we move?
Ss read the paragraph silently.

T: What the writer wrote here? What have you noticed here?

Ss: He wrote about his skills and what is special about him.

T: So, why he is a good candidate for this job, the writer appraising himself or why he thinks he is good for this job. Any question.

A student read from the TV screen aloud a summary of what information/language is in each body para.

T: Now go for the conclusion remarks. How many sentences?

Ss: One sentence.

T: What you notice? What the writer do?

T: The writer use closing remarks? How can I close a letter?

Ss: "I enclose".

T: So, we use this remark to end our email or letter. Now, read together loud.

After, can I replace "at any time convenient" with "as soon as possible".

Ss: no ;

T: because it is not formal. It should be not personal. for formal letter we prefer to use "at any time convenient to you" or they can use as Ss recall from previous lesson /year "I look forward to hear from you at any time convenient to you"

T: Any question?

Ss: No.

T: So, the closing remarks are your sincerely, yours faithfully, etc.); and then after put comma and then signature.

T: Tips for better formal language: what do I mean by formal language? What is it?

Ss: Letter is official. When we talk about job application, complain letter, we don't have abbreviation or short forms like don't, we use passive voice.

T: Ss to read from TV screen: Language use, no abbreviation, passive voice, formal tenses (passive voice), punctuation, conjunction or linking words (Ss give examples: in addition, also, moreover, however, etc.); also we should focus on the variation of tenses.

Lesson ending

T: This is the theoretical part or what do we have to know for next lesson as next lesson will be writing application letter for a job. You will be following the reading text layout,

(Note: T wrote what each paragraph in the reading text include).

T: Thank you for today. I do not know what's wrong with you today.

Ss: maybe because its Sunday teacher!

Appendix L. Coding Scheme for Teacher Interviews

1. Teachers' reactions to and their perceptions of the new elective diploma exam
 - 1.1 Teachers' reactions to the new elective diploma exam
 - 1.1.1 Negative attitudes
 - Complexity of the new elective exam
 - Content of the textbooks
 - Teachers' lack of involvement
 - 1.2 Positive attitudes
 - Awareness of the exam
 - Exam documentation
 - 1.3 Perceptions of the need for change
 - To meet the demands of higher education for the examination
 - To improve students' communicative ability in English
 - To improve teaching practices
 - To meet the demands of the new textbook
 - 1.4 Perceptions of the new test format
 - Content of the exam components
 - Question types and structure
 - 1.5 Perceptions of the pressure the new exam places on teaching
 - Difficulty of teaching reading and writing skills
 - Insufficient teaching time
 - 1.6 Perceptions of challenges in teaching
 - Level of students' proficiency
 - Lack of support and training opportunities for teachers
 - 1.7 Perceptions of the quality of the new elective exam
 - 1.7.1 Positive
 - Quality of the new exam design
 - Challenges students
 - Relevance of the textbook content to the new exam
 - Positive effect of the new exam on teaching
 - 1.8 Perceptions of the introduction of the new exam
 - 1.8.1 Negative
 - Lack of communication with MoE stakeholders

- Lack of awareness of the intended washback
2. Teachers' perceptions of the exam concerning classroom teaching behaviours
 - 2.1 Changes teachers would like to make in their teaching
 - Teaching according to the new elective diploma specification
 - Teaching according to the outcomes expected after grade 12
 - 2.2 Changes they made to improve students' learning
 - Emphasised the skills which are more likely to be tested in the exam
 - Employed new ideas and teaching methods
 - 2.3 Washback on teaching methods
 - Attention to the new exam specification in their teaching
 - Attitudes towards their teaching plan
 - Methods they used for teaching
 - 2.4 Teaching activities
 - Teaching reading
 - Teaching writing
 - Teaching listening
 3. Perceptions of the exam concerning assessment and evaluation
 - 3.1 Exam preparation lessons
 - Function
 - Materials
 - Time
 - 3.2 Continuous assessment methods
 - Emphasis on class tests
 - Continuous assessment practices in classroom
 - 3.3 Factors other than the exam that influence teaching
 - Students' fear
 - The role of the supervisors
 - The role of the principal

Appendix M. Coding Scheme for MoE Decision-Makers' Interviews

1. Enhance students' language skills
 - 1.1 The rationale for the change
 - 1.1.1 Inadequate assessment
 - 1.1.2 Low IELTS scores
 - 1.1.3 The negative impact on teaching
 - 1.2 Development of general and academic language skills
 - 1.2.1 Testing different general skills
 - 1.2.2 Preparing students for academic skills
2. Provide continuous assessment for learning
 - 2.1 Purpose of CA
 - 2.2 CA tools
 - 2.3 CA & exam assessment components
 - 2.4 Writing skills assessment in the classroom
 - 2.5 Mock exam period
 - 2.4.1 Uses of mock exams
 - 2.4.2 Content and sources
 - 2.4.3 Responsibility of the mock exams
3. Improve student-centredness and teachers' practices
 - 3.1 Emphasis on a student-centered approach
 - 3.1.1 How a student-centered approach is employed in class
 - 3.1.2 Purpose of the student-centered approach in this new change
 - 3.2 What teachers teach
 - 3.2.1 Focus on exam content
 - 3.2.2 Focus on textbook content
 - 3.2.3 Wide range of topics
 - 3.3 How teachers teach – positive washback
 - 3.3.1 “Active” or “new” ways of teaching
 - 3.3.2 Teachers as facilitators rather than dominators
 - 3.3.3 More communicative language tasks
 - 3.3.4 Professional learning communities
 - 3.3.5 Use of technology
 - 3.3.6 Different teaching strategies

Appendix N. University of York Ethics Approval

Education PhD Administrator

4 Nov 2020,

11:25

to xx

Hello Safa,

Your ethics application has been approved providing you make some minor changes, as follows:

"I would like to see two minor changes to enhance the interviewees' consent forms: the time period for anonymisation needs clarification (e.g. *when* will this happen after the interview). This date and an email address should then be added to the part of the form which says that interviewees may request and comment on a written copy of the interview. I think it is necessary to explain how this written record will be obtained (e.g. by emailing Safa) and if there is a specific period of time for this (i.e. 1-2 weeks, after which time the transcript will be anonymised and such requests will not possible)."

Please make these amendments before collecting any data. If you need further clarification, contact your 2nd reviewer xx.

I have attached the signed ethics form, and also saved a copy to your student file should you need it in the future.

Best wishes,

xx

Research Degrees Administrator
Department of Education | University of York

GENERAL RISK ASSESSMENT FORM

Section 1: Assessment Overview

Assessment Reference Number: COVID-19 Secure	Version Control	Feb 2021
---	-----------------	----------

Name of Assessor	Safa Al-Hinai	
Description of Area/ Procedure/Task being assessed	<p>The researcher will conduct interviews in Oman with 5 teachers and 4 Ministry of Education (MoE) decision-makers of the English language subject. For teachers' participants, one-on-one interviews will be conducted at the interviewees' school sites, and the MoE decision-makers will be interviewed in the Ministry site. Moreover, the researcher will attend four lessons with each of the two teachers in a traditional face to face classroom teaching. These classroom observations will be distributed over two weeks. The researcher will use a non-participant observer approach to observe closely the natural occurrence of English language teaching of the new curriculum in the classroom, to record any relevant information from the lesson, to minimize changes in the subtle behaviours of both teachers and student that may occur due to the researcher's presence, and so that participants could freely share their thoughts and practices without considering the researcher as a distraction or threat. As COVID-19 restrictions in Oman is currently under low level of strict measures and expected to be less strict by the time of conducting the interviews and observations, the targeted grade level of the research sample, grade 12 students, continue to attend face to face study at schools. In doing so, the MoE and the Ministry of Health (MoH) have asked schools across the country to follow a health and safety protocol/or guidance. Based on this guide, the MoE and MoH officials are organizing regular school site visits to monitor the implementation of these health measures in schools; as well as a doctor from the MoH was assigned to act as a school community member to check the health, safety and wellbeing of the school staff.</p>	
Location	Oman	

Section 2: Persons Affected

Who might be affected by this work? (delete as applicable)	Teachers, students and the researcher	Are any vulnerable groups affected? (delete as applicable)	No	How many people are affected? (delete as applicable)	<ul style="list-style-type: none"> - One to one interview with 5 teachers, 4 MoE officials. - Observations in Low-density schools with 16 students per class and one teacher, per day and they will study three in-school hours a day.
--	---------------------------------------	--	----	--	--

Section 3: Review

Date for Next Review of this Document	Date Document Reviewed	Reviewed by (print name)	Signature
1 st April 2021		Safa	Safa

Section 4: Risk Assessment

Risk Matrix

Hazard Severity Score		Likelihood		Probability Severity	1	2	3
Negligible Injury or Damage	1	Unlikely	1	1	LOW	MEDIUM	MEDIUM
Minor Injury or Damage	2	May Happen	2	2	MEDIUM	MEDIUM	HIGH
Major Injury or Death	3	Almost Certain	3	3	MEDIUM	HIGH	HIGH

No.	Description of Hazard	Hazard Score	Initial Likelihood Score	Initial Risk	Controls	Residual Likelihood Score	Residual Risk
1	Some of the anticipated hazards in doing face-to-face interviews may include: Going to the school sites. Finding a private, well-ventilated, and open place in the school. Coronavirus transmission.	1	1		To mitigate the risks in the interviews, the researcher will make sure that the research participants follow COVID-19 safety guidance and safety practices, such as maintaining social distance, wearing a face cover or a mask and not attend the interview if they encounter any health issues. Moreover, the researcher will conduct the interviews in May of the school year where teachers are not very busy with their day-to-day teaching or other schooling duties and the school will not be crowded with staff and students. This would enable the researcher to arrange a flexible time and a secure room for doing the interviews and will reduce the risk of meeting many people in the schools. Prior conducting the interviews, the researcher will contact the school leadership and request a private and well-ventilated room for doing the interviews. Additionally, the researcher will have a regular PCR test every week from the beginning of the school visits to eliminate any threat of the virus transmission.	1	1

2	Some of the anticipated hazards in doing classroom observations may include: Going to the school sites. Finding a private, well-ventilated classroom. Coronavirus transmission.	1	1	1	The arrangement of classroom observation will be negotiated and discussed with the teachers through the email. The researcher does not intend to discuss any issues with the teacher or the students. The researcher plans to sit at the end corner of the class, maintain social distancing and wear mask throughout the whole lesson. Another issue to safeguard all parties including the researcher, only schools in the Low-density category will be targeted to minimise the chances of transmitting the virus and maintain social distancing. Furthermore, the researcher will check in with the teacher, who will be observed in their classroom, the health measures taken in classroom before the day of the classroom visit.	1	1
---	--	---	---	---	---	---	---

Section 5: Assessment Sign-Off

Assessor's Signature	<i>Safa</i>	Position	PhD student
Print Name	Safa	Date	13/2/2021
Additional Comments			

Assessment Agreed by		Position	
Print Name		Date and Time	
Additional Comments			

Section 6: Communication of Risk Assessment

I have read and understood the contents of this risk assessment.

Name	Date	Signature
Safa Al-Hinai	13/2/2021	Safa

Appendix O. Omani Ministry of Education Approval

الرقم :
التاريخ :
18/2/2021 هـ الموافق



سلطنة عمان
وزارة التربية والتعليم

To Whom it May Concern
Subject/ Approval of Research Study

We are pleased to inform you that the Ministry of Education has approved applying the tools of the study entitled (Examining the Washback of English Elective Exam on Classroom Teaching in Grade 12 Omani Schools) presented by Safaa bint Ibrahim bin Mohammed Al-Hinaeyah. She is permitted to apply (a questionnaire, interviews, an observation form) on a sample of Elective English teachers in post-basic schools in all General Directorates. The interviews and the observation form will be conducted inside the schools. This letter has been given to her upon her request without the Ministry taking any responsibility of this, and in case of any inquiry, please contact the Technical Office for Studies and Development at 24255303 or 242555134, or email tosd@moe.om.

Yours sincerely,

Assistant Director for International Cooperation Affairs, Technical Office for Studies and Development



Appendix P. Consent Forms

MoE Decision-Makers – Interviews



Information Sheet

Examining the Washback of English Elective Exam on Classroom Teaching in Grade 12 Omani Schools

Dear participant,

My name is Safa Al-Hinai. I am currently carrying out a research project to investigate attitudinal and behavioural changes among teachers concerning the new English Elective Diploma Exam introduced in the school year 2018/2019. Specifically, this study seeks to gather information about how teachers feel about the English Elective Diploma Exam changes and the preparation they are carrying out in the classroom to cope with the changes in this exam. The study also identifies the existence of factors that seem to influence teaching beyond the exam itself. I would like to invite you to take part in this project.

What would this mean for you?

This would mean:

- Taking part in an interview lasting between 25 and 30 minutes, which will take place at a time and place convenient to you. A record of the interview (notes/recording) will be collected, but your name will not be attached. You may see and comment on this record if you wish.

Participation is voluntary

Participation is optional. If you do decide to take part, you will be given a copy of this information sheet for your records and will be asked to sign a consent form. If you change your mind at any point during the interview, you will be able to withdraw your participation without having to provide a reason. If you want to stop taking part, please inform the researcher.

Anonymity and confidentiality

The data that you provide (notes or audio recordings of the interview) will be stored by code number and will be treated confidentially. Any information that identifies you will be stored separately from the data. You are free to withdraw from the study at any time during data collection and up to two weeks.

Storing and using your data

Data will be stored on a password protected computer. Data will be fully anonymised 3 weeks after transcription of data is completed.

You will be given the opportunity to comment on a written record of your interview 1-2 weeks by emailing me at: siah501@york.ac.uk , after which time the transcript will be anonymised and such requests will not be possible . Anonymised data will be kept for three years after which time it will be destroyed.

Please note: If information is gathered that raises concerns about your safety or the safety of others, or about other concerns as perceived by the researcher, the researcher may pass on this information to another person.

Questions or concerns

If you have any questions about this participant information sheet or concerns about how your data is being processed, please feel free to contact Dr Khaled el Ebyary (khaled.elebyary@york.ac.uk) or the Chair of Ethics Committee in the Education Department, University of York via email education-research-admin@york.ac.uk. If you are still dissatisfied, please contact the University of York Data Protection Officer at dataprotection@york.ac.uk.

I hope that you will agree to take part. If you are happy to participate and for your pseudonymised data to be used in the ways listed, please keep this information sheet for your own records. Thank you for taking the time to read this information.

Yours sincerely,

Safa Al-Hinai

Information about the General Data Protection Regulation (GDPR)

Processing personal data

Under the General Data Protection Regulation (GDPR), the University has to identify a legal basis for processing personal data and, where appropriate, an additional condition for processing special category data. In line with our charter which states that we advance learning and knowledge by teaching and research, the University processes personal data for research purposes under Article 6 (1)(e) of the GDPR:

Processing is necessary for the performance of a task carried out in the public interest

Special category data is processed under Article 9 (2) (j):

Processing is necessary for archiving purposes in the public interest, or scientific and historical research purposes or statistical purposes

Research will only be undertaken where ethical approval has been obtained, where there is a clear public interest and where appropriate safeguards have been put in place to protect data.

In line with ethical expectations and in order to comply with common law duty of confidentiality, we will seek your consent to participate where appropriate. This consent will not, however, be our legal basis for processing your data under the GDPR.

Protecting and storing personal data

Information that research participants provide will be treated confidentially and shared on a need-to-know basis only. The University is committed to the principle of data protection by design and default and will collect the minimum amount of data necessary for the project. In addition we will anonymise or pseudonymise data wherever possible.

We will put in place appropriate technical and organisational measures to protect your personal data and/or special category data (for example, data may be stored in secure filing cabinets and/or on a password protected computer).

Sharing of data

The default position is that personal data will only be accessible to members of the project team. In some cases, however, the research may be of a collaborative nature and hence the data will be made accessible to others from outside the University. Information specific to the project will include details of when this is the case, who the 3rd parties are, and what they will do with the data. It is possible that personal data may be shared anonymously with others for secondary research and/or teaching purposes.

Transfer of data internationally

The default position is that data will be stored on University devices and held within the European Economic Area in full compliance with data protection legislation.

However, data may be transferred to the project partners based outside the European Economic Area. Any international transfer will be undertaken in full compliance with the GDPR.

The University has access to cloud storage provided by Google which means that data can be located at any of Google's globally spread data centres. The University has data protection compliant arrangements in place with this provider. For further information see, <https://www.york.ac.uk/it-services/google/policy/privacy/>

Your rights in relation to your data

Under the GDPR, you have a general right of access to your data, a right to rectification, erasure, restriction, objection or portability. You also have a right to withdrawal. Please note, not all rights apply

where data is processed purely for research purposes. For information see, <https://www.york.ac.uk/records-management/dp/individualsrights/>

Right to complain

If you are unhappy with the way in which your personal data has been handled, you have a right to complain to the Information Commissioner's Office. For information on reporting a concern to the Information Commissioner's Office, see



Examining the Washback of English Elective Exam on Classroom Teaching in Grade 12 Omani Schools

Consent Form

**Please tick each box, sign and return to the researcher if you agree to take part.
Please feel free to ask questions before you decide to take part**

	I agree (please tick)
I confirm that I have read and understood the information given to me about the above named research project and I understand that this will involve taking part as described above.	
I understand that participation in this study is voluntary.	
I understand that my data will not be identifiable and the data may be used pseudonymously in publications, presentations and online.	
I confirm that I have read the information about GDPR	

Name: _____

Signature: _____

Date: _____

Thank you for reading this document.
Please return this signed consent form to the researcher.

THE UNIVERSITY *of York*

Information Sheet

Examining the Washback of English Elective Exam on Classroom Teaching in Grade 12 Omani Schools

Dear school principal,

My name is Safa Al-Hinai. I am currently carrying out a research project to investigate attitudinal and behavioural changes among teachers concerning the new English Elective Diploma Exam introduced in the school year 2018/2019. Specifically, this study seeks to gather information about how teachers feel about the English Elective Diploma Exam changes and the preparation they are carrying out in the classroom to cope with the changes in this exam. The study also identifies the existence of factors that seem to influence teaching beyond the exam itself. I would like to invite you to take part in this project.

What would this mean for your school?

This would mean:

- Observations of English elective classroom lessons. A record of the observation will be made, but your teacher(s) name will not be included in any written records of the lesson in the dissertation or elsewhere.

Participation is voluntary

Participation is optional. If you do decide to take part, you will be given a copy of this information sheet for your records and will be asked to sign a consent form. If you change your mind at any point during the study, you will be able to withdraw your participation without having to provide a reason. If you want to stop taking part, please inform the researcher.

Anonymity and confidentiality

The data that your school provide (notes or audio recordings of the observation) will be stored by a code number and will be treated confidentially. Any information that identifies your teacher(s) will be stored separately from the data. Your teacher(s) are free to withdraw from the study at any time during data collection. After data collection, data will be anonymised so it will not be possible to withdraw your teacher(s) data after this point.

Storing and using your data

Data will be stored on a password protected computer. The data that this researcher collect (notes / audio recordings) may be used in an *anonymous* format in different ways (presentations, reports and online

publications) and will be destroyed when publications have resulted from the project, or 3 years after the project has completed, whichever is soonest.

Please note: If information is gathered that raises concerns about your teacher(s) safety or the safety of others, or about other concerns as perceived by the researcher, the researcher may pass on this information to another person.

Questions or concerns

If you have any questions about this participant information sheet or concerns about how your teacher(s) data is being processed, please feel free to contact Dr Khaled el Ebyary (khaled.elebyary@york.ac.uk) or the Chair of Ethics Committee in the Education Department, University of York via email education-research-admin@york.ac.uk. If you are still dissatisfied, please contact the University of York Data Protection Officer at dataprotection@york.ac.uk.

I hope that you will agree to your teacher(s) taking part. If you are happy for your school to participate and for your pseudonymised data to be used in the ways listed, please keep this information sheet for your own records. Thank you for taking the time to read this information.

Thank you for taking the time to read this information.

Yours sincerely,

Safa Al-Hinai

Information about the General Data Protection Regulation (GDPR)

Processing personal data

Under the General Data Protection Regulation (GDPR), the University has to identify a legal basis for processing personal data and, where appropriate, an additional condition for processing special category data. In line with our charter which states that we advance learning and knowledge by teaching and research, the University processes personal data for research purposes under Article 6 (1)(e) of the GDPR:

Processing is necessary for the performance of a task carried out in the public interest

Special category data is processed under Article 9 (2) (j):

Processing is necessary for archiving purposes in the public interest, or scientific and historical research purposes or statistical purposes

Research will only be undertaken where ethical approval has been obtained, where there is a clear public interest and where appropriate safeguards have been put in place to protect data.

In line with ethical expectations and in order to comply with common law duty of confidentiality, we will seek your consent to participate where appropriate. This consent will not, however, be our legal basis for processing your data under the GDPR.

Protecting and storing personal data

Information that research participants provide will be treated confidentially and shared on a need-to-know basis only. The University is committed to the principle of data protection by design and default and will collect the minimum amount of data necessary for the project. In addition we will anonymise or pseudonymise data wherever possible.

We will put in place appropriate technical and organisational measures to protect your personal data and/or special category data (for example, data may be stored in secure filing cabinets and/or on a password protected computer).

Sharing of data

The default position is that personal data will only be accessible to members of the project team. In some cases, however, the research may be of a collaborative nature and hence the data will be made accessible to others from outside the University. Information specific to the project will include details of when this is the case, who the 3rd parties are, and what they will do with the data. It is possible that personal data may be shared anonymously with others for secondary research and/or teaching purposes.

Transfer of data internationally

The default position is that data will be stored on University devices and held within the European Economic Area in full compliance with data protection legislation.

However, data may be transferred to the project partners based outside the European Economic Area. Any international transfer will be undertaken in full compliance with the GDPR.

The University has access to cloud storage provided by Google which means that data can be located at any of Google's globally spread data centres. The University has data protection compliant arrangements in place with this provider. For further information see, <https://www.york.ac.uk/it-services/google/policy/privacy/>

Your rights in relation to your data

Under the GDPR, you have a general right of access to your data, a right to rectification, erasure, restriction, objection or portability. You also have a right to withdrawal. Please note, not all rights apply where data is processed purely for research purposes. For information

see, <https://www.york.ac.uk/records-management/dp/individualsrights/>

Right to complain

If you are unhappy with the way in which your personal data has been handled, you have a right to complain to the Information Commissioner's Office. For information on reporting a concern to the Information Commissioner's Office, see



Examining the Washback of English Elective Exam on Classroom Teaching in Grade 12 Omani Schools

Consent Form

**Please tick each box, sign and return to the researcher if you agree to take part.
Please feel free to ask questions before you decide to take part.**

	I agree (please tick)
I confirm that I have read and understood the information given to me about the above named research project and I understand that this will involve taking part as described above.	
I understand that participation in this study is voluntary.	
I understand that my data will not be identifiable and the data may be used pseudonymously in publications, presentations and online.	
I confirm that I have read the information about GDPR	

Name: _____

Signature: _____

Date: _____

Thank you for reading this document.
Please return this signed consent form to the researcher.

THE UNIVERSITY *of York*

Information Sheet

Examining the Washback of English Elective Exam on Classroom Teaching in Grade 12 Omani Schools

Dear teacher,

My name is Safa Al-Hinai. I am currently carrying out a research project to investigate attitudinal and behavioural changes among teachers concerning the new English Elective Diploma Exam introduced in the school year 2018/2019. Specifically, this study seeks to gather information about how teachers feel about the English Elective Diploma Exam changes and the preparation they are carrying out in the classroom to cope with the changes in this exam. The study also identifies the existence of factors that seem to influence teaching beyond the exam itself. I would like to invite you to take part in this project

What would this mean for you?

This would mean:

- Observations of your English elective classroom lessons. A record of the observation will be made, but your name will not be included in any written records of the lesson in the dissertation or elsewhere.

Participation is voluntary

Participation is optional. If you do decide to take part, you will be given a copy of this information sheet for your records and will be asked to sign a consent form. If you change your mind at any point during observation you will be able to withdraw your participation without having to provide a reason. If you want to stop taking part, please inform the researcher via the email provided below.

Anonymity and confidentiality

The data that you provide (notes or audio recordings of the observation) will be stored by a code number and will be treated confidentially. Any information that identifies you will be stored separately from the data. You will be given the opportunity to comment on a written record of your observation. You will have two weeks to respond or make changes to this. You are free to withdraw from the study at any time during data collection. After data collection, and following your comments on observation notes and scheme, data will be anonymised so it will not be possible to withdraw your data after this point.

Storing and using your data

Data will be stored on a password protected computer. The data that this researcher collect (notes / audio recordings) may be used in an *anonymous* format in different ways (presentations, reports and online publications) and will be destroyed when publications have resulted from the project, or 3 years after the project has completed, whichever is soonest.

Please note: If information is gathered that raises concerns about your safety or the safety of others, or about other concerns as perceived by the researcher, the researcher may pass on this information to another person.

Questions or concerns

If you have any questions about this participant information sheet or concerns about how your data is being processed, please feel free to contact Dr Khaled el Ebyary (khaled.elebyary@york.ac.uk) or the Chair of Ethics Committee in the Education Department, University of York via email education-research-admin@york.ac.uk. If you are still dissatisfied, please contact the University of York Data Protection Officer at dataprotection@york.ac.uk.

I hope that you will agree to take part. If you are happy to participate and for your pseudonymised data to be used in the ways listed, please keep this information sheet for your own records. Thank you for taking the time to read this information.

Thank you for taking the time to read this information.

Yours sincerely,
Safa Al-Hinai

Information about the General Data Protection Regulation (GDPR)

Processing personal data

Under the General Data Protection Regulation (GDPR), the University has to identify a legal basis for processing personal data and, where appropriate, an additional condition for processing special category data. In line with our charter which states that we advance learning and knowledge by teaching and research, the University processes personal data for research purposes under Article 6 (1)(e) of the GDPR:

Processing is necessary for the performance of a task carried out in the public interest

Special category data is processed under Article 9 (2) (j):

Processing is necessary for archiving purposes in the public interest, or scientific and historical research purposes or statistical purposes

Research will only be undertaken where ethical approval has been obtained, where there is a clear public interest and where appropriate safeguards have been put in place to protect data.

In line with ethical expectations and in order to comply with common law duty of confidentiality, we will seek your consent to participate where appropriate. This consent will not, however, be our legal basis for processing your data under the GDPR.

Protecting and storing personal data

Information that research participants provide will be treated confidentially and shared on a need-to-know basis only. The University is committed to the principle of data protection by design and default and will collect the minimum amount of data necessary for the project. In addition we will anonymise or pseudonymise data wherever possible.

We will put in place appropriate technical and organisational measures to protect your personal data and/or special category data (for example, data may be stored in secure filing cabinets and/or on a password protected computer).

Sharing of data

The default position is that personal data will only be accessible to members of the project team. In some cases, however, the research may be of a collaborative nature and hence the data will be made accessible to others from outside the University. Information specific to the project will include details of when this is the case, who the 3rd parties are, and what they will do with the data. It is possible that personal data may be shared anonymously with others for secondary research and/or teaching purposes.

Transfer of data internationally

The default position is that data will be stored on University devices and held within the European Economic Area in full compliance with data protection legislation.

However, data may be transferred to the project partners based outside the European Economic Area. Any international transfer will be undertaken in full compliance with the GDPR.

The University has access to cloud storage provided by Google which means that data can be located at any of Google's globally spread data centres. The University has data protection compliant arrangements in place with this provider. For further information see, <https://www.york.ac.uk/it-services/google/policy/privacy/>

Your rights in relation to your data

Under the GDPR, you have a general right of access to your data, a right to rectification, erasure, restriction, objection or portability. You also have a right to withdrawal. Please note, not all rights apply

where data is processed purely for research purposes. For information see, <https://www.york.ac.uk/records-management/dp/individualsrights/>

Right to complain

If you are unhappy with the way in which your personal data has been handled, you have a right to complain to the Information Commissioner's Office. For information on reporting a concern to the Information Commissioner's Office, see



Examining the Washback of English Elective Exam on Classroom Teaching in Grade 12 Omani Schools

Consent Form

**Please tick each box, sign and return to the researcher if you agree to take part.
Please feel free to ask questions before you decide to take part.**

	I agree (please tick)
I confirm that I have read and understood the information given to me about the above named research project and I understand that this will involve taking part as described above.	
I understand that participation in this study is voluntary.	
I understand that my data will not be identifiable and the data may be used pseudonymously in publications, presentations and online.	
I confirm that I have read the information about GDPR	

Name: _____

Signature: _____

Date: _____

Thank you for reading this document.
Please return this signed consent form to the researcher.

THE UNIVERSITY *of York*

Information Sheet

Examining the Washback of English Elective Exam on Classroom Teaching in Grade 12 Omani Schools

Dear Teacher,

My name is Safa Al-Hinai. I am currently carrying out a research project to investigate attitudinal and behavioural changes among teachers concerning the new English Elective Diploma Exam introduced in the school year 2018/2019. Specifically, this study seeks to gather information about how teachers feel about the English Elective Diploma Exam changes and the preparation they are carrying out in the classroom to cope with the changes in this exam. The study also identifies the existence of factors that seem to influence teaching beyond the exam itself. I would like to invite you to take part in this project

What would this mean for you?

This would mean:

- Taking part in an interview lasting between 25 and 30 minutes, which will take place at a time and place convenient to you. A record of the interview (notes/recording) will be collected, but your name will not be attached. You may see and comment on this record if you wish.

Participation is voluntary

Participation is optional. If you do decide to take part, you will be given a copy of this information sheet for your records and will be asked to sign a consent form. If you change your mind at any point during the interview, you will be able to withdraw your participation without having to provide a reason. If you want to stop taking part, please inform the researcher.

Anonymity and confidentiality

The data that you provide (notes or audio recordings of the interview) will be stored by code number and will be treated confidentially. Any information that identifies you will be stored separately from the data. You are free to withdraw from the study at any time during data collection and up to two weeks.

Storing and using your data

Data will be stored on a password protected computer. Data will be fully anonymised 3 weeks after transcription of data is completed.

You will be given the opportunity to comment on a written record of your interview 1-2 weeks by emailing me at : siah501@york.ac.uk , after which time the transcript will be anonymised and such requests will not be possible . Anonymised data will be kept for three years after which time it will be destroyed.

Please note: If information is gathered that raises concerns about your safety or the safety of others, or about other concerns as perceived by the researcher, the researcher may pass on this information to another person.

Questions or concerns

If you have any questions about this participant information sheet or concerns about how your data is being processed, please feel free to contact Dr Khaled el Ebyary (khaled.elebyary@york.ac.uk) or the Chair of Ethics Committee in the Education Department, University of York via email education-research-admin@york.ac.uk. If you are still dissatisfied, please contact the University of York Data Protection Officer at dataprotection@york.ac.uk.

I hope that you will agree to take part. If you are happy to participate and for your pseudonymised data to be used in the ways listed, please keep this information sheet for your own records. Thank you for taking the time to read this information.

Yours sincerely,
Safa Al-Hinai

Information about the General Data Protection Regulation (GDPR)

Processing personal data

Under the General Data Protection Regulation (GDPR), the University has to identify a legal basis for processing personal data and, where appropriate, an additional condition for processing special category data. In line with our charter which states that we advance learning and knowledge by teaching and research, the University processes personal data for research purposes under Article 6 (1)(e) of the GDPR:

Processing is necessary for the performance of a task carried out in the public interest

Special category data is processed under Article 9 (2) (j):

Processing is necessary for archiving purposes in the public interest, or scientific and historical research purposes or statistical purposes

Research will only be undertaken where ethical approval has been obtained, where there is a clear public interest and where appropriate safeguards have been put in place to protect data.

In line with ethical expectations and in order to comply with common law duty of confidentiality, we will seek your consent to participate where appropriate. This consent will not, however, be our legal basis for processing your data under the GDPR.

Protecting and storing personal data

Information that research participants provide will be treated confidentially and shared on a need-to-know basis only. The University is committed to the principle of data protection by design and default and will collect the minimum amount of data necessary for the project. In addition we will anonymise or pseudonymise data wherever possible.

We will put in place appropriate technical and organisational measures to protect your personal data and/or special category data (for example, data may be stored in secure filing cabinets and/or on a password protected computer).

Sharing of data

The default position is that personal data will only be accessible to members of the project team. In some cases, however, the research may be of a collaborative nature and hence the data will be made accessible to others from outside the University. Information specific to the project will include details of when this is the case, who the 3rd parties are, and what they will do with the data. It is possible that personal data may be shared anonymously with others for secondary research and/or teaching purposes.

Transfer of data internationally

The default position is that data will be stored on University devices and held within the European Economic Area in full compliance with data protection legislation.

However, data may be transferred to the project partners based outside the European Economic Area. Any international transfer will be undertaken in full compliance with the GDPR.

The University has access to cloud storage provided by Google which means that data can be located at any of Google's globally spread data centres. The University has data protection compliant arrangements in place with this provider. For further information see,

<https://www.york.ac.uk/it-services/google/policy/privacy/>

Your rights in relation to your data

Under the GDPR, you have a general right of access to your data, a right to rectification, erasure, restriction, objection or portability. You also have a right to withdrawal. Please note, not all rights apply where data is processed purely for research purposes. For information see, <https://www.york.ac.uk/records-management/dp/individualsrights/>

Right to complain

If you are unhappy with the way in which your personal data has been handled, you have a right to complain to the Information Commissioner's Office. For information on reporting a concern to the Information Commissioner's Office, see



Examining the Washback of English Elective Exam on Classroom Teaching in Grade 12 Omani Schools

Consent Form

**Please tick each box, sign and return to the researcher if you agree to take part.
Please feel free to ask questions before you decide to take part**

	I agree (please tick)
I confirm that I have read and understood the information given to me about the above named research project and I understand that this will involve taking part as described above.	
I understand that participation in this study is voluntary.	
I understand that my data will not be identifiable and the data may be used pseudonymously in publications, presentations and online.	
I confirm that I have read the information about GDPR	

Name: _____

Signature: _____

Date: _____

Thank you for reading this document.
Please return this signed consent form to the researcher.

Appendix Q. Supplementary Tables

Table P1. Number of learning activities

Variable	Category	<i>N</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>p</i>	<i>d</i>
Gender	Female	114	8.070	2.956	-0.749	123.636	0.455	-0.120
	Male	70	8.457	3.646				
Academic qualification	Bachelor's	156	8.038	3.217	-1.783	182.000	0.076	-0.366
	Postgraduate	28	9.214	3.190				
Currently teaching elective	Yes	68	8.574	3.220	1.145	182.000	0.254	0.175
	No	116	8.009	3.234				
Training events	Yes	57	8.719	3.374	1.415	182.000	0.159	0.226
	No	127	7.992	3.153				
Variable	Category	<i>N</i>	<i>M</i>	<i>SD</i>	<i>F</i> (3, 180)	<i>p</i>	η^2	
Age	20–30	14	9.071	2.786	0.612	0.608	0.010	
	31–40	110	7.991	3.006				
	41–50	50	8.480	3.382				
	> 50	10	8.200	5.224				
Variable	Category	<i>N</i>	<i>M</i>	<i>SD</i>	<i>W</i> (3, 58.427)	<i>p</i>	η^2	
Teaching experience	0–6	15	9.533	2.476	2.299	0.087	0.026	
	7–10	35	7.400	2.902				
	11–15	66	8.258	3.149				
	> 15	68	8.309	3.550				

Table P2. Doing more examination practice

Variable	Category	No		Yes		Value	<i>df</i>	χ^2 <i>p</i>	Cramér's <i>V</i>
		<i>N</i>	%	<i>N</i>	%				
Gender	Female	51	42.9	68	57.1	.156	1	0.693	0.028
	Male	37	45.7	44	54.3				
Academic qualification	Bachelor's	70	41.9	97	58.1	1.784	1	0.182	0.094
	Postgraduate	18	54.5	15	45.5				
Age	20–30	9	52.9	8	47.1	3.593	3	0.309	0.134
	31–40	50	42.4	68	57.6				
	41–50	21	39.6	32	60.4				
	> 50	8	66.7	4	33.3				
Currently teaching elective	Yes	33	45.2	40	54.8	.068	1	0.795	0.018
	No	55	43.3	72	56.7				
Training events	Yes	33	52.4	30	47.6	2.622	1	0.105	0.114
	No	55	40.1	82	59.9				
Teaching experience	0–6	9	50.0	9	50.0	2.682	3	0.443	0.116
	7–10	13	35.1	24	64.9				
	11–15	35	50.0	35	50.0				
	> 15	31	41.3	44	58.7				

****p* < 0.001, ***p* < 0.01, **p* < 0.05

Table P3. Teaching according to the new examination specifications

Variable	Category	Median	Mean Rank	<i>U</i>	<i>z</i>	<i>p</i>
Gender	Female	4	100.24	4067.000	-1.092	0.275
	Male	4	92.01			
Academic qualification	Bachelor's	4	96.31	2558.000	0.439	0.661
	Postgraduate	4	100.77			
Currently teaching elective	Yes	4	102.74	3903.500	-1.176	0.240
	No	4	93.74			
Training events	Yes	4	97.25	3938.500	-0.044	0.965
	No	4	96.89			
Variable	Category	Median	Mean Rank	χ^2	<i>df</i>	<i>p</i>
Age	20–30	4	96.07	0.150	3	0.985
	31–40	4	96.38			
	41–50	4	97.51			
	> 50	4	102.41			
Teaching Experience	0–6	4	95.25	0.321	3	0.956
	7–10	4	94.09			
	11–15	4	96.35			
	> 15	4	99.49			

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table P4. Adopting new teaching methods

Variable	Category	Median	Mean Rank	<i>U</i>	<i>z</i>	<i>p</i>
Gender	Female	4	94.44	4745.500	0.896	0.370
	Male	4	100.94			
Academic qualification	Bachelor's	4	96.16	2582.000	0.553	0.580
	Postgraduate	4	101.57			
Currently reaching elective	Teaching	4	98.15	4224.500	-0.245	0.807
	Not teaching	4	96.35			
Training events	Training	4	95.33	4051.500	0.313	0.755
	No training	4	97.74			
Variable	Category	Median	Mean Rank	χ^2	<i>df</i>	<i>p</i>
Age	20–30	4	82.71	3.761	3	0.288
	31–40	4	94.42			
	41–50	4	107.31			
	> 50	4	94.82			
Teaching experience	0–6	4	83.31	5.393	3	0.145
	7–10	4	87.36			
	11–15	4	95.43			
	> 15	4	106.48			

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table P5. Putting more emphasis on communicative practices

Variable	Category	Median	Mean Rank	<i>U</i>	<i>z</i>	<i>p</i>
Gender	Female	4	95.64	4605.500	0.469	0.639
	Male	4	99.10			
Academic qualification	Bachelor's	4	97.33	2391.000	-0.214	0.831
	Postgraduate	4	95.20			
Currently teaching elective	Yes	4	95.81	4388.500	0.249	0.803
	No	4	97.68			
Training events	Yes	4	91.20	4295.000	1.066	0.287
	No	4	99.55			
Variable	Category	Median	Mean Rank	χ^2	<i>df</i>	<i>p</i>
Age	20–30	4	94.54	5.084	3	0.166
	31–40	4	91.08			
	41–50	4	108.28			
	> 50	4	110.82			
Teaching experience	0–6	4	95.78	3.720	3	0.293
	7–10	4	88.50			
	11–15	4	92.73			
	> 15	4	105.67			

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table P6. Putting more emphasis on receptive skills

Variable	Category	Median	Mean Rank	<i>U</i>	<i>z</i>	<i>p</i>
Gender	Female	4	102.54	3798.000	-1.932	0.053
	Male	4	88.47			
Academic qualification	Bachelor's	4	97.33	2392.000	-0.213	0.831
	Postgraduate	4	95.23			
Currently teaching elective	Yes	4	102.32	3932.500	-1.129	0.259
	No	4	93.97			
Training events	Yes	4	99.10	3829.000	-0.392	0.695
	No	4	96.07			
Variable	Category	Median	Mean Rank	χ^2	<i>df</i>	<i>p</i>
Age	20–30	4	122.36	4.509	3	0.211
	31–40	4	95.08			
	41–50	4	97.02			
	> 50	4	85.05			
Teaching experience	0–6	4	103.94	1.935	3	0.586
	7–10	4	89.41			
	11–15	4	101.71			
	> 15	4	94.92			

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table P7. Putting more emphasis on productive skills

Variable	Category	Median	Mean Rank	<i>U</i>	<i>z</i>	<i>p</i>
Gender	Female	4	102.26	3831.000	-1.877	0.061
	Male	4	88.91			
Academic qualification	Bachelor's	4	97.95	2290.500	-0.636	0.525
	Postgraduate	4	91.85			
Currently reaching elective	Yes	4	105.34	3721.000	-1.811	0.070
	No	4	92.25			
Training events	Yes	4	89.11	4418.500	1.506	0.132
	No	4	100.47			
Variable	Category	Median	Mean Rank	χ^2	<i>df</i>	<i>p</i>
Age	20–30	4	106.61	4.389	3	0.222
	31–40	4	97.82			
	41–50	4	98.57			
	> 50	4	68.82			
Teaching experience	0–6	4	92.88	0.294	3	0.961
	7–10	4	100.03			
	11–15	4	97.34			
	> 15	4	96.04			

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table P8. Employing more authentic language tasks

Variable	Category	Median	Mean Rank	<i>U</i>	<i>z</i>	<i>p</i>
Gender	Female	4	100.95	3984.00	-1.350	0.177
	Male	4	90.92			
Academic qualification	Bachelor's	4	96.79	2478.500	0.132	0.895
	Postgraduate	4	98.12			
Currently teaching elective	Yes	4	107.09	3598.500	-2.098	0.036*
	No	4	91.26			
Training events	Yes	4	97.87	3901.500	-0.160	0.873
	No	4	96.62			
Variable	Category	Median	Mean Rank	χ^2	<i>df</i>	<i>p</i>
Age	20–30	4	92.82	6.666	3	0.083
	31–40	4	96.25			
	41–50	4	106.88			
	> 50	4	64.45			
Teaching experience	0–6	4	86.53	1.1445	3	0.695
	7–10	4	102.22			
	11–15	4	94.20			
	> 15	4	99.29			

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table P9. Motivating students

Variable	Category	Median	Mean Rank	<i>U</i>	<i>z</i>	<i>p</i>
Gender	Female	4	96.12	4548.500	0.295	0.768
	Male	4	98.35			
Academic qualification	Bachelor's	4	97.52	2360.000	-0.330	0.741
	Postgraduate	4	94.17			
Currently teaching elective	Yes	4	104.39	3787.500	-1.516	0.130
	No	4	92.79			
Training events	Yes	4	99.58	3800.500	-0.466	0.641
	No	4	95.86			

Variable	Category	Median	Mean Rank	χ^2	<i>df</i>	<i>p</i>
Age	20–30	4.5	90.39	.509	3.000	0.917
	31–40	4	96.55			
	41–50	4	98.23			
	> 50	4	104.50			
Teaching experience	0–6	4	112.94	2.483	3.000	0.478
	7–10	4	92.88			
	11–15	4	92.66			
	> 15	4	99.67			

****p* < 0.001, ***p* < 0.01, **p* < 0.05

Table P10. Day-to-day observation

Variable	Category	No		Yes		Value	<i>df</i>	χ^2 <i>p</i>	Cramér's <i>V</i>
		<i>N</i>	%	<i>N</i>	%				
Gender	Female	39	32.8	80	67.2	0.388	1	0.533	0.044
	Male	30	37.0	51	63.0				
Academic qualification	Bachelor's	57	34.1	110	65.9	0.061	1	0.805	0.017
	Postgraduate	12	36.4	21	63.6				
Age	20–30	6	35.3	11	64.7	6.045	3	0.109	0.174
	31–40	39	33.1	79	66.9				
	41–50	16	30.2	37	69.8				
	> 50	8	66.7	4	33.3				
Currently teaching elective	Yes	27	37.0	46	63.0	0.314	1	0.575	0.040
	No	42	33.1	85	66.9				
Training events	Yes	23	36.5	40	63.5	0.164	1	0.685	0.029
	No	46	33.6	91	66.4				
Teaching experience	0–6	5	27.8	13	72.2	1.101	3	0.777	0.074
	7–10	14	37.8	23	62.2				
	11–15	22	31.4	48	68.6				
	> 15	28	37.3	47	62.7				

****p* < 0.001, ***p* < 0.01, **p* < 0.05

Table P11. Presentations

Variable	Category	No		Yes		Value	df	χ^2		Cramér's V
		N	%	N	%			p		
Gender	Female	64	53.8	55	46.2	0.351	1	0.553	0.042	
	Male	47	58.0	34	42.0					
Academic qualification	Bachelor's	93	55.7	74	44.3	0.015	1	0.904	0.009	
	Postgraduate	18	54.5	15	45.5					
Age	20–30	11	64.7	6	35.3	2.089	3	0.554	0.102	
	31–40	66	55.9	52	44.1					
	41–50	26	49.1	27	50.9					
	> 50	8	66.7	4	33.3					
Currently teaching elective	Yes	41	56.2	32	43.8	0.021	1	0.886	0.010	
	No	70	55.1	57	44.9					
Training events	Yes	32	50.8	31	49.2	0.825	1	0.364	0.064	
	No	79	57.7	58	42.3					
Teaching experience	0–6	10	55.6	8	44.4	3.289	3	0.349	0.128	
	7–10	24	64.9	13	35.1					
	11–15	41	58.6	29	41.4					
	> 15	36	48.0	39	52.0					

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table P12. Written work

Variable	Category	No		Yes		Value	df	χ^2		Cramér's V
		N	%	N	%			p		
Gender	Female	32	26.9	87	73.1	6.618	1	0.010*	0.182	
	Male	36	44.4	45	55.6					
Academic qualification	Bachelor's	59	35.3	108	64.7	0.797	1	0.372	0.063	
	Postgraduate	9	27.3	24	72.7					
Age	20–30	7	41.2	10	58.8	14.813	3	0.002**	0.272	
	31–40	36	30.5	82	69.5					
	41–50	15	28.3	38	71.7					
	> 50	10	83.3	2	16.7					
Currently teaching elective	Yes	26	35.6	47	64.4	0.134	1	0.714	0.026	
	No	42	33.1	85	66.9					
Attended training	Yes	25	39.7	38	60.3	1.323	1	0.250	0.081	
	No	43	31.4	94	68.6					
Teaching experience	0–6	8	44.4	10	55.6	1.783	3	0.619	0.094	
	7–10	10	27.0	27	73.0					
	11–15	25	35.7	45	64.3					
	> 15	25	33.3	50	66.7					

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table K13. Quizzes

Variable	Category	No		Yes		Value	df	χ^2 p	Cramér's V
		N	%	N	%				
Gender	Female	44	37.0	75	63.0	1.515	1	0.218	0.087
	Male	37	45.7	44	54.3				
Academic qualification	Bachelor's	67	40.1	100	59.9	0.061	1	0.805	0.017
	Postgraduate	14	42.4	19	57.6				
Age	20–30	6	35.3	11	64.7	1.814	3	0.612	0.095
	31–40	47	39.8	71	60.2				
	41–50	21	39.6	32	60.4				
	> 50	7	58.3	5	41.7				
Currently teaching elective	Yes	35	47.9	38	52.1	2.644	1	0.104	0.115
	No	46	36.2	81	63.8				
Attended training	Yes	29	46.0	34	54.0	1.168	1	0.280	0.076
	No	52	38.0	85	62.0				
Teaching experience	0–6	7	38.9	11	61.1	1.374	3	0.712	0.083
	7–10	18	48.6	19	51.4				
	11–15	26	37.1	44	62.9				
	> 15	30	40.0	45	60.0				

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table K14. Class tests

Variable	Characteristic	No		Yes		Value	df	χ^2 p	Cramér's V
		N	%	N	%				
Gender	Female	53	44.5	66	55.5	2.894	1	0.089	0.120
	Male	46	56.8	35	43.2				
Academic qualification	Bachelor's	82	49.1	85	50.9	.064	1	0.800	0.018
	Postgraduate	17	51.5	16	48.5				
Age	20–30	9	52.9	8	47.1	1.678	3	0.642	0.092
	31–40	57	48.3	61	51.7				
	41–50	25	47.2	28	52.8				
	> 50	8	66.7	4	33.3				
Currently teaching elective	Yes	34	46.6	39	53.4	.393	1	0.531	0.044
	No	65	51.2	62	48.8				
Attended training	Yes	38	60.3	25	39.7	4.305	1	0.038*	0.147
	No	61	44.5	76	55.5				
Teaching experience	0–6	8	44.4	10	55.6	2.054	3	0.561	0.101
	7–10	22	59.5	15	40.5				
	11–15	32	45.7	38	54.3				
	> 15	37	49.3	38	50.7				

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$