Gatekeepers of Ambition: Test-Taker Preparation Practices on Two High-Stakes
Language Proficiency Tests

by

Ciaran McCullough

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial
fulfillment of the requirements for the degree of

Master of Arts

In

Applied Linguistics and Discourse Studies

Carleton University

Ottawa, Ontario

Abstract

This mixed-methods study with an exploratory convergent design (Creswell, 2009) investigated the test preparation practices, reasons for test choice, perceived performance, and reported emotional experience of 33 participant test takers on two high-stakes language proficiency tests. The findings indicated a variety of practices employed in taking the tests (IELTS and TOEFL iBT), including, in order of prevalence: seeking information from former test takers, using test preparation books/practice tests, and using online resources. Perceived performance and emotional experience during reading sections were rated most negatively. The two main concerns the participant test takers had were the speeded nature of the reading and writing sections of the test, and test administration factors. These findings shed light on the mechanism of *washback* (i.e. the effect of high-stakes tests on teaching and learning). Both test preparation practices and test takers' opinions of tests have implications for the validity of inferences drawn from tests.

Dedication

To my parents, and to Kaitlyn.

Ammor magnus doctor est.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Janna Fox, for her wisdom, guidance, patience, and understanding at every stage of the writing of this thesis; she is truly an inspiration to me. Special appreciation is extended to my thesis committee members, Dr. Natasha Artemeva, Dr. Angel Arias, and Dr. Jaffer Sheyholislami for their expert feedback, advice, encouragement, kindness, and humour.

I wish to extend my appreciation to my colleagues at the Elsie MacGill Learning Centre and the ESLA unit at Carleton University, and to my contemporaries and professors in the School of Linguistics and Language Studies; it has been a pleasure and a privilege to work with such brilliant, stimulating, and scholarly people. I also extend special thanks to Ms. Peggy Hartwick, Mr. Don Myles, Mr. Mike Murphy, and to the students who participated in this study.

I would especially like to thank my parents, Peter and Bernadette McCullough, who instilled in me a love of learning from a very early age and who have encouraged me to always strive to be the best that I can be; my partner, Kaitlyn, for her endless support, encouragement, and love; my sister, Cara, who was always there with an encouraging word; her husband Lee, and my dear niece Ella Rose. Special thanks are reserved for Mr. Robert and Mrs. Caroline Poisson, who have been supportive and encouraging from the very beginning of this journey.

Finally, I wish to thank Taiyo, my faithful friend, and Stephen Malkmus and the Jicks, for providing most of the soundtrack to this endeavor.

Table of Contents

Abstract.....	ii
Dedication	iii
Acknowledgements	iv
List of Tables	viii
List of Figures.....	ix
Acronyms and Abbreviations	x
Chapter One: Introduction	1
Chapter Two: Literature Review	5
2.1 High-Stakes Testing	5
2.2 Two High-Stakes Language Proficiency Tests: TOEFL and IELTS	7
2.2.1 <i>The Test of English as a Foreign Language</i>	8
2.2.2 <i>The International English Language Testing System</i>	12
2.3 Validity	15
2.4 Washback	25
2.5 Washback and Test Preparation	30
Chapter Three: Methods.....	39
3.1 Research Design	39
3.2 Setting.....	40
3.3 Participants	41
3.4 Data Collection.....	43
3.5 Instruments	45
3.5.1 <i>Questionnaire</i>	45
3.5.2 <i>Interview</i>	48
3.6 Analysis	49
3.6.1 <i>Questionnaire Quantitative Data</i>	49
3.6.2 <i>Questionnaire Qualitative Data</i>	49
3.6.3 <i>Interview Qualitative Data</i>	49
3.6.4 <i>Qualitative Coding</i>	50
Chapter Four: Findings and Discussion	54
4.1. Questionnaire Respondent Results	54
4.2 Interview Participant Profiles	56

4.3	Reported Preparation Practices	58
4.3.1	<i>Questionnaire Respondents' Reported Preparation Practices</i>	58
4.3.2	<i>Interview Participants' Reported Preparation Practices</i>	60
4.4	Reasons for Choice of Test	76
4.4.1	<i>Questionnaire Respondents' Reported Reasons for Choice of Test</i>	76
4.4.2	<i>Interview Participants' Reported Reasons for Choice of Test</i>	78
4.4.3	<i>Research Question 1: Summary of Findings</i>	81
4.5	Perceived Effect of Preparation Practices on Test Performance	83
4.5.1	<i>Questionnaire Respondents' Perceived Effect of Preparation Practices on Test Performance</i>	83
4.5.2	<i>Interview Participants' Perceived Test Performance</i>	86
4.6	Subjective Emotional Experience During Test	94
4.6.1	<i>Questionnaire Respondents' Reported Subjective Experience During Specific Test Sections</i>	94
4.6.2	<i>Interview Participants' Reported Subjective Experience During Specific Test Sections</i>	96
4.6.3	<i>Research Question 2: Summary of Findings</i>	100
4.7	Opinions of Test	101
4.7.1	<i>Participants' Opinions of Test</i>	102
4.7.2	<i>Research Question 3: Summary of Findings</i>	108
	Chapter Five: Conclusion	110
5.1	Summary and Implications of Findings	110
5.3	Limitations of the Study	113
5.4	Future Directions	113
	References	114
	Appendix A: Ethics Approval Certificate	121
	Appendix B: Email Request to ESLA Teachers	122
	Appendix C: ESLA Class Visit Script	123
	Appendix D1: TOEFL Test Taker Questionnaire	125
	Appendix D2: IELTS Test Taker Questionnaire	132
	Appendix F: Email Response to ESLA Student Interview Volunteers	141
	Appendix G: Participant Pseudonym Options	142
	Appendix H: Semi-structured Interview Protocol	143

Appendix I: Interview Consent Form	145
Appendix J: Qualitative Coding Scheme	148
Appendix K: Primary Researcher Initial Coding Sample	151
Appendix L: Second Coder Initial Coding Sample	152

List of Tables

Table 1: <i>Messick's (1996) Facets of Validity as a Progressive Matrix</i>	17
Table 2: <i>Overall Characteristics of Questionnaire Respondents</i>	42
Table 3: <i>Characteristics of Interview Participants</i>	43
Table 4: <i>Test Preparation Practices by Source</i>	44
Table 5: <i>Interview Dates and Durations</i>	45
Table 6: <i>Questionnaire Respondent Information Breakdown by Test Taken</i>	55
Table 7: <i>Summary of Prevalence of Preparation Practice Use as a Percentage of Questionnaire Respondents Who Employed This Practice</i>	59
Table 8: <i>Interview Participants' Reported Test Preparation Practices (Questionnaire Responses)</i>	61
Table 9: <i>Questionnaire Respondents' Reasons for Test Choice</i>	77
Table 10: <i>Interview Participants' Reasons for Test Choice (Questionnaire Responses)</i> ..	78
Table 11: <i>Summary of Questionnaire Respondents' Perceived Effect on Test Performance of Specific Preparation Practices</i>	84
Table 12: <i>All Test Takers' Reported Performance Per Test Section (Number of Respondents Coded At Least Once in Each Category)</i>	86
Table 13: <i>Numbers of All Respondents Reporting Specific Categories of Subjective Emotional Experience During Different Sections of Test (Number of Respondents Coded At Least Once in Each Category)</i>	94
Table 14: <i>Ranking of Total Incidence of Emotion Words Appearing Across All Questionnaire Respondents' (IELTS and TOEFL) Qualitative Responses</i>	95

List of Figures

Figure 1: <i>Bailey's (1996) Washback Model</i>	28
Figure 2: <i>Convergent Mixed-Methods Research Design (Creswell, 2009)</i>	40

Acronyms and Abbreviations

CAEL	Canadian Academic English Language Assessment
EAP	English for Academic Purposes
ESLA	English as a Second Language for Academic Purposes
ETS	Educational Testing Service
GRE	Graduate Record Examinations
IELTS	International English Language Testing System
L1	First Language
L2	Second Language
SAT	Scholastic Aptitude Test
TOEFL (pBT/iBT)	Test of English as a Foreign Language (Paper-Based Test/Internet-Based Test)
TOEIC	Test of English for International Communication

Chapter One: Introduction

Across the world, large-scale *language proficiency tests* act as gatekeepers to millions of people who wish to study or work in, or immigrate to, countries where the dominant language of communication is different from their own first language (Chalhoub-Deville & Turner, 2000; Green, 2005; Ma, 2017; Merrifield, 2008; Micken & Motteram, 2006; Stoneman, 2006; Wall, 2000). The central characteristic of language proficiency tests is that they purport to assess “knowledge, competence, or ability in the use of a language” (Bachman, 1990, p.16), although how these terms are applied has been, and continues to be, subject to debate (Shin, 2013). Examples of large-scale language proficiency tests include the Test of English as a Foreign Language (TOEFL), International English Language Testing System (IELTS), Canadian Academic English Language (CAEL) assessment, and Test of English for International Communication (TOEIC). In the language of assessment such tests may be referred to as *high-stakes* in nature because they are critical events in the lives of the people who take them, because scores derived from performance on these tests are used to make decisions that have profound ramifications for their career and life trajectories (McNamara, 2000). Given their importance to test-takers, test score users, and assorted other stakeholders, it is perhaps unsurprising that a significant global industry based on preparation for these tests has arisen. Although numerous studies have investigated the effects of high-stakes tests on curriculum and teaching and learning at the classroom level, DeLuca, Cheng, Fox, Doe, and Lee (2013) note that “few studies have been conducted on test-taker perspectives” (p.664).

As a language teacher who has taught in several countries and been involved in the delivery of preparation courses for some of the more well-known high-stakes

proficiency tests, I am particularly interested in how these tests influence teaching and learning. This phenomenon, referred to as *washback* (or *backwash*) (Alderson & Wall, 1993), has been the subject of a considerable body of research over the previous three decades. Although numerous studies have been conducted into the effects of high-stakes language proficiency tests on classroom teaching and learning, fewer have addressed test preparation from the perspective of how test takers themselves choose the preparation practices they employ prior to taking the test (see, for example, Doe & Fox, 2011; Fox & Cheng, 2015; Ma, 2017). In this study, the term *test preparation practices* refers to a broad spectrum of activity, including but not limited to taking instructed test preparation and general English courses, using official and unofficial preparation materials, using free and paid-for online resources, and seeking knowledge about the test from friends and acquaintances.

The rationale for conducting this study is threefold. First, the independent preparation practices of test takers prior to taking high-stakes tests represents an area of interest because this appears to have been under-researched in comparison to classroom-based preparation courses (Fox & Cheng, 2015), and since test takers are the group most directly affected by the outcomes of high-stakes tests it would seem obvious that their perspectives should be considered. Second, test takers' preparation practices may have significance for the validity of test scores, based on whether they address the underlying constructs purportedly assessed by the test. Third, understanding aspects of the test experience from test-takers perspective such as reasons for choice of test, opinions about the test, and subjective experience during the test may increase understanding of the testing process and assist in developing tests that are as fair, valid, and equitable as possible.

In this study, I investigated the reported test preparation practices of students in an English as a Second Language for Academic Purposes (ESLA) unit of a Canadian university who recently sat one of two globally-administered high-stakes tests; namely the IELTS and the TOEFL iBT. In the Canadian context, results from these two tests (in addition to the CAEL Assessment) are widely used to make admissions decisions about prospective students from overseas whose first language is not English; Canadian francophone students; dual Canadian citizens and permanent residents who have studied in non-English speaking contexts; and graduates of English-medium offshore schools or Canadian high schools with less than three years of residency. As part of their application requirements, participants in this study would have been asked to provide test scores to the admissions department of the university.

The overarching aim of this study is to investigate the preparation practices that test takers employ prior to taking the TOEFL and IELTS tests, and the influence of these preparation practices on their accounts of test performance. This study was guided by the following research questions:

1. What reported test preparation practices do IELTS and TOEFL test takers employ prior to sitting these tests? What factors appear to influence test takers' choice of test?
2. How do these reported preparation practices influence their perceived performance on the tests? Do reported test preparation practices have any effect on test takers' reported subjective emotional experience on the test?
3. What opinions do test takers have of the tests, and what are the implications of these opinions with regards to validity?

To address the research questions, this thesis is divided into five chapters. Chapter one has described the background, rationale, aims and research questions central to this study. Chapter two presents background information on the IELTS and TOEFL tests and a review of the previous research literature on large-scale language proficiency tests, with a focus on the fundamental and interconnected concepts of *high-stakes testing*, *validity*, *washback*, and *test preparation*. Chapter three details the research methodology underpinning the study and describes the research setting, participants, instruments used to gather and analyze quantitative and qualitative data, data-collection procedures, and methods of analysis. Chapter four presents the findings and discussion of the study. Chapter five summarizes the findings of the study, addresses the limitations of the study, and suggests potential further avenues of study.

Chapter Two: Literature Review

This study is informed by four interconnected themes and concepts: (1) high-stakes language proficiency testing, (2) validity, (3) washback, and (4) test preparation.

Accordingly, these themes are each addressed in turn in this literature review.

2.1 High-Stakes Testing

All tests, language or otherwise, purport to assess some characteristic, ability, or skill in the test taker; however, some tests have more serious consequences than others. Tests on which little depends may be described as *low-stakes*, an example of which may be a classroom quiz that students are required to take but which has no impact on their final grades (Cole & Osterlind, 2008). Conversely, *high-stakes tests* act as gatekeepers, and are associated with significant repercussions, consequences, or outcomes; therefore, a test can be described as high-stakes depends on how significant the ramifications of success or failure are for the test taker. High-stakes tests are those where “a lot hinges on the determinations made in the language test, for example, where it is used to screen for admission to academic or work settings” (McNamara, 2000, p.48). In the broader testing domain, examples of high-stakes tests include “high school graduation tests, college undergraduate entrance tests, university graduate or professional degree entrance tests, professional licensing tests ... employment tests, and immigration and citizenship tests” (Kunnan, 2012, p.1), as scores from these are all used to make decisions which have significant impacts on test-takers’ life trajectories and opportunities. Within the narrower domain of language testing, large-scale, standardized *language proficiency tests* such as the

IELTS and TOEFL¹ are high-stakes in nature as they are used by decision-makers in university admission departments, government immigration departments, and professional licensing organizations to make decisions on whether or not individuals are admitted to programs of study, granted work visas or citizenship, or accredited by professional bodies (Chalhoub-Deville & Turner, 2000; Green, 2005; Ma, 2017; Merrifield, 2008; Mickan & Motteram, 2006; Stoneman, 2006; Wall, 2000).

High-stakes tests are not simply neutral instruments, and can be viewed from a critical perspective with regards to their social and political implications (Bachman & Palmer, 1996; Menken, 2007; Messick, 1996 Shohamy, 2001). Menken (2007) states that high stakes tests act as “a sorting mechanism in education for all students; immigrants as well as other minorities have historically been particularly vulnerable to high-stakes decisions made [based on] test scores” (p.403). Shohamy (2001) argues that high-stakes tests act as agents of power and control, and their legitimacy can be upheld or diminished over time depending on the character of the social matrix within which they are embedded. According to Messick (1996), the social implications of language testing and the use of test results are of paramount importance as contrary to being neutral instruments that simply assess ability, a central characteristic of high-stakes testing is a power balance heavily weighted towards the *test score user* [i.e. the person or organization that uses test scores to make decisions]. Messick (1989) states that

... although in some instances tests are intended to serve the needs
of the test taker (e.g., school testing with a diagnostic purpose) it is

¹ Since 2006 the TOEFL has been available in two versions: Paper-Based Test (pBT) and Internet-based Test (iBT). Although both are currently in use, the iBT is the most commonly taken test and therefore the subject of most recent study. Consequently, all subsequent references to the TOEFL in this study denote the iBT unless otherwise stated.

important to keep in mind that selection, placement, and achievement measures are primarily intended to serve the decision-making needs of the user institution. (p.9)

McNamara and Roever (2006) discuss the social dimension of language testing and state that “our conceptions of what it is that we are measuring and the things we prioritize in measurement, will reflect values, which we can assume will be social and cultural in origin, and that tests have real effects in the educational and social context in which they are used” (p.12). Although an in-depth review and analysis of the myriad ethical, social, political and ideological implications of high-stakes language proficiency tests is beyond the scope of this study, it must be borne in mind that these issues are significant, and that tests such as the IELTS and TOEFL should not automatically be accepted as being neutral, value-free instruments; furthermore, as we shall see, these issues are integral to current conceptions of *validity*.

2.2 Two High-Stakes Language Proficiency Tests: TOEFL and IELTS

The TOEFL and IELTS are two of the most globally dominant English language proficiency tests in existence today; scores derived from these tests are used for university admission purposes worldwide (Alderson, 2009; Merrifield, 2008). Although both the TOEFL and the academic version of the IELTS are used to assess the overall construct of academic language ability, Chalhoub-Deville and Turner (2000) note that they each have subtly different philosophical underpinnings. Chalhoub-Deville and Turner (2000) state that “the Cambridge exams [including IELTS] have been constructed more like an *achievement test* with strong links between the examination and teaching syllabi”, whereas “the hallmark of TOEFL...is

its psychometric qualities with a strong emphasis on reliability” (p.524). In summary, this means that the item types found on the IELTS are explicitly linked to those that may be encountered in instructional settings whereas the TOEFL is marketed as a proficiency test, existing in isolation from any instructional program and purporting to assess something essentially internal to the test taker (Chalhoub-Deville & Turner, 2000). Nonetheless, we may conclude that the IELTS and TOEFL tests both purport to assess the target construct of *academic English*, that is, English as it is used in the higher education context. The following two sections describe the structure and content of the TOEFL iBT, and the IELTS, respectively.

2.2.1 The Test of English as a Foreign Language

The TOEFL is a commercial, standardized test of English for academic purposes, and is intended to “measure the ability of non-native speakers of English to use and understand English as it is spoken, written and heard in academic settings” (Alderson, 2009, p. 621). The TOEFL is developed, owned, and administered by ETS (Education Testing Service) of Princeton, New Jersey, which also owns and administers several other high-stakes tests including the Scholastic Aptitude Test (SAT) and General Equivalency Diploma (GED). As of 2018, over 30 million people worldwide have taken the TOEFL, with the test currently administered at over 4300 test centres across the world (ETS, 2018). TOEFL scores are accepted by over 10,000 universities, colleges, agencies and other institutions in 130 countries (ETS, 2018). Test-takers pay between USD\$150 and \$300 to take the test, depending on their country of residence (ETS, 2018).

The origins of the TOEFL date back to the early 1960s. The first paper-based TOEFL test (the pBT) was introduced in 1964, and since then the TOEFL has been

repeatedly refined and developed, with its precursors, the Test of Spoken English (TSE) becoming available as an additional component in the 1970s, and the Test of Written English (TWE) developed and offered as an optional additional component in 1986 (Alderson, 2009). From 1998 to 2006, the TOEFL was available as a computer-based test (cBT), and from 2006 until the present has been available as both internet-based test (TOEFL iBT) and paper-based test (TOEFL pBT) with the format used depending on where the test is taken and the technology available (Alderson, 2009). Since the rollout of the latest version in 2006, the TOEFL has consisted of four main sections: reading, speaking, listening, and writing. Each section is scored out of 30 and these are combined to give an overall score out of 120.

The reading section is between 60 and 100 minutes in length and contains from three to five texts from university level textbooks on a variety of academic subjects. These texts may be expository, argumentative, or historical-narrative in nature (Alderson, 2009), and are each accompanied by twelve to fourteen questions designed to gauge test-takers comprehension and inferencing abilities. In the reading section, there are three main item types: four-option multiple choice, four-option sentence insertion, and ‘reading to learn’ items that have more than one correct option. ‘Reading to learn’ is further assessed by way of asking students to select important ideas from a passage from a list, and complete schematic tables with information communicated in the reading (Alderson, 2009).

The speaking section of the TOEFL is 20 minutes long and contains both independent and integrated tasks, of which there are two and four on the test respectively. For independent tasks, candidates are required to speak at length about a topic that is familiar to them. For integrated tasks, candidates are required to speak in

response to something they read and/or hear. During the integrated speaking tasks, candidates are exposed to reading passages between 75 and 100 words long, and listening passages between 60 and 120 seconds in length. Prior to responding, candidates are permitted up to 30 seconds to prepare a response, and are given up to 60 seconds in which to respond (Alderson, 2009). The speaking section is designed to assess candidates' ability to speak about topics with which they are familiar, and compare, contrast, synthesize, and summarize information from different sources spontaneously. Task responses are scored on four criteria: general, delivery, language use, and topic development. Notably, in the TOEFL iBT spoken responses are recorded by a microphone connected to the computer the candidate uses to sit the test; no human interlocutors are involved (Chapelle, Enright & Jamieson, 2008).

The listening section, which is designed to test candidates for comprehension, pragmatic awareness, and ability to synthesize information, lasts between 40 and 60 minutes and contains four to six academic-style lectures and two to three conversations. The lectures, some of which contain classroom discussion, are between three and five minutes long and contain between 500 and 800 words; conversations are each approximately three minutes long and contain between 12 and 25 conversational 'turns'. Test takers are expected to answer approximately six questions per lecture, and five questions per conversation. There are four main item types: four-option multiple choice (single answer), four-option multiple choice (multiple answer), item sequencing (putting information in the correct order), and matching of texts or objects to a table, graph, or chart (Alderson, 2009).

The writing section also contains both independent and integrated tasks; however, in this section there is only one task of each type. For the independent

writing question, candidates are required to write at length on opinions they hold or choices they have made or would make, supporting these points logically and coherently. The time limit to prepare and write responses to independent writing tasks is 30 minutes. For integrated writing tasks, candidates are given three minutes to read a text 230-300 words in length on an academic topic. The topic disappears from the screen after this time and candidates then listen to a recording of a speaker discussing the same topic as was the subject of the text, but introducing alternative or contradictory perspectives. When the listening has concluded, the previously-shown text is once again displayed on the screen. Candidates are then given 20 minutes to type a response in which they connect the main points covered in the text and those covered in the listening passage, and identify how these relate to each other (Alderson, 2009).

The TOEFL iBT has been subject to an extensive process of validation based on Kane's (2006) *interpretation and use argument* (IUA) model, a process documented in detail by Chapelle, Enright and Jamieson (2008). Furthermore, ETS funds and supports ongoing research into the TOEFL iBT and there exists a great deal of research into its validation, components, uses, and effects worldwide (Alderson, 2009; Alderson & Hamp-Lyons, 1996; Chapelle, Enright & Jamieson, 2008; DeLuca, Cheng, Fox, Doe & Li, 2013; Fox & Cheng, 2015; Kyle, Crossley & McNamara, 2016; Liu, 2014; Malone & Montee, 2014; Stricker & Attali, 2010; Wall & Horák, 2006; Ward & Liu, 1994; Yu, 2012).

2.2.2 The International English Language Testing System

The IELTS (International English Language Testing System) is another commercial, globally administered English proficiency test, and is a direct competitor to the TOEFL iBT. The IELTS is developed, owned, and administered jointly by the British Council, IDP Education, and Cambridge English language Assessment (Cambridge English Assessment, 2018). There are two versions of the IELTS: an ‘Academic’ version and a ‘General’ version, and each year over three million individuals sit some version of the IELTS (Cambridge Assessment English, 2018) at over 1,200 locations throughout the world (IELTS, 2017a). The academic version is intended for individuals who require proof of their ability to study in an environment where English is the language of instruction or who wish to take a language proficiency test score to gain professional accreditation, whereas the general version is intended for those who intend to migrate to an English-speaking country or who require a certain test score to gain entry to work, training, or secondary education in an English-speaking country; particularly Australia, Canada, or the UK (IELTS, 2017a). The speaking and listening task specifications are identical across both test versions; however, the content of the reading and writing task specifications differ across the two versions, with these sections reflecting academic content on the academic version of the IELTS, and general content on the general version (IELTS, 2017b). The academic IELTS is the version of interest in this study, and from this point all references to the IELTS will mean the academic version of the test.

Like the TOEFL iBT, the IELTS is divided into four main sections: reading, speaking, listening, and writing. The test takes two hours and 45 minutes to complete, with 30 minutes for listening, 60 minutes for reading, 60 minutes for writing, and 11-14 minutes for speaking (IELTS, 2017b). Performance is scored on a scale of from 0

(did not attempt the test) to 9 (expert user) at half-band increments (i.e., it is possible to achieve a score of e.g., 0.5, 1.5, 2.5, etc.). IELTS score reports detail both overall and individual section scores according to this 0-9 scale (IELTS, 2018).

As on the TOEFL, the reading section of the IELTS academic version is intended to assess academic reading ability. The IELTS reading section contains three texts which range “from the descriptive and factual to the discursive and analytical” (IELTS, 2018, p.3), and are selected based on their representation of academic and professional written material. Test takers are required to answer a total of 40 questions on the texts, and questions are intended to assess reading skills such as reading for detail, comprehending logical argument and opinion, reading for gist, detail, and main ideas, and understanding opinions, attitudes, and purpose (IELTS, 2018).

The speaking section of the IELTS consists of three sections, and lasts between 11 and 14 minutes. Significantly, the IELTS speaking section is conducted in person with a human interviewer, in contrast to the TOEFL speaking section during which test takers speak into a computer microphone without the presence of a human interlocutor. With regards to content, part one of the IELTS speaking section mirrors the independent speaking section on the TOEFL, with the examiner asking questions about a broad range of topics that are familiar to the student; these may include work, studies, interests, and family (IELTS, 2018). In part two, candidates are given a card on which is printed a question prompting students to speak about a specific topic. Candidates are allowed one minute to prepare an answer and are required to speak about the specified topic for up to two minutes, with the examiner asking one or two follow-up questions about the same topic (IELTS, 2018). In part three of the IELTS speaking section, candidates are asked additional questions about the topic from part

two, encouraging deeper and more abstract discussion of the topic. Part three lasts from four to five minutes.

During the listening section of the IELTS, candidates hear four recordings of speech, after which they are required to answer questions relating to the content. The four recordings each have different characteristics: the first depicts a conversation between two people in an everyday social context; the second is a monologue relating to an everyday matter; the third is a conversation between a maximum of four people in an academic context; and the fourth is a monologue on an academic topic (IELTS, 2017b).

The writing section of the academic version of the IELTS, which lasts 60 minutes, contains tasks on topics that are “of general interest and suitable for test takers entering undergraduate and postgraduate studies or seeking professional registration” (IELTS, 2018, p.3). This section consists of two tasks: in Task one, candidates are required to describe, summarize, or explain a graph, table, chart, or diagram (IELTS, 2018). This task is intended to assess candidates’ ability to describe, understand, or explain data, a process, an event, or an object. For Task two, candidates are required to construct an essay “in response to a point of view, argument, or problem (IELTS, 2018, p.3). Responses must be written in a suitably academic style. Unlike with the TOEFL, there appears to be relatively little detailed information on exactly how the IELTS is validated, although IELTS does support research relating to the test through its IELTS Research Reports series (IELTS, 2018).

If tests such as the TOEFL and IELTS purport to assess test takers’ readiness to function in an academic environment where the language of instruction is English, how can we be sure that they achieve this? Brooks and Swain (2014) state that “the

use of such proficiency tests (e.g., TOEFL iBT) as evidence of language facility for admission decisions is based on the claim that scores on the tests can predict test takers' linguistic readiness (or not) to cope with the language demands of academic studies" (Swain, 2014, p.354). Thus, these tests must be subjected to an ongoing process of *validation*.

2.3 Validity

Tests are used to make defensible decisions based on *inferences*, or logical conclusions drawn from test takers' performance on certain measures, and these inferences must be underpinned by reason, logic, and evidence if a test is to be considered *valid* (Messick, 1989). However, validity is a highly complex topic, and understanding of it continues to develop (Cizek, 2012; Cronbach & Meehl, 1955; Kane, 2013; McNamara & Roever, 2006; Messick, 1989). The following section represents a selected historical account of validity theory with regards to high-stakes language tests, and its implications with regards to the topic of this study.

An influential contribution to current conceptualizations of the nature of validity in testing was made by Cronbach and Meehl (1955), who established a basis for validating psychological tests and made the point that evidence should be sought for test interpretation, refuting validity claims in early arguments for the validity of the TOEFL that were advanced mainly based on a simple reliability coefficient. Cronbach and Meehl (1955) proposed that validity can be viewed across four dimensions: *predictive*, *concurrent*, *content*, and *construct validity*. Predictive and concurrent validity can both be classified under the larger category of *Criterion-oriented* measures, and are concerned with whether a test addresses certain underlying abilities; for example, grammar knowledge in the case of language tests. The

predictive validity component refers to how well a test predicts performance in some specified domain; for example, if a candidate performs well on a test of academic writing, this should predict their performance on other academic writing tasks external to the test. *Concurrent validity* refers to the correspondence on scores on different tests purporting to measure the same ability, for example, substituting a cloze task for an open-ended writing task when assessing grammar usage. *Content validity* refers to how well the test items represent the target domain; in the case of the TOEFL and IELTS Academic tests, the stated target domain is ‘academic English’ (although whether such a domain exists may be open to question). *Construct validity* applies “whenever a test is to be interpreted as a measure of some attribute or quality which is not *operationally defined*, [that is], whenever no criterion or universe of content is accepted as entirely adequate” (Cronbach & Meehl, 1955, p.282). In the context of language testing, no test will ever be able to represent the entirety of content implied by a domain classification, for example, ‘academic English’, therefore construct validity is a central concern. Following Cronbach and Meehl (1955), understanding of validity underwent a further advance in the 1990s, spearheaded by Samuel Messick.

Messick (1989) represents a significant and deeply influential advance in the understanding of validity. According to McNamara (2009), the two main ways Messick developed the understanding of validity were by “proposing a new understanding of how inferences made based on tests must be challenged, and in drawing attention to the consequences of test use” (p.31). In addition to the strands of validity described by Cronbach and Meehl (1955), Messick (1989) argued that any model of validity should be extended beyond empirical questions of concurrent, predictive, content, and construct validity to include consideration of the “appropriateness, meaningfulness, and usefulness of score-based inferences”

(Messick, 1989, p.5). This development elevated validity from being simply a set of measures based on the test itself, and placed the inferences based on these measures, and the *consequences* of the test and test score use at the centre of validity; this final addition to the validity lexicon has come to be referred to as *consequential validity*. Messick's (1989) conception of the dimensions of validity are represented by a progressive matrix, as shown in Table 1.

Table 1: *Messick's (1996) Facets of Validity as a Progressive Matrix (p.10)*

	<i>Test Interpretation</i>	<i>Test Use</i>
Evidential Basis	Construct validity (CV)	Construct validity + Relevance / utility (R/U)
Consequential Basis	CV + Value implications (VI)	CV + R/U + VI + Social consequences

In Messick's (1989) progressive matrix model, the columns are labelled *test interpretation* (left column) and *test use* (right column), and the rows are labelled *evidential basis* (top row) and *consequential basis* (bottom row). From top-left to bottom-right, the four cells represent (a) *construct validity*, (b) *evidential basis of test use*, (c) *consequential basis of test interpretation*, and (d) *consequential basis of test use* (Cizek, 2012). According to Messick's (1989) unitary model, construct validity, at the intersection of *evidential basis* and *test interpretation* (cell a), represents the basis of validity. As the construct is impossible to observe under test-conditions, tests must somehow model this real-world behaviour to make inferences about a test-taker's ability that are based on convincing evidence. Cell (b), at the intersection of *evidential basis* and *test use*, indicates that the *uses* of tests must also be underpinned by

evidence. Cell (c), at the intersection of *consequential basis* and *test interpretation*, “highlights the proposition that value considerations underlie score meaning and are inherent at every stage of test development, administration, and reporting, beginning with the very conceptualization of a test” (Cizek, 2012, p.33). Cell (d), at the intersection of *consequential basis* and *test use*, represents consideration of the broader social implications of test use; or according to Messick (1989), “the functional worth of scores in terms of social consequences and their use” (p.84).

According to Messick’s (1989) progressive matrix model of validity, *construct validity* underpins all other facets of validity as these exist in a nested arrangement. Although Messick’s (1989) model has been highly influential, Cizek (2012) argues that this unitary conceptualization is flawed. According to Cizek (2012), incorporating *test interpretation* and *test use* into one unitary model of validity entails accepting a definition of validity “where one aspect (i.e., accuracy of score inferences) is incompatible with the other (i.e., the justification for actions taken based on test results” (p.34). To illustrate this point, Cizek (2012) uses the analogy of a blood test which could predict, with a high degree of accuracy, whether an individual would in the future develop an illness which would result in a slow and painful death. In this analogy, the validity of the *score inference* (whether the test accurately detects the presence of the latent disease) is not connected to the appropriateness or otherwise of the test’s *use*; that is, the score validity of the test may be substantially supported by evidence, even if its deployment may have overall negative consequences such as the subject of the test potentially experiencing severe mental distress and therefore rendered invalid with regards to Messick’s (1989) consequential validity facet. Cizek (2012) claims that Messick’s (1989) linking of test interpretation and test use in a unitary model of validity should be reconsidered, and argues in favour of a “a crisp

conceptual distinction between two fundamental concerns—the *validity of a test score inference* and the *justification for a test's use*” (p.37).

In a later paper developing his model of validity, Messick (1996) reiterated that validity is “not a property of the test or assessment as such, but rather of the *meaning* of the test scores” (p. 245), and introduced a set of six aspects that may be considered components of an overarching concept of validity. These are: the *content* aspect (determining the boundaries of the construct domain, i.e. the skills, knowledge etc. to be assessed); the *substantive* aspect (the appropriateness of sampling for the target domain); the *structural* aspect (the relation of the internal structure of the assessment to the internal structure of the construct domain); the *generalizability* aspect (how representative the assessment content is of the content and processes of the construct domain); the *external* aspect (whether the assessment constructs “rationally account for the external pattern of correlations” (p. 251); and the *consequential* aspect (the consequences of test score interpretation in personal, social, political and economic terms).

Bachman (2007) argues that in the context of language tests, the construct may be defined in three general ways: ability-focused, task-focused, and interaction-focused. Bachman (2007) charts the history of approaches to defining the construct in language testing, and states that since the early 1960s, there has existed a dialectic between “a focus on *language ability* [or *trait*] as the construct of interest and a focus on *task* or *context* as the construct” (p.43). The ability/trait approach broadly corresponds to a Chomskyan, ‘competence-based’ model of language ability as something which is internal to the individual, that is, the abilities or traits that they possess. On the other hand, the task/context-based approach situates the construct as

being concerned with the kinds of tasks learners can perform, and may, therefore, be viewed as more constructivist in character (Creswell, 2009). These two strands are reconciled by Chapelle (as cited in Bachman, 2007), who argues for an *interactionalist* approach to construct definition which views ability and task-based definitions as reconcilable. In the interactionalist view, Chapelle (as cited in Bachman, 2007) argues that the interactionalist approach “must specify relevant aspects of both trait and context” (p.58), but that this is not a simple summation of the two; rather, “trait components [cannot] be defined in context-independent, absolute terms, and contextual features cannot be defined without reference to their impact on underlying characteristics [of language users or test takers]” (p.58).

McNamara and Roever (2006) propose that test score validation may be imagined as consisting of a sequential *chain of inferences*. These are: eliciting student performance (task design), scoring the performances (scoring procedures), ascertaining how typical the score is (reliability, generalizability), what the score means (interpretation), and using the score to make decisions (consequences) (McNamara & Roever, 2006). Kane (2013) also subscribes to the *chain of inferences* model of validation with regards to the *interpretation and use* of test scores.

Kane (2013) describes validation with regards to tests as “an evaluation of the coherence and completeness of an *interpretation and use argument* [IUA] and of the plausibility of its inferences and assumptions” (p.1). Kane (2013) states that “...to validate an interpretation or use of measurements is to evaluate the rationale, or argument, for the proposed conclusions or decisions” (p.64), and that “...validity is not a property of the test. Rather, it is a property of the proposed interpretations and uses of the test scores” (Kane, 2013, p. 3). Kane (2013) notes that tests cannot be

viewed as neutral instruments that simply assess some set of attributes in isolation, but as instruments that produce results that are then used for specific purposes, for example, in the case of high-stakes language tests such as the TOEFL test, granting admission to a course of higher study. Therefore, the interpretation and use of test results must be considered when discussing validity.

According to Kane's (2013) model, the process of validation proceeds by making the case for inferences made at each step of the testing process; these can be viewed as a series of 'bridges' which must be crossed for a test to be considered valid. The first of these is the *scoring inference*, which refers to the conversion of an observation (i.e. performance on a test component) into a valid score. The second is the *generalization inference*, which refers to how well the scores reflect general performance in a test setting. The third is the *extrapolation inference*, which refers to the use of scores to predict extra-test performance in the target domain. The fourth inference concerns *implications of test use*, which corresponds to Messick's (1989) concept of *consequential validity*, or Cheng's (2007) concept of *impact*. Kane's (2006) IUA approach has been influential, with the most recent version of the TOEFL iBT undergoing a rigorous and extensive process of validation by Chapelle, Enright, and Jamieson (2008).

Under Kane's (2013) argument-based approach to validation, "interpretations can change over time as uses change and as new evidence becomes available" (Kane, 2013, p. 45); when interpretations and uses change, new evidence may be required to validate these. According to Kane (2013), the more ambitious the claim being made by the assessment the greater the requirement for evidential support for these claims; furthermore, ambitious claims, while more useful than less-ambitious claims, are more

difficult to validate. Significantly, Kane (2013) argues that “negative consequences can render a score use unacceptable” (p.1), and the validation of a score interpretation has no automatic bearing on the validation of a score use. For example, an instance of potentially invalid score use would be the use of TOEFL scores (an assessment of academic language proficiency) by an employer to select candidates for a position that did not require the use of academic English. Kane (2013) integrates the empirical aspect of test validation with the consequential aspect, inextricably incorporating this into the validation network. In summary, Kane (2013) states that “...to validate an interpretation or use of measurements is to evaluate the rationale, or argument, for the proposed conclusions or decisions” (p.64). Michael Kane’s work, dating back to the early 1990s and continuing today, “is seen as an extension of a long tradition and history of validity in educational and psychological measurement... [and] draws on argumentation theory and logical reasoning” (Dr. A. Arias, personal communication, September 11, 2018).

With regards to the consequential aspect of validity, Chalhoub-Deville (2016) and Zumbo and Hubley (2016) reiterate Shohamy’s (2001) point that testing exists within a social and political matrix, and that test scores, and more importantly the use of test scores, extend beyond the individual level and into the social and political realms, which they argue has implications for validity. McNamara (2009) distinguishes two views vis-a-vis the social and political role of tests.

One holds that language-testing practice can be made ethical and stresses the individual responsibility of testers to ensure that it is. The other sees tests as essentially sociopolitical constructs, which, because [they are] designed as instruments of power and control

[and] must therefore be subjected to the same kind of critique as are all other political structures in society (McNamara, 2009, p.43).

In line with this view, Kane (2013) agrees that tests cannot be viewed as neutral instruments that simply assess some set of attributes in isolation, but as an instrument that produces results that are used for a specific purpose; for example, in the case of high-stakes language tests such as the TOEFL test, filtering candidates for admission to a course of higher study. Therefore, it follows that the *use* of the test results must be carefully considered when discussing validity.

Adopting a strong task/context perspective to construct definition, McNamara (2009) argues that the idea of a psychometrically-defined language ability as an ‘inner essence’ may be misguided and instead adopts a perspective informed by the ideas of Judith Butler, viewing language ability as *performative*, as opposed to being a representation of something internal. Butler (1993, as cited in McNamara, 2009) posits a difference between *performativity*, or the construction of a sense of something inner by certain acts, and *expression*, which is “the outward manifestation of something inner” (McNamara, 2009, p.41). This dichotomy corresponds to a long-running debate within the language testing community (Bachman, 2007), which likely has its roots in more fundamental ideological and ontological arguments.

Messick’s (1989) consequential aspect of validity is considered by Chalhoub-Deville (2016), who argues that “the social dimension to validity” (p. 454) has been neglected. Chalhoub-Deville (2016) references both Messick’s approach and Kane’s IUA (interpretation / use argument) approaches to validity, but goes further than either in arguing that any consideration of validity must account for “the interconnections of policy mandates, testing research, and societal consequences” (p. 454). Furthermore,

Chalhoub-Deville (2016) situates language testing in the context of a ‘global education reform movement’ (GERM), and calls for “a social orientation to validity research” (p. 460). Essentially, Chalhoub-Deville (2016) suggests that exclusively construct-related measures of validity neglect the socio-political aspect of the use of test scores, and encourages a ‘theory of action’ (TOA) approach as a means of accommodating “consequential investigations at the aggregate and system level” (p. 463).

Messick (1996) discusses the concepts of authenticity and directness in relation to testing, with authenticity referring to the closeness of the test content to the real-world tasks represented by the test content, and directness referring to the desirable situation wherein the test content is not artificially restricted by the test format. Further threats to validity described by Messick (1996) include *construct under-representation*, whereby the focal testing constructs are too narrow in relation to the target tasks or skills, and *construct-irrelevant variance*, a problem caused by assessments being too broad, and “containing excess reliable variance that is irrelevant to the interpreted constructs” (p. 244). Construct-irrelevant variance may arise due to (for example) the format of the test, such as when test-takers without the requisite computer skills are asked to take a computer-based test (Fox & Cheng, 2015). In such an example, the assessment of the construct (language proficiency) is undermined by difficulties that test-takers may have with the format of the test, and therefore the resulting test scores cannot be accepted as a valid measure of the construct.

In the current study, the topic of interest is the preparation practices test takers employ prior to taking the TOEFL and IELTS high-stakes language proficiency tests,

and their effect on test takers' accounts of performance and subjective experiences during these tests. So far in this section, we have discussed foundational notions of validity, and their application to high-stakes proficiency tests. We will now turn to the concept of *washback*. As we shall see, this concept has complex implications for validity, and offers a framework through which we can examine the issues surrounding student preparation for high-stakes tests.

2.4 Washback

As described in the validity section of this literature review, following Messick (1989) and Kane (2013), the concept of validity may be considered to encompass the effects of a test on teaching and learning. This effect has been subject to a considerable body of research over the previous 30 years, a program that was arguably prompted by Alderson and Wall (1993), who postulated the existence of a phenomenon they and others label *washback* (or *backwash*). Washback, a concept which may be viewed as falling under the category of *consequential validity* (Messick, 1989), refers to the effects that the existence of a test has on the activities of those who are affected by the test (or *stakeholders*). Four main contributions to washback research are central to this study: Alderson and Wall (1993), Bailey (1996), Cheng, Watanabe, and Curtis (2004), and Green (2007). In this section, these contributions will be discussed, as well as other significant contributions to the washback literature with a focus on washback to the learner.

Alderson and Wall (1993) initiated a program of research on washback that continues to the present day (Alderson & Hamp-Lyons, 1996; Bailey, 1996; Cheng, 2005; Cheng, Watanabe, and Curtis, 2007; Green, 2013; Messick, 1996). In their milestone paper, Alderson and Wall (1993) considered the idea that tests may cause

“teachers and learners [to] do things they would not necessarily otherwise do because of a test” (p.117), and formulated 15 hypotheses regarding washback with the intention of inspiring future research. These are:

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

As a factor that could be argued to be contingent on consequential validity, washback may be considered as part of the broader category of *impact* (Cheng & Curtis, 2004). Impact refers to the ramifications of testing on stakeholders and society with implications as diverse as the political, social, and financial (Bachman & Palmer, 1996; Cheng, 2008; Hamp-Lyons, 1997). Bailey (1996) proposed that tests may have effects on teaching and learning that can be positive or negative, and these can differ for various stakeholder groups (Fig.1). Building on the work of Alderson and Wall (1993) and Hughes (as cited in Alderson & Wall, 1993), Bailey's (1996) model conceptualizes washback as acting in terms of *participants*, *processes*, and *products*. Participants refers to any group of that may be influenced by a test, for example students, teachers, materials designers and curriculum designers, and researchers. Processes refers to how the test influences the activity of the participants. In Bailey's (1996) model, products may include learning, teaching, new materials and curricula, and research results (see next page).

Figure 1: Bailey's (1996) Washback Model (p.294)²

Messick (1996) discusses the possibility of washback being *positive* or *negative*, and states that *positive washback* may occur when the test “includes authentic and direct samples of the communicative behaviours of listening, speaking, reading, and writing of the language being learnt” (p. 241). Conversely, *negative washback* refers to negative effects of tests on learner and teacher behaviour which may have little connection with the target construct. With regards to the connection between washback and validity, according to Messick (1996), washback can only be considered to have a bearing on validity if it can be evidentially shown “to be an effect of the test and not of other forces operative on the educational scene” (p. 242).

Numerous studies have been conducted into the washback effects of high-stakes tests at both the macro-systemic level and the micro school, classroom, and

² Included under the terms of the Copyright Act of Canada, Section 29 (Fair Dealing).

individual levels. Wall and Alderson's (1993) study informed Alderson and Wall's (1993) Washback hypotheses. Wall and Alderson (1993) investigated the influence of the redesigning of a high-stakes English examination in Sri Lanka that was intended as an agent of educational and curricular change, and based on their findings formulated guidelines for policymakers and educators who intend to implement such changes through the introduction of tests. This work informed Cheng (2005) who, in a similar study, examined the effects of an examination in Hong Kong and observed that this test in question affected stakeholders both at the macro-level, "including different parties or levels of stakeholders within the Hong Kong educational context, [and] at the micro-level, in terms of classroom teaching and learning, including aspects of [teachers' and learners'] attitudes" (Cheng, 2014, p.353).

Green (2013) defines washback as "the impact that a test has on the teaching and learning done in preparation for it" (p. 40). Green (2013) notes that a distinction may be drawn between the *extent* (Bachman & Palmer, 1996) or *intensity* (Cheng, 2005) of washback and its *direction*, which can be beneficial, damaging, or not impactful. Hughes (as cited in Green (2013) suggested that washback should only be anticipated where participants are (1) motivated to succeed on the test, (2) believe they know how to be successful and (3) believe they have sufficient resources to succeed. Green (2007) Proposes a model that conceptualizes three aspects of washback: *direction*, *variability*, and *intensity*. In terms of direction, the model proposes that an overlap between the focal construct and the test design characteristics implies the potential for positive washback. However, where the test design characteristics and focal construct do not overlap, negative washback may occur. *Variability* refers to the effect of participant characteristics and values (e.g., knowledge and understanding of the test and its demands, acceptance of these

demands, and resources available to cope with these demands (Green, 2007). *Intensity* refers to how strong the washback effect is likely to be, and is conceptualized by Green (2007) as being defined by two dimensions: perception of test importance, and perceptions of test difficulty. With regards to factors influencing washback intensity, Green's (2007) model predicts that if a test is perceived as important but easy, no washback will occur, and if a test is perceived as unimportant, little washback will occur, whether it is perceived as easy, challenging, or difficult. For washback to occur, a test must be perceived to be important and challenging, and not easy or unachievable (Green, 2007).

Informed by Green's (2007) washback model, we shall thus now turn our attention to the concept of *test preparation*, and the washback effects of the TOEFL and IELTS on test takers directly. We will consider types of test preparation, and the possible associated *direction*, *variability*, and *intensity* of washback in these instances.

2.5 Washback and Test Preparation

If washback encompasses the effect of tests on teaching and learning, what forms might this take? This is a complex question, but for the purposes of this study we will make a distinction between washback to teachers and washback to learners, although realistically this line may be somewhat blurred. Cheng (2008) refers to *consequences* of high stakes testing as being "both intended or unintended and positive or negative aspects of instruction [on] students, teachers, and the school (p.349). This framing mirrors Messick's (1982) consequential aspect of validity; therefore, we may connect washback to validity in this way. Aside from the effect of test on teaching and learning, consequences and impact can manifest at multiple levels, from the systematic effect across society and the education system (Wall &

Horák, 2008) to the individual financial burden borne by test takers (Alsagoafi, 2018). However, since the current study relates to test taker preparation practices, we shall focus on that specific aspect. Before looking at specific studies of test preparation from either the classroom or student perspective, we must first consider the meaning of *test preparation*.

Messick (1982) defines *test preparation* as “any intervention procedure specifically undertaken to improve test scores, whether by improving the skills measured by the test or by improving the skills for taking the test, or both” (p.70). Anastasi (1981) proposed three main categories of test preparation, depending on the nature of the instruction; these are *test-taking orientation*, *coaching*, and *training in applicable skills*. *Test-taking orientation* refers to familiarization with the test format, question types, and procedure; this is intended to reduce anxiety on the test. *Coaching* refers to the use of practice exercises analogous to the test items with the purpose of enhancing performance. *Training in applicable skills*, the most desirable mode of test preparation, aims to improve performance in the target construct (e.g., using general academic exercises to train for the TOEFL and IELTS, which have academic English as their construct).

Smith (1991) lists eight general types of test preparation that may be employed when preparing students to take high-stakes tests. These are:

- (a) ordinary curriculum with no special preparation,
- (b) teaching test-taking skills,
- (c) exhortation,
- (d) teaching content known to be covered by the test,

- (e) teaching to the test in format and content,
- (f) stress inoculation,
- (g) practicing test or parallel test items, and
- (h) cheating (p.521).

According to Smith (1991), *an ordinary curriculum* reflects little to no orientation towards the test. *Teaching test-taking skills* involves orienting test takers towards specific skills that are applicable to the test, for example, strategies that may be used on certain types of item. *Exhortation* refers to general encouragement through general motivational techniques. *Teaching content* refers to incorporating content into the curriculum that is known or predicted to be on the test. *Teaching to the test in format and content* involves familiarizing test takers with the structure of the test and how items are presented, as well as the type of content that is predicted to be found on the test. *Stress inoculation* refers to the teaching of techniques that test takers may use to reduce stress during the test. *Practicing test or parallel test items* involves practicing the exact types of questions that may be found on the test. *Cheating*, according to Smith (1991), may be considered to include “any test preparation practice that [artificially] inflates scores” (p.537). This final category is particularly complex, as various stakeholders may have different ideas with regards to what constitutes cheating and therefore the line between acceptable and unacceptable test preparation practices may blur, depending on how one defines cheating.

The use of test preparation is not uncontroversial. The literature appears to be divided between those who take an ambivalent-at-best view of the effectiveness or ethics (or both) of focused preparation for high-stakes tests (Green, 2005; Hamp-

Lyons, 1999; Liu, 2014; Zhengdong, 2009) and those who are more optimistic about various types of test preparation (Brown, 1998; Wadden & Hilke, 1999). Somewhat pessimistically, Messick (1996) noted that “if it occurs, washback is likely to be orientated towards the achievement of high test scores as opposed to the achievement of facile domain skills” (p 245) which may lead to the possibility that certain preparation practices may contribute to the artificial raising of scores by improving performance on the test independent of performance in the construct domain; this is referred to as *score inflation*. Despite considerable debate in the literature, it seems to be generally accepted that activities geared towards improving skills in the construct domain, as opposed to the test domain, are desirable (Ma, 2017; Messick, 1996). Conversely, Messick (1996) suggests that the validated proficiency tests “can be subverted by test preparation practices or coaching emphasizing *testwiseness* strategies that might increase test scores without correspondingly improving the skills measured by the test” (p.246); therefore, the coached test performance may invalidate the interpretation and use of the coached scores.

What teachers do in the classroom when their instruction is oriented towards a test has been subject to a significant amount of research since Alderson and Wall (1993). Previous studies have been conducted into the effects of high-stakes tests on classroom instruction, including both for IELTS (Brown, 1998; Elder & O’Loughlin, 2003; Green, 2007; Zhengdong, 2009) and TOEFL (Fox & Cheng, 2015; Gebril, 2018; Green, 2005; Kim, 2007; Roberts, 2002; Smith, 1991). With regards to taught courses, evidence that test-oriented preparation courses influence test scores is mixed. Zhengdong (2009) investigated the effect of IELTS test preparation courses on IELTS test performance, and found no evidence that such preparation courses influence IELTS test scores. Green (2007) investigated the effect of dedicated test preparation

classes on performance on the writing section of the IELTS and found that when compared with general academic language writing classes and combined test preparation and general academic writing classes, dedicated IELTS writing preparation classes were not associated with any clear performance advantage. However, Brown (1998), in a study comparing the pre-test and post-test scores of students taking instructed EAP and IELTS preparation courses, found that students who received specific IELTS-focused preparation instruction received higher marks on the IELTS than those who took general EAP courses.

Despite the variability in the literature regarding the effectiveness of preparation courses, some researchers report that such courses may confer an advantage in that test takers may learn strategic behaviours and knowledge that are advantageous on the test, and may also extend to the wider construct domain. Issitt (2008) investigated three types of strategies that may be employed by test takers. The three strategies are aimed at *psychological preparation*, *general academic English awareness*, and *test awareness* based on published IELTS marking criteria. Issitt (2008) reports that based on speaking scores in the IELTS, there is some evidence to suggest that these three types of strategies are useful in preparing students for the speaking section.

Fox and Cheng (2015) note that although test-takers are the group most affected by high-stakes proficiency testing, their role as “principal stakeholders in language testing has not always been recognised” (p. 66). Fox and Cheng (2015) suggest that performance by test-takers may differ in terms of test-taking strategies, behaviours and perceptions, prior knowledge, anxiety, and motivation. Mikan and Motteram (2009) investigated the test preparation activities employed by candidates

prior to taking the IELTS test, with the aim of formulating “a description of candidate experiences and practices outside of language programs” (p.5). Mician and Motteram (2009) observed that many subjects reported using practice materials (particularly published tests) to prepare. In their study, Mician and Motteram (2009) report that most candidates had prepared alone (although some reported shared preparation practices including “studying and talking together, doing practice tests, and sharing resources” (p.10). Overall, out of 78 respondents, only six reported having attended English lessons while preparing for the IELTS. Of the total number, 37 had studied or practiced for the test, 13 had taken practice tests, 8 had used internet or library resources, and five did not prepare at all. Regarding this last figure, Mician and Motteram (2009) reported that “a number of candidates had repeatedly taken the IELTS test with some admitting to not knowing how to improve their scores” (p.1). Of the 78 respondents, 41 reported having used an IELTS preparation textbook, 54 sat practice tests (one respondent reporting having taken 100 such tests).

Ma (2017) researched the preparation techniques of Chinese students for high-stakes English language tests with a focus on the “nature, perceived effects, and perceived value of test preparation practices for the TOEFL iBT and the College English Test Band 4 (CET4)” (p.ii). Ma (2017) reported that overall, three general practices that helped Chinese students’ perceived success on the test were *test familiarization*, *stress management*, and *English skill improvement*; however, these factors were influenced by candidates’ individual characteristics, perceptions, and situated context, in concurrence with Green’s (2013) points regarding *variability* of washback.

With regards to the types of skills test takers may gain from their preparation practices, some are in line with the stated construct of language proficiency, whereas others may threaten the validity of the test. Mickan and Motteram (2009) and Doe and Fox (2011) both draw a distinction between *language-use strategies*, which promote and assist engagement with language itself, and *testwiseness strategies*, which relate to techniques and methods that are specific to the form of the test in question. Gebril (2018) connects *teaching to the test*, which “refers to those activities that merely focus on test content and ignore real learning” (p.5) with *testwiseness*, which can be defined as “skills used by test takers to attempt to answer a question without mastering the implied knowledge, [which may promote] skills that may help students answer test questions without adequate understanding of the target content or skills” (p.6); this in turn may contribute to *score pollution*, or test scores that are corrupted with regard to construct validity. Cohen (2007) argues that awareness of test taking strategies is central to considerations of validity, as it allows a distinction to be drawn between language-based learner strategies and test-taking (*test management* and *testwiseness*) strategies.

Gosa, as cited in Xie and Andrews (2012), found that student’s views on teaching and learning activities were influenced most strongly by the students’ expectation of assessment. Xie and Andrews’ (2012) study extends understanding of this, and found that candidates’ perceptions of both test design and test use influenced their preparation practices. Fox and Cheng (2015) investigated stakeholder accounts of taking the TOEFL iBT, including test-preparation practices employed by test takers prior to the test. In their study, they report that common test preparation practices included looking at online practice tests and preparation guides, talking to friends, and taking preparation courses. Doe and Fox (2011) state that test preparation may also

take other forms, including “studying alone with preparation books as guides, or seeking explicit or implicit test-preparation tips from an English for Academic Purposes (EAP) course” (p.30).

Ma (2017) notes that “the threat that test preparation poses to validity has been categorized into two main aspects: *score inflation* and *construct-irrelevant variance*” (p.42). *Construct-irrelevant variance* refers to other factors influencing test performance that are external to the testing of the construct. An example of construct irrelevant variance is illustrated in Fox and Cheng (2007) who describe a case where in a test of writing, test-takers may have been prevented from performing the task not because they could not write, but because they did not understand a word used in the question, therefore negatively affecting the construct validity of the test. Fox and Cheng (2015) report that for the TOEFL, another possible source of construct irrelevant variance is the computer-based nature of the test, in that computer ability is not a stated construct of the test but a certain level of computer skills is required to navigate the test interface and complete test items. Therefore, test takers who are not familiar with computers may achieve misleading scores as this factor may have prevented them from fully representing their language skills.

As specified at the beginning of this chapter, this study is guided by the overarching and intertwined concepts of high-stakes language proficiency testing, validity, washback, and test preparation practices. The literature described to this point provides a background information on these topics from which to proceed with the research questions central to this study. To reiterate, the main research questions guiding this study are:

1. What reported test preparation practices do IELTS and TOEFL test takers employ prior to sitting these tests? What factors appear to influence test takers' choice of test?
2. How do these reported test preparation practices influence test takers' perceived performance on the tests? Do test preparation practices have any effect on test takers' reported subjective emotional experience on the test?
3. What opinions do test takers have of the tests, and what are the implications of these opinions with regards to validity?

The next chapter describes the methods employed in this study to address these research questions.

Chapter Three: Methods

3.1 Research Design

The current study employs a combination of qualitative and quantitative aspects, and can therefore be described as employing a *mixed-methods* approach. According to Sandelowski (as cited in Dörnyei, 2007), mixed-methods approaches enable researchers to (a) achieve a fuller understanding of a target phenomenon, and (b) verify one set of findings against the other” (p.164). The second of these goals may be referred to as *triangulation*, in that it involves cross-validation of conclusions by comparing results obtained via different data gathering methods. Dörnyei (2007) states that “triangulation [is] an effective strategy to ensure research validity: if a finding survives a series of tests with different methods, it can be regarded as more valid than a hypothesis tested only with the help of a single method” (p.165). In this study, I am concerned with what preparation practices test takers employed prior to taking the test, how they assessed their own performance and subjective experience on the test, and their reasons for taking a specific test. To fully investigate these issues, I adopted an *exploratory convergent mixed-methods* research design (Creswell, 2009). In this design paradigm, the quantitative and qualitative data are collected and analyzed simultaneously, and the results are synthesized to provide an overall picture. Accordingly, this study was conducted using two main data-collection methods: (1) a questionnaire containing both quantitative and qualitative components, and (2) semi-structured interviews which generated qualitative data. As the quantitative strand of data collection and analysis is less dominant than the qualitative in this study, the design can be characterized as quantitative + QUALITATIVE in nature (Creswell, 2009). Figure 2 (next page) shows a schematic representation of the convergent mixed-methods research design.

Figure 2: *Convergent Mixed-Methods Research Design* (Creswell, 2009, p.37)

3.2 Setting

The setting for the data collection phase of this study was the English as a Second Language for Academic Purposes (ESLA) unit at a mid-sized Canadian university. This ESLA unit, run under the auspices of the School of Linguistics and Language Studies at the university, is intended for English as a Second Language (ESL) students who did not achieve specified cutoff scores on entry-contingent high-stakes language proficiency tests which would have allowed them to enroll in their primary degree program full-time. Students in the ESLA unit take courses that target academic English skills in the four main domains (reading, writing, speaking, and listening), as well as research and study skills. ESLA classes are grouped into three levels: ESLA 1300 (introductory), ESLA 1500 (intermediate), and ESLA 1900 (advanced), with students ideally advancing through the levels over successive semesters (although students with higher test scores, but that still fall under the cutoff for entry into full-load degree courses, may skip 1300 and/or 1500 levels). Courses are

for-credit and are completed in one semester, although failure of a level necessitates that students repeat the level if they wish to remain enrolled at the university. A limited number of for-credit courses from students' degree courses may be taken alongside ESLA courses.

3.3 Participants

All ESLA students, as part of their application process to the university, sit one of the major high-stakes language proficiency tests (TOEFL, IELTS, or CAEL). Scores from these tests are submitted to the university admissions department, which uses them to decide whether the student possesses the requisite language skills to enter full-time academic study, or needs to take ESLA courses prior to full-time registration in their degree program. Therefore, the ESLA student population represented an ideal group through which to study test preparation practices as they all would have taken at least one of these tests prior to enrollment.

On receipt of CUREB ethics approval (Appendix A), seven ESLA class teachers were contacted via email (Appendix B) and asked for permission to visit their ESLA class so that I could explain my study and distribute questionnaires to students who volunteered to participate. Of the teachers I contacted, three agreed to grant access to their classes for the purposes of the study. I visited the classes, informed the student groups about the aims of the study using a pre-written script (Appendix C), and distributed questionnaires (Appendices D1 and D2) and combined information/consent forms (Appendix E) to volunteer questionnaire respondents. A total of 33 completed questionnaires were collected: 27 from prior IELTS test takers, and 6 from prior TOEFL test takers (Table 2).

Table 2: *Overall Characteristics of Questionnaire Respondents*

No.	Pseudonym	Test Taken	Age	Gender	Home Country	Times test taken
1	Pearl	IELTS	37	F	Turkey	1
2	Anastasia	TOEFL	19	F	Serbia	3
3	Elida	IELTS	20	F	Syria	1
4	Nelson	IELTS	19	M	Palestine	2
5	Chan	TOEFL	24	F	China	3
6	Sung	IELTS	20	M	China	3
7	Edwin	IELTS	22	M	Senegal	3
8	Matt	IELTS	19	M	China	5
9	Damon	IELTS	21	M	China	3
10	Maxwell	IELTS	19	M	Palestine	6
11	Britta	IELTS	25	F	Ethiopia	1
12	Gisele	IELTS	22	F	Palestine	2
13	Fay	IELTS	21	F	Iran	1
14	Lucas	IELTS	19	M	China	5
15	Ashley	IELTS	19	F	China	4
16	Brett	IELTS	19	M	China	2
17	Forest	IELTS	20	M	China	1
18	Michael	TOEFL	19	M	China	4
19	Josiah	TOEFL	21	M	Pakistan	1
20	Cathy	TOEFL	20	F	Kenya	1
21	Hatem	TOEFL	18	M	Egypt	2
22	Paris	IELTS	20	F	China	2
23	Lacie	IELTS	19	F	China	6
24	Aliya	IELTS	22	F	Libya	1
25	Ally	IELTS	21	F	Ethiopia	2
26	Dennis	IELTS	19	M	China	4
27	Yuki	IELTS	19	F	Japan	2
28	Lewis	IELTS	18	M	China	4
29	Fatima	IELTS	21	F	Syria	2
30	Adham	IELTS	18	M	Kuwait	2

31	Sebastian	IELTS	19	M	China	4
32	Riley	IELTS	18	F	China	4
33	Fernando	IELTS	19	M	China	4

All participants had taken either the IELTS or TOEFL test in the previous two years, and were ESLA 1900 (advanced-level) students with one exception (Pearl).

At the end of each questionnaire, I included a solicitation for volunteers to participate in the interview portion of the study. From the questionnaire participant pool, five who had completed the questionnaire were recruited to participate in the interview portion of the study. In the interest of balance, I decided to recruit three participants who had taken the IELTS and two who had taken the TOEFL; consequently, the first three IELTS and the first two TOEFL respondents who volunteered via email were informed of their acceptance via email (see Appendix F). Table 3 shows the characteristics of the five interview participants.

Table 3: *Characteristics of Interview Participants*

No.	Pseudonym	Test Taken	Age	Gender	Home Country	Times Test Taken
1	Pearl	IELTS	37	F	Turkey	1
2	Anastasia	TOEFL	19	F	Serbia	3
3	Elida	IELTS	20	F	Syria	1
4	Nelson	IELTS	19	M	Palestine	2
5	Chan	TOEFL	24	F	China	3

3.4 Data Collection

For the purposes of this study, I initially classified the main types of test preparation practices described in the literature into categories to facilitate the

construction of the questionnaire items relating to preparation practices. This allowed me to present participants with a range of options from which to choose, which I considered a more effective method of data collection for this aspect of the study because it did not require questionnaire respondents to generate these categories themselves; furthermore, I considered this approach appropriate because it would facilitate the initial coding and metrics for preparation practices. The main preparation practice categories are shown in Table 4, along with citations to the source papers supporting their inclusion.

Table 4: *Test Preparation Practices by Source*

Preparation Practice Type	Source(s)
Test Preparation Classes	Brown (1998), Fox & Cheng (2015); Green (2007)
General / EAP Classes	Mickan & Motteram (2009)
Published Test Materials / Textbooks	Mickan & Motteram (2009); Roberts (2002)
Practice Tests	Roberts (2002)
Self-study	Doe & Fox (2011)
Online Resources	Mickan & Motteram (2009)
Talking to Friends	Fox & Cheng (2015)

Data collection commenced on March 14, 2018, and was completed on April 6, 2018. Three ESLA 1900 (upper-level) classes were each visited twice between the dates of March 12, 2018 and March 30, 2018, with questionnaires distributed during the first visit and collected during the second; this approach was chosen to allow participants the opportunity to take the questionnaire home and complete it without time pressure. During the second class visit, students who returned completed

questionnaires were rewarded with a gift of a study notebook as a token of thanks. On collection of all questionnaires, participants were randomly assigned a participant number and a pseudonym from a predetermined list (Appendix G).

For the semi-structured interviews, meetings were arranged via email with questionnaire respondents who volunteered for the interview portion at times convenient for both the participant and the investigator; these were conducted at appropriately quiet locations on campus, following a specified semi-structured interview protocol (Appendix H). Five interviews lasting from between 26 and 36 minutes were conducted with interview volunteers. Details of the dates and durations of the interviews are shown in Table 5.

Table 5: *Interview Dates and Durations*

Participant Pseudonym	Date of Interview	Duration of Interview (Mins)
Pearl	21 March 2018	26:38
Anastasia	30 March 2018	35:21
Chan	30 March 2018	31:51
Nelson	03 April 2018	25:58
Elida	06 April 2018	28:49

3.5 Instruments

Two instruments were used to collect data for the study: a questionnaire and a semi-structured interview.

3.5.1 Questionnaire

The questionnaire (Appendices D1 and D2) was designed to elicit information from participants regarding the test preparation practices they had employed prior to

taking the test, as well as general biographical information, self-reported experiences during the different test sections, reported effect of preparation practices on test performance, and their reasons for choosing the test they sat. A range of categories of test preparation practices were drawn up, based on findings from a variety of sources (Brown, 1998; Doe & Fox, 2011; Fox & Cheng, 2015; Green, 2007; Ma, 2017; Mickan & Motteram, 2009; Roberts, 2002). The questionnaire was printed and distributed with either TOEFL (Appendix D1) or IELTS (Appendix D2) identification although the content was identical except for this; consequently, both are therefore referred to as *the questionnaire*. 1 to 6 Likert-type interval response formats (Likert, 1932) were used for questions that asked respondents for subjective ratings, for example, how much they felt specific test preparation practices improved their performance on the test.

The questionnaire (Appendices D1 and D2) was divided into one opening section and 14 main questions. The opening section (un-numbered) elicited information relating to participants' background information (name, age, gender, home country, languages spoken, and degree program at the university).

- Question 1 asked how many years of English education (full-time or part-time) participants had had prior to taking the test.
- Question 2 asked how long participants had spent living in an English-speaking country.
- Question 3 asked for the reason participants had taken the test.
- Question 4 asked how many times participants had taken the test.
- Question 5 asked participants to rate on a Likert-type response format of 1 to 6 (1 = not at all good, 6 = excellent) how good they thought the test they took

was as a measure of (a) academic English proficiency, and (b) general English proficiency.

- Question 6 asked participants where they had taken the test.
- Question 7 asked participants whether they had prepared for the test.
- Question 8 asked participants to specify their preparation practices from a specified list drawn up using information uncovered during the literature review (using official and unofficial preparation textbooks, taking instructed courses, using free or paid-for online resources, searching online forums, talking to others who had taken the test, using media, using authentic materials).
- Questions 9 and 10 asked participants to rate on a 6-point Likert-type response format how much they thought their chosen preparation practices helped their (a) performance and (b) confidence on the test, respectively (1 = not at all, 6 = a lot).
- Question 11 asked participants to rank from 1 to 6 (with 1 being most important and 6 being least important) their opinion of the importance of specific skills (reading, speaking, listening, writing, grammar, and vocabulary) in succeeding on the test.
- Question 12 asked participants to specify their overall and section scores for the test they took.
- Question 13 asked participants to rate on a scale of 1 to 6 (1 = not at all, 6 = a lot) how much they thought the test they took prepared them for their studies.
- Question 14 asked participants to describe, in short answer form, their subjective feelings during the four sections of the test (reading, writing, speaking, and listening).

A combined information sheet and consent form (Appendix E) was attached to the front of each questionnaire, containing information on the aims of the study, the date of ethics approval, and researcher and supervisor contact information. Participants were asked to sign and date this consent form to affirm their consent to participate in the study.

3.5.2 Interview

The interview followed a *semi-structured* format, as specified by Charmaz and Belgrave (2012), in which a set of focused questions relevant to the research area are specified in the interview protocol (see Appendix H), but room is allowed for digressions that are of interest. The purpose of the interview was to collect more detailed information on participants' test preparation practices, how their preparation practices influenced their performance and subjective experience on different sections of the test, and their opinions of the test. Participants were also asked why they chose the test they sat and not another. Interviews began with general discussion of the participant's background and English experience before leading to the more in-depth questions, with the semi-structured format allowing for digressions into pertinent topics for both the researcher and the participants. As with the questionnaire, participants were asked to read and sign a combined information and consent form (Appendix I) granting the investigator permission to record their responses, along with information on the aims of the study, the date of ethics clearance, an affirmation of participation, information on data protection, and researcher and supervisor contact details. Interview volunteers were rewarded with an Amazon Canada voucher to the value of \$25 for their participation.

3.6 Analysis

The data analysis in this study was conducted for the quantitative data generated by the quantitative items on the questionnaire, and for the qualitative data generated by the qualitative questionnaire items and the interviews, in a *convergent* manner (Creswell, 2009); In accordance with the research design the analysis of both types of data was conducted non-sequentially.

3.6.1 Questionnaire Quantitative Data

All questionnaires were assigned a participant number and a pseudonym. Quantitative questionnaire responses were tabulated in a Microsoft Excel spreadsheet, with rows on the spreadsheet corresponding to respondents and columns to question response data. SPSS Version 25, was used to obtain descriptive statistics for each category of data.

3.6.2 Questionnaire Qualitative Data

Two items on the questionnaire generated qualitative data. These questions were Question 9: “Why did you choose to take the IELTS/TOEFL and not a different English proficiency test?”, and Question 19: “Please describe the feelings you experienced while taking the different sections of the test”. The responses to these two questions were transcribed from the questionnaires into an aggregate Microsoft Word document containing all respondents’ responses to Question 9, and the same for each separate section of Question 19 (reading, listening, speaking, and writing).

3.6.3 Interview Qualitative Data

Interviews were transcribed from the digital recording device (a password-protected Apple iPhone) used to record the audio into a Microsoft Word document.

These files were then printed in a format conducive to manual *qualitative coding*, with the dialogue in the left third of each page and the centre and right thirds left blank.

3.6.4 Qualitative Coding

The qualitative questionnaire and interview data were subjected to a process of *qualitative coding*. Saldaña (2009) states that “a code in qualitative inquiry is most often a word or short phrase [assigned by the researcher] that symbolically assigns a summative, salient, essence-capturing, and/or evocative attribute for a portion of language-based or visual data” (p.3). Qualitative coding thus enables the analysis of data and extraction of meaning from texts through a focused, principled means of inquiry, allowing for the detection of patterns and categorization of units of textual data. For the qualitative data in this study, a combination of coding techniques was applied to extract meaning in accordance with the research questions. *Structural coding* (p.84), which is used to construct a taxonomy of categories based on certain criteria, was used to inductively ascertain the reasons for test choice. *Affective* (or *emotion*) coding (p.105) enables the identification of subjective experiential emotion information in the data, and was used to inductively identify subjective experiences reported by questionnaire respondents and interview participants on different sections of the test, and information relating to performance on the test. *Provisional coding* (p.144) entails coding data according to a pre-identified list of categories; this technique, informed by the categories of test preparation practice detailed in the literature review, was used to identify and classify information relating to the reported preparation techniques of participants. Finally, *in-vivo coding* (p.91) was used to extract meaningful quotes from participants across all aspects of the study.

Using this combination of coding techniques, which can overall be described as an *eclectic coding* approach (Saldaña, 2009), the qualitative questionnaire and interview data were coded in two cycles. First-cycle coding involves looking for fragments of relevant information from the data, and second-cycle coding involves categorizing this information (Saldaña, 2009). As previously stated, both inductive (emergent) and deductive coding techniques were employed depending on the data type and specific question. I constructed an initial provisional coding scheme based on the literature review and research questions. The questionnaire qualitative data and the interviews were coded using the relevant techniques, and based on a combination of inductive and deductive coding I then constructed a final overall coding scheme that was used for all qualitative data in the study (Appendix J). The coding scheme was divided into eight main code categories:

- (1) *opinions of test,*
- (2) *performance on test,*
- (3) *preparation practices,*
- (4) *reasons for choosing test,*
- (5) *strategies,*
- (6) *subjective experience during test,*
- (7) *speededness, and*
- (8) *repeated test taking.*

Of these top-level codes, *opinions of test*, *performance on test*, and *subjective experience of test* each contained four second-level codes, one each for general, reading, writing, speaking, and listening elements of the test. These second-level codes each had three third-level codes: positive, negative, and neutral, corresponding to the categorization of the coded content. Therefore, if an interview participant

mentioned feeling that they performed poorly on the reading section of the test, this was coded as *performance on test + reading + negative*. The *preparation practices* first-level code had a number of sub-codes specific to types of test preparation; this taxonomy was deductively applied based on information gathered during the literature review. The *reasons for choosing test* code category was constructed inductively based on respondents' questionnaire responses to the relevant question. *The strategies, speededness, and repeated test taking* first level codes appeared after first cycle coding, as these aspects of the tests and test preparation repeatedly appeared in participant responses during the first cycle of coding.

In the interest of coding reliability, one non-coded interview was passed to a second experienced coder, who independently coded a data sample in the form of one complete interview transcription based on the coding scheme with the aim of assessing *intercoder agreement* (Creswell, 2009), which may also be referred to as a measure of *interrater reliability*. For this purpose, the two independently-coded interviews were compared for similarity of coding using a hit/miss protocol. This technique entails comparing the coded interviews and registering a 'hit' each time both coders identify the same head code from the text, and a 'miss' when one coder registers a code but the other does not. Analysis of the hit/miss data revealed that the 'hit' agreement figure was 80.2%, that is, both coders agreed with the coding choice 80.2% of the time. Creswell (2009) notes that "the consistency of the coding should agree at least 80% of the time for good qualitative reliability" (p.191); therefore, given that this threshold was met, we can conclude that the coding reliability level was acceptable (see Appendices K and L for samples of primary researcher and second coder coding, respectively).

Having ascertained the reliability of the coding, the interview and questionnaire qualitative transcribed data were imported into the NVivo qualitative analysis program for the purposes of analysis. NVivo allows the collation and encoding of qualitative data as well as word frequency counts and cross-tabulation of data and matrix searches, and is therefore a useful tool to the qualitative researcher. The coding scheme (Appendix I) was entered into NVivo as coding 'nodes', and the qualitative data Word files were converted to PDF format, imported to NVivo, and recoded in digital form according to the final scheme. This allowed the manipulation of the data in the form of cross tabulations, matrix search queries, and word counts, as well as categorization of qualitative data into an easily-searchable and indexed form.

Chapter Four: Findings and Discussion

So far, I have outlined the background to the study, summarized and discussed relevant literature, and described the methods used in the data collection and analysis phase. This chapter presents and discusses the findings of the study, and is organized as follows. First, relevant information regarding the participants in the study is presented. General information regarding the total questionnaire respondent group is presented in Table 6, followed by brief descriptions of the five interview participants. Following this, findings for each main area of the study according to the research questions are presented and discussed for both the overall questionnaire group, and each individual interview respondent. In accordance with the convergent *quantitative* + *QUALITATIVE*³ research design (Creswell, 2009), (see Figure 2). The quantitative data are analyzed using general descriptive statistics, and the qualitative data using coding frequencies based on the coding scheme (Appendix J) and relevant in-vivo quotes from interview participants.

This chapter is organized following the order of the research questions. Following brief descriptions of the questionnaire and interview respondent groups the quantitative and qualitative data are presented, analyzed, and findings discussed for each research question in turn (research questions one, two, and three).

4.1. Questionnaire Respondent Results

Table 6 (next page) shows the ratios of the gender, age, and reported first language of the questionnaire respondent group by test taken. The gender and age proportions were similar between the two groups. However, as is evident from the data, the questionnaire respondents overwhelmingly reported sitting the IELTS, with

³ The capitalization here indicates qualitative dominance (Creswell, 2009).

27 respondents (81.8%) having sat the IELTS and 6 (18.2%) the TOEFL. The largest first language (L1) group in the sample was Chinese [Mandarin and Cantonese] (48.5%), followed by Arabic (24%).

Table 6: *Questionnaire Respondent Information Breakdown by Test Taken*

Characteristic	Combined Test Takers (TOEFL & IELTS, n=33)	IELTS Test Takers (n=27)	TOEFL Test Takers (n=6)
Males/Females (%)	51.5/48.5	52/48	50/50
Age (Mean/Median)	20.5/19	20.6/19	20.2/19.5
First Language: (%)			
Chinese (Mandarin & Cantonese)	48.5%	51.9%	33.3%
Arabic	24%	25.9%	16.7%
Amharic	6%	7.4%	-
Farsi	3%	3.7%	-
Fulani	3%	3.7%	-
Japanese	3%	3.7%	-
Pashto	3%	-	16.7%
Serbian	3%	-	16.7%
Somali	3%	-	16.7%
Turkish	3%	3.7%	-
Number of Times Test Taken (Mean/Median)	2.8/2	2.9/2	2.3/2.5

As can be seen in Table 6, there was a wide disparity between the numbers of respondents for each test, with IELTS respondents representing 81.8% (n=27) of the respondent group, and TOEFL respondents representing 18.2% (n=6). In total, 33 participants responded to the questionnaire, which represents a rather small sample size. Interestingly, repeated test taking appears to be a feature of the total group, with respondents reporting having taken either the IELTS or TOEFL a mean number of 2.8

times. This is important to note, as it is a characteristic of the group that may have a bearing on why they chose to respond to the questionnaire during the data collection phase. Clearly, this group was motivated to participate in the study whereas others were not, so it should be noted that the respondent group may or may not be representative of the overall ESLA student or general IELTS/TOEFL test taker populations.

4.2 Interview Participant Profiles

This section contains a brief profile of each participant for the interview section of the study, with information elicited during the interviews. In the qualitative data analysis, I return to each interview participant and report their responses as they relate to each research question.

Pearl (IELTS)

The first participant to be interviewed, Pearl, was a native of Turkey and spoke Turkish as her first language. Pearl was unique in the interview participant group in that she was not an ESLA student, and rather was participating in an ESLA 1900 class as an assistant to the class teacher as part of the practicum requirement for her Master's degree in Applied Linguistics and Discourse Studies at the university. Nonetheless, she volunteered to participate in the interview portion of the study because she had sat the IELTS as part of her degree application process and was interested in the topic of the study. She had studied English as a subject throughout high school, and following her high school graduation she trained as an English teacher and taught at the high school level in Turkey for fifteen years. Pearl had also served as a speaking examiner for a Cambridge Exams-administered English proficiency test in Turkey, which gave her a deep insight into the language testing and

language education in general. She stated that she prepared to take the IELTS while teaching full-time, and took the test only once. Pearl sat the IELTS in Turkey.

Nelson (IELTS)

Nelson, was born and spent part of his early life in Palestine but emigrated to Saudi Arabia with his family as a child, and spoke Arabic as his first language. He attended an international school in Saudi Arabia where English was the exclusive language of instruction for 12 years prior to his high school graduation. At school, in addition to studying the main academic subjects in English, he also took Arabic and French language classes. His program at the university was an undergraduate degree in Biomedical and Electrical Engineering. Nelson sat the IELTS twice in Saudi Arabia prior to his acceptance to the university.

Elida (IELTS)

Elida was born and grew up in Syria, spoke Arabic as her first language, and was studying for a degree in Social Sciences at the university. She had studied English as a subject at school in Syria from the beginning of high school onwards. She had spent relatively little time living in an English-speaking country prior to taking the exam. Elida stated that she had taken the IELTS once, in Canada, prior to her admission to the university.

Anastasia (TOEFL)

Anastasia was born and grew up in Serbia, spoke Serbian as her first language, and was enrolled in a Communications and Media degree at the university. She stated that she had started studying English in kindergarten, and that her English study had continued throughout her primary and secondary education. For the latter part of her secondary-level education Anastasia had attended an all-girls boarding school in

Maine, USA, where she sat her SATs prior to commencing her university studies in Canada. She stated that she had sat the TOEFL three times, prior to her acceptance to the university.

Chan (TOEFL)

Chan was from China, spoke Mandarin as her first language, and was studying Communications and Media as her undergraduate degree subject. She stated had studied English as a subject throughout her education from kindergarten until graduating high school in China, and that this had been mandatory throughout her pre-university education. She also stated that she had taken after-school and weekend courses, including English, for all core subjects at elementary and high school. Chan reported that she had sat the TOEFL three times, achieving the exact same overall score each time.

4.3 Reported Preparation Practices

In response to the research question: *What test preparation practices do IELTS and TOEFL test takers employ prior to sitting these tests?* this section describes the reported preparation practices of questionnaire respondents and interview participants.

4.3.1 Questionnaire Respondents' Reported Preparation Practices

Table 7 (next page) shows the percentage of questionnaire respondents reporting specific categories of preparation practice. Preparation practice categories have been sorted in descending order of prevalence for the total test taker group (please see the following page).

Table 7: *Summary of Prevalence of Preparation Practice Use as a Percentage of Questionnaire Respondents Who Employed This Practice*

Preparation Practice Type	TOEFL and IELTS Test Takers (n=33)	IELTS Test Takers Only (n=27)	TOEFL Test Takers Only (n=6)
Talking to friends who had previously taken the test	69.7% (n=23)	70.4% (n=19)	66.7% (n=4)
Official Preparation Books/Tests	66.7% (n=22)	66.7% (n=18)	66.7% (n=4)
Free Online Materials	63.6% (n=21)	63% (n=17)	66.7% (n=4)
Audio/Visual Media	57.6% (n=19)	59.3% (n=16)	50% (n=3)
Time with English-speaking friends	48.5% (n=16)	51.9% (n=14)	33.3% (n=2)
Part-time Instructed Test Preparation Course	42.4% (n=14)	40.7% (n=11)	50% (n=3)
Authentic Texts	42.4% (n=14)	44.4% (n=12)	33.3% (n=2)
Unofficial Preparation Books / Tests	39.4%(n=13)	37% (n=10)	50% (n=3)
Online Forums	36.4% (n=12)	33.3% (n=9)	50% (n=3)
Paid Online Materials	24.2% (n=8)	18.5% (n=5)	50% (n=3)
General Instructed English / EAP Course	18.2% (n=6)	18.5% (n=5)	16.7% (n=1)
Full-time Instructed Test Preparation Course	15.2% (n=5)	18.5% (n=5)	0% (n=0)

The data for test preparation practices for the total questionnaire group shows that *talking to friends who had previously taken the test* was the most common reported preparation practice category, with 69.7% (n=23) of the 33 respondents

reporting having done this prior to taking the test. This was followed by *official preparation books/tests* (66.7%, n=22) and *free online materials* (63.6%, n=21). For the total sample, the three least commonly reported preparation practices were *full-time instructed test preparation course* (15.2%, n=5), *general instructed English/EAP course* (18.2%, n=6), and *paid online materials* (24.2%, n=8).

The top result for the combined group, talking to friends who had taken the test, echoes Fox and Cheng's (2015) finding that this was a common practice among test takers, and emphasizes the idea that high-stakes language proficiency tests are socially-situated constructs. It is therefore perhaps unsurprising that test takers seek the advice and insight of others prior to taking the test. The findings regarding published test materials and textbooks echo Mician and Motteram (2009) and Roberts (2002), who both identify these as common modes of test preparation. This also reflects the fact that there exists a vast industry based on the production and distribution of such materials (Hamp-Lyons, 1999; Wadden & Hilke, 1999). The prevalence of the use of online resources for test preparation concurs with Mician and Motteram's (2009) findings, although given how common this practice appears to be, more research is required into its exact origins, nature and mechanisms.

4.3.2 Interview Participants' Reported Preparation Practices

The reported preparation practices data for interview participants are drawn from two sources: the questionnaires that they completed as part of the overall questionnaire respondent group, and the content of their interviews. Table 8 (next page) shows the questionnaire-reported preparation practices for interview participants.

Table 8: *Interview Participants' Reported Test Preparation Practices (Questionnaire Responses)*

Preparation Practice Category	Pearl (IELTS)	Nelson (IELTS)	Elida (IELTS)	Anastasia (TOEFL)	Chan (TOEFL)
Official Preparation Books/Tests		✓	✓	✓	✓
Unofficial Preparation Books/Tests		✓		✓	✓
Part-time Instructed Test Preparation Course				✓	✓
Full-time Instructed Test Preparation Course					
General Instructed English / EAP Course				✓	
Free Online Materials	✓	✓	✓	✓	✓
Paid Online Materials				✓	✓
Online Forums				✓	✓
Talking to friends who had previously taken the test		✓	✓	✓	✓
Time with English-speaking friends			✓	✓	
Audio/Visual Media			✓	✓	✓
Authentic Texts				✓	

Interview participants' top three test preparation practices appear to generally reflect patterns of preparation practices reported by the overall questionnaire sample group, although interview participants reported *free online materials* as their most common preparation practice (n=5), followed by *talking to friends who had previously*

taken the test (n=4) and *official preparation books/tests* (n=4). The least common preparation practices reported by the interview participant group on the questionnaire were the use of *authentic texts* (n=1) and *general instructed English / EAP courses* (n=1). Among the interview participants, Pearl reported employing the fewest types of preparation practice (*free online materials* only); therefore, she may be an outlier with regards to the total group, although this may have been a function of her specific circumstances. In contrast, Anastasia reported using every category of test preparation practice except for *full-time instructed test preparation course*. We shall now turn to each interview participant in turn, and give an overall description of their individual approaches to test preparation using relevant in-vivo quotes from the participants themselves. The interview data offer a fine-grained illustration of idiosyncratic test preparation practices by respondent, and show that while individual patterns of test preparation practices differ, certain themes (e.g., the use of strategies) tend to recur across participants.

Pearl

As previously stated, Pearl reported using the fewest types of test preparation, relying mainly on freely-available materials sourced online. She stated:

Pearl: Well, at the time I was working really hard, I was teaching, and so I didn't have much time to get prepared properly, unfortunately, but I wanted [to] so the only thing I could do was to do some tests online, to have a look at the previous tests and the question types, and... I watched some webinars... about the content of the test, and I, uh, did some listening tests which were online again.

However, Pearl reported a significant error that she made during her preparation for the IELTS:

Pearl: I made a mistake, I only did those listening tasks which were prepared for the general, uh, IELTS...

Interviewer: Not the academic?

Pearl: Not the academic, unfortunately I realized the difference during the test!

Pearl stated that she paid special attention to the writing section of the IELTS during her preparation, and used online videos for advice:

Pearl: I was not used to... doing academic writing, so I focused on this too, and for the writing I watched some online videos giving some tips about what kind of structures to use, what kind of vocabulary to focus [on]...

Pearl warned about the varying quality of such freely-available online videos, stating that she had to be careful to select ones that were most useful:

Interviewer: And were those online videos, um, sort of, just YouTube videos that you could watch for free, or...

Pearl: Yeah, they were YouTube videos but there are thousands of them, so you need to be careful about which ones to pick up because there are some good quality ones, I don't remember the names but, uh, I followed those ones.

For writing, Pearl reported that she practiced her academic writing based on topics that she sourced from online resources, and checked her writing herself:

Pearl: For the writing, there were hundreds of topics, so I tried to, not write all of them, of course, but I tried to get, think some deeper

ideas, to... I mean, if I'm asked about this topic then, in order to be ready, I tried to think about them, elaborate on those topics.

Interviewer: Did you get any feedback on your writing, um, before the test?

Pearl: Hmm... I don't know... only myself. I checked my writing as my teacher!

With regards to her preparation for the speaking section, Pearl reported that she practiced her speaking skills as much as possible, by herself if necessary:

Pearl: I also checked speaking topics, and I tried to speak to the wall [laughs], to the door, and to my husband, sometimes on my own, just to try to be fluent, because you see we don't normally speak English in Turkey, so outside the classroom I didn't use to practice at all, so I tried to practice on my own.

Interviewer: So, it was, anything you could speak to at all you would speak to?

Pearl: Yeah, basically! Especially to the mirror, it works very well, I recommend [it]!

For the reading section of the IELTS, Pearl stated that she did not do any preparation at all:

Pearl: For reading I did not do any preparation because I thought that was going to be the easiest part of it, which turned out to be the opposite!

Overall, Pearl was extremely self-reliant when it came to her test preparation practices, relying on online resources as her only external source of insights into the structure of the IELTS and strategies that she could utilize to maximize her

performance on the test. Furthermore, she drew on her significant experience as a language teacher and former test examiner, as well as her own resourcefulness in considering what each section of the test was intended to assess.

Nelson

In his questionnaire responses to the question asking respondents to specify their test preparation practices, Nelson reported using official and unofficial preparation books and tests, free online materials, and talking to friends who had previously taken the IELTS. However, he stated that he used online resources, particularly online videos for most of his preparation, and revealed that he learned test taking strategies from these that he considered useful:

Interviewer: So, would you say that most of your preparation was online, through webinars and practice materials you found through the internet?

Nelson: Yes, the YouTube videos were the most things that prepared me for the IELTS test.

Interviewer: Do you remember what sort of sources they were from, or was it more like “a bit from here, a bit from there”?

Nelson: No, I guess it’s a big school, I don’t remember the name of the channel but I can look it up for you... Yeah, it was “Learn English with Anna”, that was the channel which I used.

Interviewer: Thanks. So, you talked about strategies you learned from that, so things like note taking...

Nelson: Yeah, and vocabulary words, and how to scan paragraphs, underlining dates, names, verbs, and so on.

Interviewer: And did you have to pay for that or was it free?

Nelson: No, it was completely free.

With regards to preparation books, Nelson stated that he used one official IELTS test preparation booklet which he received for free when he registered for the test:

Nelson: I didn't buy any, but this book was offered for free by the British Council when I first reserved for the test they gave me this book, so I watched the videos and then I practiced from this book, and it also has an answer key at the end, yeah... it was given for free in order to practice. It was only for April, for a limited time you got it for free when booking the test, it was kind of a special offer.

Like Pearl, Nelson drew largely from online resources to gain an understanding of the requirements of the IELTS and to learn strategies that he felt would help him on the test. He also used official preparation books, to practice his skills and gain a deeper understanding of the tasks he would face during the test.

Elida

In her questionnaire responses, Elida reported using official preparation books, free online materials, talking to friends who had previously taken the test, spending time with English-speaking friends, and audio/visual media. Elida reported that she attended a part-time IELTS preparation course, though notably this was three years before she took the test. Consequently, she did not indicate this on the questionnaire in her preparation practices prior to taking the test; however, this can be considered relevant information and so was discussed when it came to light during the interview.

Elida: Me and my friends, we were a group of three, we took some classes with a teacher, but not in Canada, abroad, and we took two days a week classes. So, it was general, and we tried to have some

idea about how the test would be, how everything's going to be, and, uh, we did some speaking and writing, and he was correcting us, so it was a good experience but I didn't do the test then because I wanted to do it before university, so... but then when I prepared before the exam it was different because I prepared for it alone... I took two months of classes so it wasn't too much, but it was good enough to have a good idea about it, and to just get over the book.

Elida reported that the classes were generally focused on skills:

Elida: It was just reading the test, practicing skimming and scanning. And some skills, you know, it's all about skills so we took some skills, like, we learned how to write, how to organize our ideas when we need to write or talk, or when we listen, when we should pay attention to the speaker.

In contrast, immediately prior to taking the test, Elida reported that she prepared alone using mostly online resources:

Elida: Before the exam I studied on my own, so I was depending on some YouTube or some links I have, I didn't study from the book, because I didn't have time to get over all the book, it's big.

Interviewer: So, was there an official book that you were using?

Elida: Well, it was links that someone I knew who'd took it before sent me, like, it was questions from the book, but in a specific way that's it's easier to study, and with the questions and answers—I needed the answers because no-one was correcting me, so I didn't know if I did well or not.

Interviewer: Right, so it was just yourself?

Elida: Yes, it was just myself, and the good thing was doing lots of preparation tests before the big test, and from the links there was uh... answers. So, I checked my answers and I compared, and I learned. And for the speaking I used YouTube, because there are good teachers that give good advice and they ask questions, [and] you answer them...

Elida stated that she found the online videos particularly useful in preparing for the speaking section, specifically with regards to strategies that could be used in constructing a spoken answer:

Elida: I [would] think about [a response], and then they would say “it’s better for you to talk this way, not that way, use this example”, and there was also the real interview, it’s really good, you can see it and take ideas, like “oh, he did that really well”.

Interviewer: Was that kind of, for example, structuring, or what kind of, bits of phrases to use etc.?

Elida: Yeah, exactly, how to start a sentence, how to give an example, how you check the time, sometimes you need to talk for two minutes and that’s not easy on the subject you are talking about.

Elida also reported practicing her speaking skills by recording herself and listening to her speaking to identify ways to improve:

Elida: I recorded myself and listened to myself, to see when I was panicking, or thinking too much...

Elida’s pattern of test preparation practices was unusual in that she had taken an instructed test preparation course, but this was three years before she sat the test. She used self-practice and critique to prepare for the speaking sections, and drew on

online resources to familiarize herself with the test. Perhaps the most salient quote from Elida was “it’s all about skills”, which encapsulated her perception of the test and explains her focus on skills and strategies.

Anastasia

Anastasia reported the greatest number of different preparation practice types in her questionnaire responses, omitting only the *full-time instructed test preparation course* option. She took the TOEFL three times in total, and reported different preparation approaches each time:

Anastasia: I took [the TOEFL] three times – first time I didn’t prepare at all, I just went there, sort of like, let’s see, I never took TOEFL, I don’t know what it’s about, I didn’t even know how long it’s going to take

Interviewer: Really? Wow.

Anastasia: Nothing! I just Googled the night before to see what is it about, and my friends were like, “it’s not that hard, but it depends what test you get”, so I was just like, let’s go, let’s waste \$200, why not! ... [The] second time I took [the TOEFL] – between [the] first and second time I was preparing a lot.

Prior to her second sitting of the TOEFL, Anastasia enrolled in distance test preparation classes run from Serbia, and she attended these online while simultaneously studying at boarding school in Maine:

Anastasia: I found, uh, it’s a like private English school, so I found it online and I ask if I can do Skype with them so I had classes online... I had, like two or three classes per week, so, and I was preparing a lot, she gave me a lot of homework, [the teacher] told

me, like, strategies to do, and everything, I took it second time and I had enough for [the university's entrance] requirements... it was one hour, the class was one hour, I think I had, like, two or three classes a week, and then a week before test I think it was every day, or something like that.

About the content of the classes, Anastasia stated that:

Anastasia: They taught me strategies for readings, so that was pretty much all I had to do, practice with those strategies and... I mean, they couldn't teach me anything else for readings, except that I had to practice on my own, listenings... uh, we practiced together I think, it was listening, so, and she gave me lot of homework to do on listenings, like compare the answers, if they're right or wrong, and pretty much on note taking for that, speaking, I practiced a lot for that because I can't fit everything in one minute or 30 seconds, like [laughs] how do you want me to do it? Do you want me to speak fast? So, I had to put down on paper what I had to say, what's important, what's not important, writings, uh, also I had, oh my god, at that time I think I wrote hundreds and hundreds of essays per week, I was writing like crazy at that time.

Anastasia reported that she utilized both paid and free online resources:

Anastasia: [I used] videos on YouTube for note-taking strategies, things like that.

Interviewer: For the TOEFL specifically?

Anastasia: Yes, I can't remember, it was something like that, note... yeah, something like note-taking it was called, specifically for

TOEFL, so I watched that, I was Googling all the time TOEFL strategies, how to prepare for TOEFL... I paid for just [online content relating to] reading strategies, for the rest I didn't want anything because I was comfortable with those ones, but I needed one for reading specifically because reading was my weakest point, uh... and the rest were free, yeah.

In relation to social media sources, Anastasia stated that she did not use online forums, but sought information about the TOEFL from friends over social media networks:

Anastasia: I was talking to my friends that already took it, so if you consider that as a social media...[laughs] because I talked with them through social media, so it's pretty much, yeah, I used it, but for... getting more information, specifically from some web page or anything, no.

Overall, Anastasia recounted casting a wide net in sourcing ways to prepare for the TOEFL, an approach she described as stemming from her strong motivation to achieve the necessary score on the test:

Anastasia: Some [other sources] were from official TOEFL website, and some were from random websites, I considered everything just to pass the test, give me everything just to pass it, and to finish with it.

Overall, Anastasia employed an 'omnivorous' approach to her test preparation, employing just about any method she could find, and reflecting her strong motivation to succeed on the test. The fact that she took the test three times further reflects her motivation to maximize her score.

Chan

In her questionnaire responses, in terms of her preparation practices prior to taking the TOEFL Chan reported the use of both official and unofficial preparation books and tests, taking a part-time instructed test preparation classes, free and paid online materials, online forums, talking to friends who had previously taken the test, and audio/visual media. However, her most significant preparation practice was enrolling in a part-time instructed test preparation course in her native China; consequently, this covered the most discussion time during the interview:

Chan: I took lessons, and had four teachers, for speaking, reading, listening and writing, and prepared for two months to take the first exam.

Interviewer: So, what exact preparation methods did you use? You said you took courses, were they full-time, every day courses?

Chan: Yes, like, today is listening, tomorrow is reading...

Interviewer: And were those all-day courses?

Chan: Three or four hours a day.

Interviewer: And how long did you do that for?

Chan: From December to March. Maybe about 5 months.

Regarding her assessment of the classes, Chan stated that she had faith in their methods:

Chan: They [the school] have a really mature model for that, with a lot of materials, and some techniques to deal with the exam. So, in the first class they gave me lots of materials and told me how to get good scores on that, and some mock exams, mock questions. And

yeah, they are really good teachers, so I wasn't nervous at all because I believed in them.

Chan stated that she learned specific skills in the classes aimed at improving performance on specific sections of the test. For speaking preparation, Chan said:

Chan: They [the teachers] gave me some speaking models, and I prepared much on the listening part, because we need to listen to some videos and make some notes and make my response, so I focused on the listening parts.

Interviewer: You mentioned 'speaking models', what did you mean by that?

Chan: Um, just like, how to begin your, like beginning sentences, and transitions, and conclusion sentences, things like that.

Interestingly, two test-related skills Chan learned during the test preparation classes was the use of shorthand for the listening sections, and the practice of folding a piece of paper 6 times to regulate space for note-taking:

Chan: First, my teachers [taught] me some symbols

Interviewer: Like shorthand?

Chan: Yeah! Most of the time I used the symbols and made notes, and they told me some, like, when I listen to the [audio], if the professors said "interestingly..." then I need to pay attention to the following sentences. If I heard some words that I don't know, I need to write down the pronunciation of that, and I can guess when I get the questions. [For the listening section] we get A4 paper, and my teacher told me to fold it 3 times because we have 6 listenings [and to use these folded portions to make notes on each listening].

Chan revealed that her teachers advised a specific approach to multiple choice questions:

Chan: There are multiple choice [questions], so sometimes if I don't know the questions, maybe I should guess the four choices, like "this is impossible, that is impossible" ...

For the reading section, Chan's teachers advised rote vocabulary learning as a preparation technique:

Interviewer: So, did your teachers prepare you in any special ways for reading?

Chan: They gave me a really long word list. Like a book! And [I had to] memorize all of them.

Interviewer: So, memorize the word and definition?

Chan: Yeah, I think that's the foundation, not only for the reading part, each of my teachers gave me a word list, like, for the listening part they gave me the word list but classified by discipline, like Psychology or Biology.

Chan also described certain strategies that her test preparation course teachers taught her for the writing section:

Chan: There's two parts of writing, first listen to the video, but my teacher told me to focus on the second writing, the big essay. I don't know if this is right, because my teacher told me how to write the first paragraph, and how to write the second paragraph's first sentences, and how to write a conclusion, like it's also a model...

Interviewer: A general structure?

Chan: Yeah, and in the first paragraph, the first sentence is a general introduction, and think I put a sentence, this is “however this is a controversial problem, but from my perspective, blah, blah...”, so I put this sentence in every essay.

Overall, Chan focused on the test preparation course, and the advice of the teachers for it, as the basis of her approach to the test. Unlike any of the other respondents, she reported unique skills like folding the note paper three times to allocate space for taking notes on each of the six listening sections, the use of shorthand in taking notes, and vocabulary memorization. Chan’s approach represents a significant example of the use of specific strategies that may be employed during the test to maximize performance, although some of these may represent test taking strategies as opposed to language use strategies.

Interview participants’ reported test preparation practices reflect the variety of practices found in the overall test taker group, with both idiosyncratic patterns of test preparation practice, ranging from the use of self-study and online resources in Pearl’s case, to an omnivorous, wide-ranging set of preparation practices in Anastasia’s case. However, in common with the questionnaire group, online resources were a salient component of the interview participants’ test preparation. Although Mickan and Motteram (2009) addressed this factor, it appears that there is a need for more research into this facet of test preparation.

The finding that all interview respondents reported using specific strategies to aid their test performance concurs with previous research into the prevalence of strategy use by test takers on high-stakes language proficiency tests (Barkaoui, Brooks, Swain, & Lapkin, 2012; Cohen, 2007; Doe & Fox, 2011). With regards to

strategies, the questionnaire and interview groups reported practices appear to suggest a combination of the types of strategies described by Issitt (2008), which address psychological preparation, general academic English awareness, and test awareness or orientation. Furthermore, the findings in this section suggest a combination of *language use* and *testwise* strategies as described by Doe and Fox (2011). An example of the former is Pearl's preparation for the speaking section, wherein she freely practiced her speaking on various topics to maximize her performance on this section of the IELTS through addressing the focal construct; this could be considered *positive washback* (Messick, 1996). Conversely, Chan mentioned the use of testwise strategies on the multiple-choice sections of the TOEFL, a practice that Gebril (2018) and others note may cause artificial score inflation and therefore be classed as *negative washback* (Green, 2007) with implications for the construct validity (Messick, 1989) of the test scores on the test.

4.4 Reasons for Choice of Test

In response to the portion of research question one that asks: *What factors appear to influence test takers' choice of test?*, this section examines test takers' reported reasons for choosing a specific test (TOEFL or IELTS). It should be noted that because the small size of the TOEFL sample this should be considered to have limited generalizability; however, TOEFL-related responses are still considered.

4.4.1 Questionnaire Respondents' Reported Reasons for Choice of Test

Table 9 (next page) shows questionnaire respondents' stated reason for taking the test, as coded using the *structural coding* approach (Saldaña, 2009) described in the Methods section [n.b. some respondents gave more than one reason for their choice of test and if this was the case, both reasons were recorded; therefore,

numerical totals add up to more than 33 for the combined test taker group, and 27 and 6 for the IELTS and TOEFL groups respectively].

Table 9: *Questionnaire Respondents' Reasons for Test Choice*

Reason for Taking Test	All Respondents (n=33) (%)	IELTS Respondents (n=27) (%)	TOEFL Respondents (n=6) (%)
Easier than others	9 (27.3%)	9 (33.3%)	0 (0.0%)
Familiar	9 (27.3%)	8 (29.6%)	1 (16.7%)
Common	6 (18.2%)	6 (22.2%)	0 (0.0%)
Recommended	4 (12.1%)	2 (7.4%)	2 (33.3%)
Required	4 (12.1%)	3 (11.1%)	1 (16.7%)
Computer-based	2 (6.1%)	0 (0.0%)	2 (33.3%)
Paper-based	2 (6.1%)	2 (7.4%)	0 (0.0%)
Human examiner	2 (6.1%)	2 (7.4%)	0 (0.0%)
Available	2 (6.1%)	0 (0.0%)	2 (33.3%)
Popular	2 (6.1%)	2 (7.4%)	0 (0.0%)
Better	2 (6.1%)	2 (7.4%)	0 (0.0%)
Content	1 (3.0%)	1 (3.7%)	0 (0.0%)
Convenient	1 (3.0%)	1 (3.7%)	0 (0.0%)
Free to take	1 (3.0%)	1 (3.7%)	0 (0.0%)

From Table 9, we may make several observations. It appears that more than a quarter of IELTS respondents (33.3%) reported choosing this test because they perceived it as *easier* than alternative tests. With regards to other reasons for choosing the IELTS, 29.6% of IELTS respondents reported that they perceived it as *more familiar* than other tests, and 22% chose the test because they considered it more *common*. On the other hand, the most common reasons for choosing TOEFL were split equally between *recommended* (33.3%), *available* (33.3%), and *computer-based*

(33.3%). None of the TOEFL respondents reported the perception that it was *easier* than other similar tests, in contrast with the IELTS test taker group.

4.4.2 Interview Participants' Reported Reasons for Choice of Test

As qualitatively coded from both interviews and questionnaire data, table 10 shows interview participants' stated reasons for choosing the specific test they sat.

Table 10: *Interview Participants' Reasons for Test Choice (Questionnaire Responses)*

Reason for Test Choice	Pearl (IELTS)	Nelson (IELTS)	Elida (IELTS)	Anastasia (TOEFL)	Chan (TOEFL)
Easier than Others	✓	✓	✓		
Familiar	✓		✓		
Common					
Recommended		✓			✓
Required					✓
Computer-based					
Paper-based					
Human Examiner		✓			
Available				✓	
Popular	✓				
Better					
Content			✓		
Convenient					
Free to Take					

Based on Table 10, two main observations may be made about reasons for test choice among interview participants. First, all IELTS test takers interviewed reported the perception that the IELTS is easier than other language proficiency tests. Second,

the only reasons the two TOEFL test takers gave for their choice of test was that it was recommended, required, or the only one available, although it should be noted that this may be a function of the small TOEFL group sample size. Individual respondents' reasons are now considered in detail.

Pearl

Pearl reported choosing the IELTS because she perceived it as easier, more familiar, and popular:

Pearl: Maybe this [the IELTS] was the most popular one, and, uh, I thought this was more doable than the other ones, like TOEFL [laughs]. I don't know why I had this prejudice, but maybe I also had a student who, uh, I helped, to uh, who had this exam years ago, but it was a general IELTS, so maybe that's why I felt closer with this one... maybe more people are taking this exam than the other ones... I knew the question types more than the others, they were more, uh, they... I mean, this was the most popular one, so became across the question types many types, even when I was getting for my university exam we had many questions from the previous IELTS tests, so I felt that I could do this better.

Nelson

Nelson stated that he chose the IELTS because he perceived it as easier than other language proficiency tests, it was recommended to him, and the speaking section was conducted with a human examiner.

Nelson: Many people told me that IELTS is easier than the TOEFL, and it's easier because you are communicating with a human, instead of a computer, because in the speaking part of TOEFL you

are speaking to a computer, that's why they told me it might be a better choice... I researched and asked people who took both tests [IELTS and TOEFL] before and they told me that IELTS is easier.

Elida

Elida reported that she chose the IELTS because she perceived it as easier/having content that she felt would be more suited to her strengths, and more familiar:

Elida: I chose it [the IELTS] because I heard it's more, like, there's no science and math and all that academic stuff in it, although there is are academic readings and subjects, but, uh I felt more comfortable doing the one without science and math and all those stuff, and also, because I had a part-time teacher that gave me lessons about IELTS, but that was, like, three years ago.

Anastasia

Anastasia took the TOEFL solely because it was the only test available to her at the time when she needed to take it.

I was [at boarding school] in Maine, so at the time it was only TOEFL offered in my high school, none of [the] other [tests]... I didn't have [any] other choice unless I wanted to travel down to Boston and take another test.

Chan

On her questionnaire response, Chan wrote that the university required TOEFL. However, in her interview she stated that she took the TOEFL because it was recommended by her teachers based on her perceived strengths:

Chan: My teacher told me these were my choices, IELTS and TOEFL, and she asked me which part of English you think you are good at, and I thought maybe listening, and so she said maybe you should take the TOEFL exam, because listening is really important in the TOEFL.

Interviewer: So, your teacher pointed you towards the TOEFL?

Chan: Well yeah, but it's really... I think it wasn't a hard choice, there were only two choices! I didn't really think deeply about that.

We will now consider research question 1, and summarize the findings of the previous two sections as they relate to this question.

4.4.3 Research Question 1: Summary of Findings

What reported test preparation practices do IELTS and TOEFL test takers employ prior to sitting these tests? What factors appear to influence test takers' choice of test?

Among the combined questionnaire respondent group (n=33), the six most common ways respondents reported preparing to take the IELTS or TOEFL were:

- *talking to friends who had previously taken the test*, (69.7%, n=23),
- *official preparation books/tests*, (66.7% n=22)
- *using free online materials* (63.6% n=21),
- *using audio/visual media* (57.6% n=19),
- *spending time with English-speaking friends* (48.5% n=16), and
- *enrolling in a part-time instructed test preparation course* (42.4% n=14).

These findings were supported by the interview data, with four out of five participants reporting *talking to friends who had previously taken the test* and using *official preparation books/tests*, and all five respondents reporting the use of *free online materials*. Interview participants reported a variety of patterns of preparation practices, with extremes represented by Pearl, who relied solely on online materials and her own knowledge and experience as a teacher to guide her preparation; and Anastasia, who employed a wide variety of preparation practices. Differences between the test taker groups (IELTS and TOEFL) were impossible to assess given the small size of the TOEFL sample group (6) in comparison to the IELTS group (27).

As is apparent from consideration of the reported test taking practices of the questionnaire respondents and the interview participants, the IELTS and TOEFL appear to exert a *washback* effect that manifests as the use of a variety of test preparation behaviours by students. These test preparation practices appear motivated by test takers' understanding of the stakes associated with the test, and manifested in individually unique combinations, but with several appearing to be more common, including seeking the advice of friends and acquaintances, using preparation books and tests, and availing of online resources. As discussed in the literature review section of this study, test preparation practices must be carefully evaluated to ensure they do not threaten the validity of tests due to *testwise* strategies that artificially inflate test scores (Anastasi, 1981; Messick, 1996; Miyasaka, 2000), although in practice it is likely that test takers will utilize any option available to maximize their scores.

With regards to reasons for choosing a specific test, 33% of IELTS respondents reported that they chose this test because they perceived it as easier than other tests, a finding that has implications for the putative equivalence between the

IELTS and TOEFL with regards to test score users. 29.6% of IELTS respondents reported that they chose the test because they perceived it as more familiar than other tests, and 22.2% chose it because it was more common. The data for the TOEFL group should be cautiously interpreted due to the small sample size of respondents who had sat the TOEFL; however, none of the TOEFL respondents reported that facility was a reason for choosing the test.

Having considered preparation practices prior to sitting these tests, and factors that influence test takers' choice of test, in the following sections I explore how these tests taking practices appear to influence participant test takers' accounts of their performance and their subjective emotional experience during the test. After the findings are presented and discussed, I consider research question 2, and summarize the findings as they relate to this question.

4.5 Perceived Effect of Preparation Practices on Test Performance

In response to the portion of research question two that asks: *How do these preparation practices influence perceived performance on the test?*, an unexpected potential link was discovered between reported performance on the test and subjective emotional experience for specific parts of the test. This finding is elaborated on after consideration of the findings.

4.5.1 Questionnaire Respondents' Perceived Effect of Preparation Practices on Test Performance

Table 11 (next page) shows questionnaire respondents' evaluations of test preparation practices effect on performance on the test for the combined, IELTS, and TOEFL groups, and is sorted from highest to lowest effect for the combined group

means for each preparation practice category according to aggregated 1 to 6 Likert-type ratings

Table 11: *Summary of Questionnaire Respondents' Perceived Effect on Test Performance of Specific Preparation Practices (Likert-type response format, 1 to 6)*

Preparation Practice Type	TOEFL and IELTS Test Takers (Mean/Median/n)	IELTS Test Takers Only (Mean/Median/n)	TOEFL Test Takers Only (Mean/Median/n)
Paid Online Materials	5.2 / 5.5 (n=6)	5/5 (n=4)	5.5/5.5 (n=2)
Unofficial Preparation Books/Tests	5.1 / 5 (n=13)	5.2/5 (n=10)	4.7/5 (n=3)
Part-time Instructed Test Preparation Course	5 / 5 (n=13)	4.9/5 (n=10)	5.3/5 (n=3)
Official Preparation Books/Tests	4.9 / 5 (n=23)	5/5 (n=19)	4.5/5 (n=4)
Free Online Materials	4.9 / 5 (n=20)	4.8/5 (n=17)	5/5 (n=3)
Time with English-speaking friends	4.9 / 5 (n=16)	4.8/5 (n=14)	5.5/5.5 (n=2)
Full-time Instructed Test Preparation Course	4.8 / 5 (n=5)	4.8/5 (n=5)	-
Talking to friends who had previously taken the test	4.5 / 5 (n=21)	4.4/4.5 (n=18)	5/5 (n=3)
Audio/Visual Media	4.2 / 4 (n=19)	4.1/4 (n=16)	4.7/4 (n=3)
General Instructed English/EAP Course	4 / 4 (n=7)	3.8/4 (n=6)	5/5 (n=1)
Online Forums	3.8 / 4 (n=11)	3.6/4 (n=8)	5/5 (n=3)

Authentic Texts	3.7 / 4 (n=7)	3.8/4 (n=5)	3.5/3.5 (n=2)
-----------------	---------------	-------------	---------------

The similarity in the figures for participants' perceptions of test performance as described by 1 to 6 Likert-type responses was surprising, with one scale point difference covering the top nine practices, and 1.5 points covering all practices. However, this may be because there is a possibility that Likert response formats are not necessarily true interval scales. Jamieson (2004) notes that "the response categories in Likert scales have a rank order, but the intervals between values cannot be presumed equal" (p.1217). Therefore, these findings should be interpreted cautiously. In addition, it should be noted that reported practice effects were only considered for those participants who reported using these practices prior to taking the test.

Table 12 (next page) shows all participants' reported assessment of their performance on the test, as coded from the qualitative questionnaire and interview data. It should be noted that for questionnaire respondents, in response to the question *Please describe the feelings you experienced while taking the different sections of the test*, some reported information that was ascertained to relate to performance as opposed to emotional experience; this was therefore categorized as *performance* under the *structural coding* (Saldaña, 2009) category during qualitative coding. Figures refer to the number of participants reporting performance for each category at least once.

Table 12: *Test Takers' Reported Performance Per Test Section (Number of Respondents Coded At Least Once in Each Category) (n=33)*

Performance Assessment Category	Section of Test				
	General	Listening	Speaking	Reading	Writing
Positive	0	7	11	2	6
Neutral/Mixed	2	2	0	3	3
Negative	0	7	5	12	9

From Table 12, it is apparent that the speaking section of the tests was associated with the greatest number of positive subjective performance perceptions across all respondents, followed by listening; conversely, the reading section, followed by writing, had the greatest number of individuals coded as reporting negative performance perceptions. It should be noted, however, that not all respondents registered an opinion regarding performance on the sections of the test; these data were gathered from questionnaire respondents' responses to a question on the questionnaire regarding subjective experience, to which many responded with information that was coded as relating to performance instead.

4.5.2 Interview Participants' Perceived Test Performance

Pearl

Pearl reported that she felt her preparation practices for the writing aided her performance on the IELTS:

Pearl: For writing, yeah, I focused on academic writing, and that worked I can say, because before that I didn't know anything about how to write academically, so... if I hadn't worked on writing

before I took the exam I wouldn't have been able to do anything,
and I would have got a much lower score.

Even though Pearl did not do any specific preparation for the reading, she was pleased with her performance on this section of the test:

Pearl: [It] was to my surprise that I got an 8.5 [out of 9 on the]
reading [section], because it was really hard.

Interestingly, Pearl reported that the timed nature of the test [speededness] influenced her perceived test performance, particularly during the reading and writing sections:

Pearl: The reading passages, they were very, very hard and we had
very limited time... [during the writing section] I was really ready to
write, because you know I prepared very well, but I was not
satisfied, I was not happy with what I wrote and I had to erase and
write again, erase and write again, so I had some timing problems,
so towards the end of the time I had to be, uh, less careful [about]
what I was writing.

With regards to the speaking section, Pearl reported that she was satisfied with her performance, although about the independent speaking task she had this to say:

Pearl: I just tried to produce something which sounded silly to
myself, maybe that was the reason of getting, uh... Yeah, it's not
bad, 7.5, still... for those speaking... yeah, she, she asked me some
more questions so to elaborate on, uh, what I was telling, and
luckily...she had another question on another topic, which I could
talk more, but I, yeah, I was okay.

Pearl's error in practicing for the listening section of the general IELTS, as opposed to the academic version, was reflected in her perceived performance on the test:

Pearl: As I told you I made this mistake of focusing on general IELTS listening texts so I found them, oh, these are very easy to understand so I don't need to go any further! [laughs]... which was my mistake and which resulted in getting a 7, the lowest... of the scores

Nelson

Nelson related that he had the most difficulty with the reading and writing sections of the IELTS:

Nelson: I only had trouble in the writing part and the reading part, I wasn't actually confident while taking these parts of the test.

In elaborating on his performance during the writing section, Nelson had this to say:

Nelson: For the writing part, the first part was the one which has the diagram and I lost marks on this, because I don't know how to analyze a diagram, it was about the internet speed in Asia and USA in a graph, like, how is the speed in this country and that country, I was unable to analyze it, I had to write 150 words in the first part, I wrote way more than 150 because I wanted to include every single detail, but for the second part I was well-prepared because I had something similar in my school, it was approximately the same prompt but different things.

Like Pearl, Nelson mentioned the timed nature of the reading section as having an adverse effect on his performance, particularly on the reading section of the test:

Nelson: I practiced before, but I also faced a lot of problems in the test because I thought the time was going to be enough, I faced difficulties at the end, I wasn't able to finish, because there were

three readings, the first one is the easiest, the second one is medium, and the last one was the hardest... if I had 15 extra minutes I might have done better on the reading part, I might have answered more accurately on some parts, and also that helps me in solving the questions accurately.

Nelson felt that he performed best on the speaking section of the IELTS, and attributed some of this performance directly to his preparation:

Nelson: For the speaking section, it was actually the best part, and I got the best score of all sections, from the videos I watched I learned they don't want you to use hard words, you don't have to show them you know how to speak English but you should use the proper sentence lengths and so on. The part I was most prepared for was the speaking test, and it was easiest for me, I got a 7.0 [out of 9].

Nelson also reported being reasonably pleased with his performance on the listening, and again directly linked his preparation practice to his performance:

Nelson: I did three practice tests online. I expected the grade I got, it was 6.5, it was the same thing in the test which I did. The first one was 6, the second and third were 6.5. It was somehow consistent.

Elida

Elida summed up her performance on the sections of the test as follows, mentioning the timed nature of the reading section:

Elida: For me I was good at speaking, I don't know if you noticed that! And I love writing, so the listening was a little bit more difficult for me, and the reading... because you don't usually read

the three articles in three hours or less, so I needed to practice that, and also, you need to answer the questions, yeah.

Elida recounted that she experienced a moment of doubt during the speaking section with regards to the content of the question, but she realized that the content isn't as important as the use of language in answering the prompt:

Elida: I faced one problem, when I understood the question wrong, so I wanted to talk about something, I prepared to talk about something, but then I noticed in the last ten seconds “Oh, what if she’s talking about that, not this! What if she means this thing!” So I changed everything, I changed all my ideas I changed everything, all my ideas, but the problem was if I told her about what I was preparing that would be better, so I felt sad about this, because my speaking was not as good as I expected, but I got a really good mark in that, maybe I didn’t do a lot of mistakes, but they don’t really care about the subject, whether it’s good or not, so it was not really a good subject, but it was good speaking so I got a good mark.

On the listening section, Elida stated that she experienced a lapse in concentration, but she felt that this did not significantly affect her score for this section:

Elida: I faced one problem in the test, which was... I lost myself, I was listening and I felt that I... there was one blank where I didn’t write and he’s talking about the rest, so I didn’t know what to do!

Anastasia

With regards to the writing section, Anastasia reported that strategies she had learned through her preparation aided her performance on this section of the TOEFL:

Anastasia: Writing, I didn't find that hard ...[the] first time that I took [the TOEFL] I wrote it in the way I don't know how, because I didn't prepare for that one, so and then for the second and third time I knew strategy how to start, how to finish, what to put in each one, so pretty much I knew strategies so it was easy, for me.

And,

Anastasia: That helped a lot, because, um, my prof told me, like, strategies, how to start introduction, first, third and second, uh, first second and third supporting point for second writing

Anastasia reported that she felt that she had performed well on the listening section, although she voiced concerns about the length of the listening extracts:

Anastasia: Listenings weren't that hard, it's just they were so long, so you lose concentration and you have, how many, like... ten listenings or something?

She also reported that note-taking was a major part of her approach to the listening elements of the test, both in the integrated tasks and the listening section of the TOEFL:

Anastasia: I'm fast at note taking, so I write everything, and it was pretty much easy to find the answer because I wrote every single thing that they said, I mean maybe some non-important ones I did, but mostly I wrote everything so I had the answers, so it wasn't that hard

Anastasia reported that she found the speaking section easy; however, like Pearl, Nelson and Elida, she voiced concerns about speededness:

Anastasia: Speaking, it's simply easy...only the time limit, I would expand it to two minutes [laughs] because I can't... you have only thirty seconds to prepare, but it wasn't that hard, like once you get into a routine of practicing to speak fast and [putting] all the information in, that's fine.

Anastasia explicitly stated that strategies she learned from her preparation aided her performance on the independent speaking tasks particularly:

Anastasia: I had on paper, like writing three reasons, three supporting points that I'm going to mention, like refer to question or what was it, can't remember actually, specifically questions, but yeah, first two were basic questions about myself, so I always had at least two supporting points and if I can come up with a third one in those 15 seconds of preparation that's good, if not I just briefly explain those two, and... yeah, so it helped me a lot, those preparation and classes.

Anastasia reported the most difficulty during the reading section of the TOEFL:

Anastasia: It was really hard and going back to text and looking, trying to find the answer, and the questions are really tricky, honestly, all of them. So maybe like a few of them are easy like to put your score up, pretty much, but all the rest of them are like, "what do you want me to do?" I don't understand, where do you want me to find the answer?" Because it can be A, C and D also, but it's not... here were also some questions where nothing could work. I feel like nobody knows the answer [laughs], that's... only the people who made it.

Chan

In contrast to all other participants, Chan rated the speaking section of the TOEFL as the most difficult for her:

Chan: In our high school or primary school we focused much on the reading and listening, not on speaking.

Interviewer: So, was it more difficult because of that?

Chan: Yeah, it's my weakest part. I don't put much faith in it, if I got 22 or 21 [out of 30 on the speaking section] that's great, I'm happy.

Chan also reported experiencing difficulties on the reading section of the test, a problem she attributed to her perceived lack of vocabulary:

Chan: Reading is difficult for me, because I don't like to memorize vocabulary, so there's a lot of words I don't know, so I think that's why I get low scores on reading, so I think that vocabulary is the most important part in reading.

With regards to the listening section, Chan stated that she tried to use the note-taking skills she had learned during her test preparation course, but found this difficult in practice:

Chan: I was too busy to make notes, and at first I think that's really useful, but after that I think that distracts me from really listening to what the professor said, so I make notes, really good notes, organized, but after that I saw the notes and did not really understand what I write, so I think that's distracting me.

Overall, the findings for the interview participants mirror the questionnaire respondents' impressions of their performance on the test, with reading appearing to be associated with the greatest number of negative reports, followed by writing.

4.6 Subjective Emotional Experience During Test

The following section describes the data for test takers' subjective emotional experience during the different sections of the test. As the findings below illustrate, there was an association between test takers accounts of performance and reported subjective emotional experience.

4.6.1 Questionnaire Respondents' Reported Subjective Experience During Specific Test Sections

Table 13 shows data coded using both questionnaire respondents' replies to the questionnaire item "Please describe the feelings you experienced while taking the different sections of the IELTS test", and interview data. This was coded using a *structural coding* (Saldaña, 2009) approach. Figures refer to the number of participants reporting each emotion category at least once.

Table 13: *Numbers of All Respondents Reporting Specific Categories of Subjective Emotional Experience During Different Sections of Test (Number of Respondents Coded At Least Once in Each Category) (n=33)*

Emotion Category	Section of Test				
	General	Listening	Speaking	Reading	Writing
Positive	0	10	16	3	5
Neutral/Mixed	1	4	5	2	4
Negative	1	17	9	20	15

It is immediately apparent from the data that reading was associated with the greatest number of questionnaire respondents reporting negative emotions (20), followed by listening (17), writing (15) and speaking (9). Conversely, the speaking section was associated with the greatest number of positive emotions (16), followed by listening (10), writing (5) and reading (3).

Table 14 shows the incidence of emotion words appearing in all questionnaire respondents' replies to the open-ended question: *Please describe the feelings you experienced while taking different sections of the test.*

Table 14: *Ranking of Total Incidence of Emotion Words Appearing Across All Questionnaire Respondents' (IELTS and TOEFL) Qualitative Responses*

Word Rank	Listening	Speaking	Reading	Writing
1	confident (8)	confident (12)	confused (6)	confident (5)
2	nervous (7)	nervous (9)	nervous (5)	nervous (5)
3	prepared (4)	prepared (4)	prepared (4)	confused (4)
4	anxious (2)	comfortable (3)	confident (3)	prepared (4)
5	confused (1)	afraid (2)	afraid (2)	anxious (1)
6	demotivated (1)	anxious (2)	anxious (2)	comfortable (1)
7	excited (1)	scared (2)	scared (2)	panicked (1)
8	panicked (1)	confused (1)	stress (2)	satisfied (1)
9	relaxed (1)	happy (1)	excited (1)	tired (1)
10	terrified (1)	silly (1)	hopeless (1)	-

In this analysis, stemmed words have been included, meaning that different word forms (e.g., confusion, confuse, confused) have been classified as the adjective form (e.g., confused). Although this method of analysis excludes the context of the words, one main observation can be made: the two most commonly occurring emotion

word for reading have a negative connotation (confused, 5; and nervous, 9), and the most commonly occurring words with regards to writing are confident (5) and nervous (5). As Cassaday and Johnson (2002) note, *test anxiety* [negative emotional experience] during the test is likely to have a detrimental impact on test scores independent of other factors. Therefore, test takers' reported subjective experience on the reading section may have a feedback effect, with negative emotional experience leading to sub-optimal performance.

4.6.2 Interview Participants' Reported Subjective Experience During Specific Test Sections

This section describes, in detail, the subjective emotional experiences reported by participants during their interviews.

Pearl

Pearl reported some negative feelings on the listening and writing sections of the IELTS:

Pearl: The part that I was most, uh, let's say scared [laughs] was listening... and writing, because I was not used to doing academic writing.

With regards to the speaking section, Pearl described a moment of panic stemming from being briefly confounded by one question:

Pearl: The examiner, uh, wanted me to talk about, uh, a topic that was not easy to be productive immediately, so I need to give a lot of thought, which... which caused a feeling of panic. I don't remember the question exactly but it was like, uh, something like "a surprising present you bought for someone", you see, it's not easy to

answer...because you have to think about it before you answer this,
you see? I'm sure that there were many surprising presents but I just,
it was not easy to make up one at that moment, or to remember one
at that moment

Nevertheless, this feeling of panic passed and Pearl reported feeling better about the next question:

Pearl: Luckily, she [the speaking examiner] had another question on another topic, which I could talk more, but I, yeah, I was okay, I felt confident.

Pearl described feeling strong negative emotions during the reading section, stemming from her perception of the difficulty of the texts:

Pearl: I could understand but I didn't realize because of panic because obviously, I could, I could do it very well... but still, it was really torturing, it was a torturing hour for me [laughs] so, it was so hard the reading passages, so hard and long, with a lot of vocabulary, yeah.

Due to mistakenly expecting a pause between the third and fourth sections of the test, Pearl reported that she experienced some negative emotions during the listening section:

Pearl: [As a result] I was so nervous you see, so I missed half of the last section, and I was, I lost my motivation of course, but that was my mistake because after I finished the test I immediately googled because I was sure I was going to complain about this, but I checked and then I saw that there was no pause between [these sections].

Nelson

Nelson reported feeling a lack of confidence during the reading and writing sections of the IELTS.

Nelson: I only had trouble in the writing part and the reading part, I wasn't actually confident while taking these parts of the test.

Although he generally felt good about the speaking section of the test, Nelson reported that when speaking to the examiner:

Nelson: It was natural at the beginning, until the last part when I was a bit confused about the topic

For the listening section, Nelson stated that his preparation practices had contributed to a feeling of confidence:

Nelson: I think listening was the one I saw most improvement, because at the beginning I wasn't able to listen and take notes at the same time, it needed a lot of time to practice on this, I practiced in different classes on this, like math, chemistry, biology, biology was the most thing because it has a lot of definitions, so I practiced writing what the teacher says in order to improve my ability in taking the test.

Elida

Elida reported some general nervousness about taking the IELTS, but overall nothing excessive at the beginning. She reported high confidence during the speaking section of the test, but also some worry about the question content:

Elida: I was confident about [the speaking section] because I felt like it's my best part—I enjoy speaking English and I don't feel like

I have a problem with that, but for sure I was also worried about the subject and the questions she's going to ask me, the examiner.

Elida reported some negative feelings during the reading section of the test, caused by her impression that her skills were insufficient for the task at hand:

Elida: I also searched for skills to improve reading, like look for key sentences, find dates and words, underline it, those skills, this is how I improved it but I was nervous about it in the exam because I felt like that wasn't enough for me, I needed more. But this was all I could do.

Anastasia

Anastasia used fewer emotion words in her answers than other participants, and often answered questions intended to probe emotional responses with commentary about her performance. However, she did describe the feelings she experienced while preparing for and taking the test in general terms, stating:

Anastasia: It [taking the test] very stressful for, especially, like, high school age, and when you're thinking, like, this is either going to get me into university or not... [it was] the most stressful time of my life.

Anastasia reported feeling annoyance during the speaking section of the test due to the distraction of hearing other test takers speak:

Anastasia: Everybody is pretty much doing it at the same time, speaking, so you can hear other people, and that's annoying.

Chan

Like Anastasia, Chan responded to questions intended to probe emotional responses with descriptions of her performance, and used few emotion words in her

responses. In her interview, Chan reported feeling few negative emotional experiences during the TOEFL, which she attributed to how her preparation course teachers prepared her to take the test:

Chan: They are really good teachers, so I wasn't nervous at all because I believed in them.

Interviewer: So, you weren't nervous when you were taking the test?

Chan: Not really, no!

However, Chan described feeling distracted during the test by some of the techniques she had been taught:

I was too busy to make notes, and at first I [thought] that's really useful, but after that I think that distracts me from really listening to what the professor said, so I make notes, really good notes organized, but after that I saw the notes and did not really understand what I write, so I think that's distracting me.

4.6.3 Research Question 2: Summary of Findings

How do these preparation practices influence their perceived performance on the tests? Do test preparation practices have any effect on test takers' subjective emotional experience on the test?

The evidence that specific test preparation practices improved perceived performance on specific sections of the tests was inconclusive, as the influence of reported test preparation practice on reported test section performance (as described by a 1 to 6 point Likert-type statement) did not show considerable difference. However, there is a possibility that this may be related either to the small sample size, or as mentioned earlier to the properties of Likert-type response formats themselves; as Jamieson (2004) notes, Likert data should be cautiously interpreted.

Interview respondents reported that specific practices that improved their strategic awareness of how to approach the test improved their perceived performance on the test; all interview participants emphasized the role of test taking skills and strategies as having an influence on their test performance. These skills and strategies generally appeared not to threaten test construct validity, and included such skills as note-taking, discourse structuring; however, several participants stated that they considered test-taking skills to have a greater influence on test success than language use ability. Notably, one participant (Chan) mentioned the use of multiple choice testwiseness strategies.

With regards to the second part of research question 2, for the participant group, emotions on the test appeared more aligned with perceived performance than test preparation practices. It is notable that, when compared, the tables displaying *test takers' reported performance per test section as coded from qualitative questionnaire data* (Table 12), and *Numbers of All Respondents Reporting Specific Categories of Subjective Emotional Experience During Different Sections of Test* (Table 13) show that the speaking section had the highest number of positively coded reports on both, and the reading section the highest number of negatively coded reports. These findings support several previous studies, for example Tsai and Li (2012) that suggest that the reading sections of high-stakes language proficiency tests are a particular cause of anxiety among test takers.

4.7 Opinions of Test

This section describes salient findings with regards to participants' opinions of the test they took, and particularly specific concerns that were raised.

4.7.1 Participants' Opinions of Test

Data for all respondents' opinions of the test were coded using interview data, as well as answers to the questionnaire item "Please describe the feelings you experienced while taking different sections of the test". As in previous sections, although this question sought to elicit information on emotional experience, respondents included other information in their answers. One salient finding was that out of all respondents, 19 participants, representing 57.6% of the total sample (16 IELTS and 3 TOEFL) remarked on having difficulty with *speededness*, almost exclusively in relation to the reading and writing sections. According to Lu and Sireci (2007), speededness refers to "the situation where the time limits on a standardized test do not allow substantial numbers of examinees to fully consider all test items" (p.29). Shown below are selected examples of such responses by section:

Reading

Forest: It is the hard part for me, the time is too short.

Aliya: I was nervous and scared because I don't like reading and we have limited time so it was hard.

Gisele: It was very long and it took a long time to understand. The first two readings were okay but I did not have time to finish the third one, so I felt that I didn't prepare well.

Writing

Ally: The writing was hard for me because it takes me time to start thinking what to write about, so thinking took most of my test time and I didn't have enough time to finish everything I wanted to write about.

Adham: I had a hard time to write in an academic way and to complete the two sections in an hour.

Fay: Not prepared. Some issues. Ran out of time. Bad time management.

I shall now turn to the opinions of the test voiced by individual interview participants. Because space constraints preclude an in-depth exploration of all interview participants' opinions of the test they took in fine detail, we shall consider each interview participant, their general opinions of the test, and anything they considered particularly problematic.

Pearl

Although Pearl stated that she considered the IELTS a good assessment of academic English ability, she made several observations about factors that she found problematic; these related to speededness (Fox & Cheng, 2007; Lu & Sireci, 2007), the test environment, and the speaking examiner. With regards to the test in general and speededness, Pearl had this to say:

Pearl: I think the reading [section] does what it's supposed to do, because the texts are academic level, I agree with that, for listening yes, it does what it's supposed to do, for writing too, yeah, it absolutely does what it's supposed to, and, no, I think all these sections do what they're supposed to but the only problem is that, uh, if, the fact that you cannot do well in this test does not indicate that you are not good at these sections because you have a time constraint.

Pearl described the room in which she took the test, and stated that she did not consider the environment ideal for taking such an important test:

Pearl: For the listening I want to add something else, because the problem with IELTS listening is that you have to take, uh, this test in a very crowded classroom...[during] this listening test I was seated in a room of hundreds of people, maybe, I couldn't count it but it was at least a hundred, but they played, uh... they played the texts using those huge...

Interviewer: Oh, PA systems, big speakers?

Pearl: Yeah, uh-huh, big speakers, and then sometimes, it was wintertime...and people were sneezing, and [sniffs] doing this all the time, so it was really hard to concentrate.

Pearl stated that during the speaking section, she was somewhat taken aback by the examiner's spoken English ability.

Pearl: My examiner, the, uh, I mean her English was not that good. I mean the pronunciation, her word choice. Of course, she followed the, uh...

Interviewer: The script?

Pearl: Yeah, script, but then I was really productive, talked a lot, and, but then she didn't seem so happy with that, I don't know why.

Nelson

Nelson reported generally positive impressions of the IELTS as an assessment of the ability to use English in academic settings:

Nelson: Actually, it's a really good test for universities, because if people don't know how to speak English they will find a lot of

difficulties with communicating with professors, taking notes, doing assignments and so on, so it's actually something mandatory for a student when applying to university to be well prepared in speaking English, and understanding English.

With regards to the test environment, Nelson described his experience as positive and without problems:

Nelson: For the listening reading and writing, it was a big hall with around 150 people. It was fine, because each table was far from the next ones so it was like you were sitting by yourself.

However, Nelson also remarked negatively on the speededness of the reading section:

Nelson: Actually, for me it was the time for the reading part, the time was the thing I thought was unfair because I wanted more time, I wanted at least 15 extra minutes, for the reading part...if I had 15 extra minutes I might have done better on the reading part, I might have answered more accurately on some parts, and also that helps me in solving the questions accurately.

Elida

Despite having generally positive impressions of the IELTS, Elida made interesting observations relating to test taking skills and speededness:

Elida: I know you maybe get good vocabulary from preparing, but at the same time I think that any person, even if he doesn't have good English or good vocabulary, he can take a good mark in IELTS, because I think it's more about skills than English. I think it's more about skills.

And:

Elida: Sometimes in my own first language, if I have [to read] three big articles in one hour, I can't even do it sometimes! It depends on the article.

Elida reported that the room in which she sat the test was suitable for its purpose:

Elida: It was small, but there weren't too many people so it was fine, and it was a good atmosphere.

Anastasia

Anastasia reported generally negative opinions about the TOEFL, mostly relating to the stress of taking the test, and the fact that she took three tests which she perceived to differ in difficulty:

Anastasia: I took it three times, I had three completely different tests, one was easy, another one was extremely hard... I had some tests, or like, readings where they were harder for understanding generally, vocabulary was extremely hard [compared to] previous ones, so that was weird

In addition, Anastasia voiced negative opinions about how the structure of the TOEFL differs from test to test:

Anastasia: I hated, you know how for example if you have, like four readings you have like, three listenings? I hated that. Why can't you put, like, three-three or four-four, do always the same, because you never know, like, you click... to open the test and you find out how many readings you have.

Anastasia also expressed dissatisfaction with the test environment:

Anastasia: it was hard because in TOEFL, so everybody's in the same room, and everybody is pretty much doing it at the same time, speaking, so you can hear other people, and that's annoying.

Furthermore, Anastasia reported that the computer interface format of the TOEFL may be problematic:

Anastasia: Staring at a computer constantly for three hours, you're not blinking, you're just reading, reading, reading, writing, writing, writing, so that's... I find it very hard. I find it harder to read on computer.

With regards to speededness, Anastasia's main concern was with the speaking section:

Anastasia: For me it [the listening section] was easy, only the time limit, I would expand it to two minutes because I can't... you have only thirty seconds to prepare, but it wasn't that hard, like once you get into a routine of practicing to speak very fast, and put all the information in.

Chan

Overall, Chan expressed positive opinions about the TOEFL:

Chan: I think the listening is good, because the video is really similar to the [classes] I take here [at the university], so that's good, and the writing, I think that's good too, because the essay is similar to what I do here, and uh, reading, I don't think reading.... Because reading is from different disciplines, and I only took communication and media now, so I still don't know much professional words about my program so I need to learn here, but there [is] some general vocabulary [that is] useful

However, Chan expressed concerns about the speededness of the speaking section, as well as the test room environment:

Chan: The speaking, I don't think it's good because I need to talk to a computer, and I need to pay attention to the time limit, and I don't know how they are evaluating my speaking level, so I think that's not comprehensive. And during the exam, at one moment everybody starts talking, and my teacher told me before the exam, if I took maybe later I could maybe listen to some people's first questions! But I think, people sitting near me, they want to hear me!

4.7.2 Research Question 3: Summary of Findings

What opinions do test takers have of the test(s), and what are the implications of these opinions with regards to validity?

The findings show that the most salient patterns of opinions apparent in the data with regards to the tests fell under two categories: *speededness* (Fox & Cheng, 2007; Lu & Sireci, 2007) and adverse test conditions potentially causing *construct-irrelevant variance* (Fox & Cheng, 2015; Haladyna & Downing, 2004; Ma, 2017; Messick, 1996). Fox and Cheng (2015) report that speededness contributes negatively to test takers' subjective emotional experience during tests, while Lu and Sireci (2007) raise the point that if speededness is a factor contributing to test scores but is not one of the constructs being assessed in a test, this has serious negative implications with regards to validity. Over half of all respondents (57.6%) reported difficulty with the time limits on the reading and/or writing sections of the test, findings that echo Fox and Cheng (2007), and Lu and Sireci (2007), who note that "speededness [in tests that do not have speed of response as a focal construct] introduces a severe threat to the validity of interpretations based on test scores" (p.29).

The findings regarding adverse test conditions also have implications with regards to validity. Interview participants reported such problems as distraction from other test takers, crowded test rooms, computer fatigue in the case of the TOEFL, and problems with the speaking examiner in the case of the IELTS. There is a possibility that these could potentially give rise to *construct-irrelevant variance* (see, for example, Fox & Cheng, 2007; Ma, 2017; Messick, 1996), which has implications for the validity of the tests and the interpretation of test scores.

In the next and final chapter of this study, I will summarize the research findings, discuss their implications with regards to validity, washback, and test preparation, and consider the study's limitations.

Chapter Five: Conclusion

The core aim of this study was to investigate the test preparation practices that test takers report employing when preparing to take two similar high-stakes language proficiency tests (IELTS and TOEFL). In addition, the study sought to shed light on how test preparation practices informed test takers' reported performance on the tests, reported subjective emotional experiences during the tests, reported reasons for choosing a specific test, and opinions of the tests. The research was guided by the overarching concept of *washback* (Alderson & Wall, 1993; Bailey, 1996; Cheng, Curtis, & Watanabe, 2007; Green, 2007) which postulates that tests influence teaching and learning; and the concepts of *high-stakes tests* (Kunnan, 2012), *test preparation* (Anastasi, 1981; Messick, 1982; Smith, 1991), and *validity* (Cronback & Meehl, 1955; McNamara & Roever, 2006; Messick, 1989; Kane, 2013). To address the research questions guiding the study, an exploratory convergent mixed-methods research design (Creswell, 2009) was employed, with data gathered using a questionnaire incorporating both quantitative and qualitative items, and interviews with selected participants. The next section briefly summarizes the findings of the study with regards to the stated research questions. Following this, the implications of the study are considered, followed by its limitations. Finally, directions for further research are suggested.

5.1 Summary and Implications of Findings

Several salient findings are apparent from the results of this study. First, test takers employ a variety of test preparation practices when preparing for high-stakes language proficiency tests, in this case the TOEFL and IELTS. In this study, the most common of these were seeking knowledge from friends and acquaintances who were

familiar with the test, followed by official test preparation books and tests, and free online resources. These findings support the washback hypothesis that high-stakes tests will influence learners and learning (Alderson and Wall, 1993), and illustrate students' motivation to succeed on these tests. However, we must be aware that washback can be *negative*, *positive*, or *neutral*, and can vary over time and between individuals (Cheng, 2007). Therefore, teachers, materials developers and producers, and indeed students, should strive to promote and engage in practices that address the stated constructs as opposed to those that may interfere with validity, and test takers should be encouraged to critically assess the information they receive from a range of sources. Realistically, test takers will most likely primarily aim to achieve high test scores, because high-stakes tests like the IELTS and TOEFL act as *gatekeepers of ambition*, and represent a hurdle that must be overcome to enable them to follow their dreams. Nevertheless, validity matters, and students may be best served by engaging in preparation practices that will serve them well both on the test, and afterwards in their academic careers. With regards to the reasons test takers choose a specific test, it appears that there is a common opinion (among IELTS test takers) that the IELTS is easier than the TOEFL. This is worthy of investigation because if this claim can be supported, this has implications for the validity of inferences drawn from both tests: if they purport to assess the same focal constructs, then it should necessarily follow that one should not be easier or more difficult than the other.

The second main finding of this study is that there appears to be a connection between students' reported performance on sections of the tests and their reported emotional experience during those sections. Although this study aimed to investigate the effects of certain test preparation practices on performance during specific sections of the tests, no evidence was found to suggest any differences; however, it seems clear

from the data that students report both negative emotional experience and performance self-evaluation on the reading sections of the tests specifically. This suggests that in choosing their preparation practices, test takers may be well advised to take account of these aspects by focusing on emotion management and developing the requisite reading skills to succeed on the test and beyond.

Finally, the results of the study suggest that the test environment may be a source of construct-irrelevant variance that may have a negative effect on test scores, and therefore potentially represent a threat to the validity of these scores. Testing companies should strive to make their tests as fair and equitable as possible by ensuring that the environments in which students take their tests are fit for purpose and do not threaten the tests scores' validity. Furthermore, the issues of speededness and construct-irrelevant variance are a significant concern with regards to the validity of inferences drawn from IELTS and TOEFL scores. Speededness appears to be a salient concern, as from test takers' perspectives the data appear to suggest that this could be considered a barrier to the valid demonstration of language proficiency.

Overall, as a teacher, this study has given me an insight into the stakes at play in high-stakes language proficiency tests, how test takers prepare to take these tests, why they choose their specific test preparation practices, and what opinions they have of the tests. The salience of strategies in this study has illustrated of how these may be positive or negative, and as a teacher, it is useful to remember that the construct extends into, and arises from, the academic world in which students aim to perform. Therefore, teachers and materials developers should aim to encourage students to develop the skills that will help them not just to pass a high-stakes test, but to succeed in their academic and professional careers.

5.3 Limitations of the Study

The study had several significant limitations. First, the overall sample size was small, at a total of 33 questionnaire respondents, therefore the generalizability of the findings is somewhat limited. Second, the discrepancy between the sizes of the IELTS and TOEFL groups precluded any meaningful comparison of these tests along the dimensions addressed in the research questions, so for quantitative measures the groups were combined. Third, the tests were treated as largely analogous because they purport to assess the same constructs; however, due to their subtle differences in their philosophical underpinnings (Chalhoub-Deville & Turner, 2000) this may be open to question. Fourth, the sample group for this study consisted exclusively of students in an upper-level EAP program at one Canadian university; this also limits the generalizability of the study.

5.4 Future Directions

As Fox and Cheng (2015) note, more research needs to be conducted into testing from the perspective of test takers. With regards to this study, it may be worth expanding the questions asked to incorporate other high-stakes language proficiency tests not covered in this study, and to examine differences between the IELTS, TOEFL, and other such high-stakes language proficiency tests based on the foundational research questions of this study. In addition, it is apparent that more research needs to be conducted with regards to the use of online resources in test takers' preparation practice.

References

- Alderson, J. C. (2009). Test review: Test of English as a Foreign Language: Internet-based Test (TOEFL iBT). *Language Testing*, 26(4), 621-631. doi:10.1177/0265532209346371
- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13(3), 280-297. doi:10.1177/026553229601300304
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129. doi:10.1093/applin/14.2.115
- Alsagoafi, A. (2018). IELTS economic washback: a case study on English major students at King Faisal University in Al-Hasa, Saudi Arabia. *Language Testing in Asia*, 8(1), 2-13. doi:10.1186/s40468-018-0058-3
- Anastasi, A. (1981). Coaching, test sophistication, and developed abilities. *American Psychologist*, 36(10), 1086-1093. doi:10.1037//0003-066x.36.10.1086
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language testing*, 13(3), 257-279. doi:10.1177/026553229601300303
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F. (2007). What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment. In Fox, J. (Ed.). *Language testing reconsidered* (41-71). Ottawa, ON: University of Ottawa Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford: Oxford University Press.
- Barkaoui, K., Brooks, L., Swain, M., & Lapkin, S. (2012). Test-takers' strategic behaviors in independent and integrated speaking tasks. *Applied Linguistics*, 34(3), 304-324. doi:10.1093/applin/ams046
- Brooks, L., & Swain, M. (2014). Contextualizing Performances: Comparing Performances During TOEFL iBT and Real-Life Academic Speaking Activities. *Language Assessment Quarterly*, 11(4), 353-373. doi:10.1080/15434303.2014.947532
- Brown, J. D. H. (1998). Does IELTS preparation work? An application of the content-adaptive model of language program evaluation. In S. Wood (Ed.), *IELTS Research Reports Volume 1*(2) (20-37). Retrieved from IELTS Research Reports website: <https://www.ielts.org/teaching-and-research/research-reports/volume-01-report-2>

- Cambridge Assessment English (2018). *IELTS (International Language Testing System)*. Retrieved from <http://www.cambridgeenglish.org/exams-and-tests/ielts/>
- Cassaday, J. C., Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary educational psychology* 27(2), 270-295. doi:10.1006/ceps.2001.1094
- Chalhoub-Deville, M. (2016) Validity theory: Reform policies, accountability testing, and consequences. *Language Testing*, 33(4), 453-472. doi: 10.1177/0265532215593312
- Chalhoub-Deville, M., & Turner, C. E. (2000). What to look for in ESL admission tests: Cambridge certificate exams, IELTS, and TOEFL. *System*, 28(4), 523-539. doi:10.1016/s0346-251x(00)00036-1
- Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (Eds.). (2011). *Building a validity argument for the Test of English as a Foreign Language*. New York, NY: Routledge.
- Cheng, L. (2008). Washback, impact and consequences. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education (2nd ed.)*: Language testing and assessment (349–364). New York, NY: Springer. doi:10.1007/978-0-387-30424-3_186
- Cheng, L. (2014). Consequences, impact, and washback. In A. J. Kunnan (Ed.), *The companion to language assessment* (1130–1146). Hoboken, NJ: John Wiley & Sons. doi:10.1002/9781118411360.wbcla071
- Cheng, L., Watanabe, Y., & Curtis, A. (2004). *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Cizek, G. J. (2012). Defining and distinguishing validity: Interpretations of score meaning and justifications of test use. *Psychological Methods*, 17(1), 31-43. doi:10.1037/a0026975
- Cohen, A. D. (2007). Test-taking strategies. *Language Testing*, pp. 89-111.
- Cole, N., Osterlind, S. (2008). Investigating differences between low- and high-stakes test performance on a general education exam. *The Journal of General Education*, 57(2), 119-130. doi:10.1353/jge.0.0018
- Creswell, J.W. (2009). *Research design: Quantitative, qualitative, and mixed-methods approaches*. London: SAGE Publications.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, 52(4), 281-302. doi:10.1037/h0040957

- DeLuca, C., Cheng, L., Fox, J., Doe, C., & Li, M. (2013). Putting testing researchers to the test: An exploratory study on the TOEFL iBT. *System*, 41(3), 663-676. doi:10.1016/j.system.2013.07.010
- Doe, C. & Fox, J. (2011). Exploring the testing process: Three test takers' observed and reported strategy use over time and testing contexts. *The Canadian Modern Language Review*, 67(1), 29-54. doi:10.3138/cmlr.67.1.029
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.
- Elder, C., & O' Loughlin, K. (2003). Investigating the relationship between intensive English language study and band score gain on IELTS. In R. Tulloh (Ed.), *IELTS Research Reports Volume 4*(6) (207-254). Retrieved from IELTS Research Reports website: <https://www.ielts.org/teaching-and-research/research-reports/volume-04-report-6>
- ETS (2018). *Who accepts TOEFL scores?* Retrieved from https://www.ets.org/toefl/ibt/about/who_accepts_scores, May 29, 2018.
- Fox, J. & Cheng, L (2007). Did we take the same test? Differing accounts of the Ontario Secondary School Literacy Test by first (L1) and second (L2) language test takers. *Assessment in Education*, 14(1), 9-26. doi:10.1080/09695940701272773
- Fox, J., & Cheng, L. (2015). Walk a Mile in My Shoes: Stakeholder Accounts of Testing Experience with a Computer-Administered Test. *TESL Canada Journal*, 32(9), 65-86. doi:10.18806/tesl.v32i0.1218
- Gebril, A. (2018). Test preparation in the accountability era: Toward a learning-oriented approach. *TESOL Journal*, 9(1), 4-16. doi:10.1002/tesj.302
- Green, A. (2007). Washback to learning outcomes: A comparative study of IELTS preparation and university pre-sessional language courses. *Assessment in education: principles, policy & practice*. 14(1), 75-97. doi:10.1080/09695940701272880
- Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practices*, 23(1), 17-27. doi:10.1111/j.1745-3992.2004.tb00149.x
- Hamp-Lyons, L. (1997). Washback, impact and validity: Ethical concerns. *Language Testing*, 14(3), 295-303. doi:10.1177/026553229701400306
- Hamp-Lyons, L. (1999). Ethical test preparation practice: The case of the TOEFL. *TESOL Quarterly*, 32(2), 329-337. doi:10.2307/3587587
- IELTS (2017a). *What is IELTS?* Retrieved from <https://www.ielts.org/what-is-ielts/ielts-introduction>

- IELTS (2017b). *Test format*. Retrieved from <https://www.ielts.org/about-the-test/test-format>
- IELTS (2018). *Guide for teachers: Test format, scoring and preparing students for the test*. Retrieved from <https://www.ielts.org/-/media/publications/guide-for-teachers/ielts-guide-for-teachers-2017-uk.ashx>
- Issitt, S. (2008). Improving scores on the IELTS speaking test. *ELT Journal*, 62(2), 131–137. doi:10.1093/elt/ccl055
- Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education* 38(1212-1218). doi:10.1111/j.1365-2929.2004.02012.x
- Kane, M. (2006). Validation. In R. Brennan (Ed.), *Educational measurement (4th ed.)*. (17–64). Westport, CT: American Council on Education and Praeger.
- Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, 50, 1–73. doi:10.1111/jedm.12000
- Kim, J. (2007). *The effectiveness of test-takers' participation in development of an innovative web-based speaking test for international teaching assistants at American colleges*. (Doctoral dissertation, University of Illinois at Urbana-Champaign).
- Kunnan, A. J. (2012). High-Stakes Language Testing. In *The Encyclopedia of Applied Linguistics*, C. Chapelle (Ed.). doi:10.1002/9781405198431.wbeal0504
- Kyle, K., Crossley, S. A., & McNamara, D. S. (2016). Construct validity in TOEFL iBT speaking tasks: Insights from natural language processing. *Language Testing*, 33(3), 319-340. doi:10.1177/0265532215587391
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*, 22, pp. 5-55
- Liu, O. L. (2014). Investigating the Relationship between Test Preparation and TOEFL iBT Performance. Research Report. ETS RR-14-15. *ETS Research Report Series*. doi:10.1002/ets2.12016
- Lu, Y., & Sireci, S. G. (2007). Validity issues in test speededness. *Educational Measurement: Issues and Practice*, 26(4), 29-37. doi:10.1111/j.1745-3992.2007.00106.x
- Ma, J. (2017). *Understanding test preparation phenomenon through Chinese students journey to success on high-stakes English language tests*. (Unpublished doctoral thesis). Kingston, ON: Queen's University.

- Malone, M. E., & Montee, M. (2014). Stakeholders' Beliefs About the TOEFL iBT® Test as a Measure of Academic Language Ability. *ETS Research Report Series*, 2014(2), 1-51. doi:10.1002/ets2.12039
- McNamara, T.F. (2000). *Language Testing*. Oxford: Oxford University Press.
- McNamara, T. (2009). Language tests and social policy. In Hogan-Brun, G., Mar-Molinero, C., & Stevenson, P. (Eds.). *Discourses on language and integration: Critical perspectives on language testing regimes in Europe* (Vol. 33). Amsterdam: John Benjamins Publishing.
- McNamara, T.F. & Roever, C. (2006). Validity and the social dimension of language testing, pp. 9-42. *Language Testing: The social dimension*. Oxford: Oxford University Press.
- Menken, K. (2008). High-stakes tests as de facto language policies in education. In E. Shohamy and N. Hornberger (Eds.), *Encyclopedia of language and education* 7 (401-414). Alphen aan den Rijn: Kluwer.
- Merrifield, G. (2008). An impact study into the use of IELTS as an entry criterion for professional associations-Australia, New Zealand and the USA. *International English Language Testing System (IELTS) Research Reports 2008*, 8(1).
- Messick, S. (1982). Issues of effectiveness and equity in the coaching controversy: Implications for educational and testing practice. *Educational Psychologists*, 17(2), 67–91. doi:10.1080/00461528209529246
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.), pp. 13-103. New York: American Council on Education and Macmillan Publishing Co.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241– 256. doi:10.1177/026553229601300302
- Mickan, P., & Motteram, J. (2009). The preparation practices of IELTS candidates: Case studies. *IELTS Research Reports Volume 10*(5). Retrieved from IELTS Research Reports website: <https://www.ielts.org/teaching-and-research/research-reports/volume-10-report-5>
- Miyasaka, J. R. (2000). *A framework for evaluating the validity of test preparation practices*. Paper presented at the annual meeting of the American Educational Research Association, Chicago. Retrieved from ERIC database (ED454256)
- Roberts, M. (2002). TOEFL preparation: What are our Korean students doing and why. *The Korea TESOL Journal*, 5(1), 81-106.
- Saldaña, J. (2009). *The coding manual for qualitative researchers*. Los Angeles, CA: SAGE Publications, Inc.

- Shin, S. (2013). Proficiency Scales. In *The encyclopaedia of applied linguistics*, C. Chapelle (Ed.), pp. 1-7. doi:10.1002/9781405198431.wbeal1423
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. London: Longman.
- Smith, M. L. (1991). Meanings of test preparation. *American Educational Research Journal*, 28(3), 521–542. doi:10.3102/00028312028003521
- Stoneman, B. W. H. (2006). *The impact of an exit English test on Hong Kong undergraduates: a study investigating the effects of test status on students' test preparation behaviours*. (Unpublished Doctoral dissertation). The Hong Kong Polytechnic University, Hong Kong.
- Stricker, L. J., & Attali, Y. (2010). Test takers' attitudes about the TOEFL IBT. *ETS Research Report Series*, 2010(1). doi:10.1002/j.2333-8504.2010.tb02209.x
- Tsai, Y. C., & Li, Y. C. (2012). Test anxiety and foreign language reading anxiety in a reading-proficiency test. *Journal of Social Sciences*, 8(1), 95. doi:10.3844/jssp.2012.95.103
- Wadden, P., & Hilke, R. (1999). Comments on Liz Hamp-Lyons "Ethical Test Preparation Practice: The Case of the TOEFL" Polemic Gone Astray: A Corrective to Recent Criticism of TOEFL Preparation. *TESOL Quarterly*, 33(2), 263-270. doi:10.2307/3587720
- Wall, D. (2000). The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled? *System*, 28(4), 499-509. doi:10.1016/s0346-251x(00)00035-x
- Wall, D., & Alderson, J. C. (1993). Examining washback: the Sri Lankan impact study. *Language testing*, 10(1), 41-69. doi:10.1177/026553229301000103
- Wall, D. & Horák, T. (2006) The impact of changes in the TOEFL examination on teaching and learning in Central and Eastern Europe: Phase 1, The baseline study. *TOEFL Monograph Series*, MS – 34. Princeton, NJ: Educational Testing Service. doi:10.1002/j.2333-8504.2006.tb02024.x
- Wall, D. & Horák, T. (2008). *The impact of changes in the TOEFL examination on teaching and learning in a sample of countries in Europe: Phase 3, The role of the coursebook. Phase 4, Describing change*. Manuscript submitted for publication, Lancaster University. doi:10.1002/j.2333-8504.2011.tb02277.x
- Ward, A. M., & Liu, X. (1994). *The relationship between summarization skills and TOEFL scores*. Paper presented at the annual meeting of the Teachers of English to Speakers of Other Languages, Baltimore, MD. Retrieved from ERIC database. (ED 394332)

- Xie, Q., & Andrews, S. (2013). Do test design and uses influence test preparation? Testing a model of washback with Structural Equation Modeling. *Language Testing*, 30(1), 49-70. doi:10.1177/0265532212442634
- Yu, G. (2012). *Preparing for TOEFL iBT speaking tasks: Test-takers' experiences and expectations*. Paper presented at 34th Language Testing Research Colloquium, Princeton, New Jersey.
- Zhengdong, G. (2009). IELTS preparation course and student IELTS performance: A case study in Hong Kong. *RELC journal*, 40(1), 23-41. doi:10.1177/0033688208101449
- Zumbo, B. & Hubley, A.M. (2016). Bringing consequences and side effects of testing and assessment to the foreground. *Assessment in Education: Principles, Policy & Practice*, 23(2), 299-303. doi:10.1080/0969594x.2016.1141169

Appendix A: Ethics Approval Certificate

Office of Research Ethics and Compliance
5110 Human Computer Interaction Bldg | 1125 Colonel By Drive
| Ottawa, Ontario K1S 5B6
613-520-2600 Ext: 2517
ethics@carleton.ca

CERTIFICATION OF INSTITUTIONAL ETHICS CLEARANCE

The Carleton University Research Ethics Board-A (CUREB-A) has granted ethics clearance for the research project described below and research may now proceed. CUREB-A is constituted and operates in compliance with the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans* (TCPS2).

Ethics Protocol Clearance ID: Project # 108517

Project Team Members: Mr. Ciaran McCullough (Primary Investigator)
Janna Fox (Research Supervisor)

Project Title: Gatekeepers of Aspiration: The Effects of Test Preparation Techniques on Actual and Perceived Test-Taker Performance and Confidence on High- Stakes Language Proficiency Tests

Funding Source (If applicable):

Effective: February 27, 2018

Expires: February 28, 2019

Restrictions:

This certification is subject to the following conditions:

1. Clearance is granted only for the research and purposes described in the application.
2. Any modification to the approved research must be submitted to CUREB-A via a Change to Protocol Form. All changes must be cleared prior to the continuance of the research.
3. An Annual Status Report for the renewal of ethics clearance must be submitted and cleared by the renewal date listed above. Failure to submit the Annual Status Report will result in the closure of the file. If funding is associated, funds will be frozen.
4. A closure request must be sent to CUREB-A when the research is complete or terminated.
5. Should any participant suffer adversely from their participation in the project you are required to report the matter to CUREB-A.

Failure to conduct the research in accordance with the principles of the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans 2nd edition* and the *Carleton University Policies and Procedures for the Ethical Conduct of Research* may result in the suspension or termination of the research project.

Upon reasonable request, it is the policy of CUREB, for cleared protocols, to release the name of the PI, the title of the project, and the date of clearance and any renewal(s).

Please contact the Research Compliance Coordinators, at ethics@carleton.ca, if you have any questions.

CLEARED BY:

Date: February 27, 2018

Andy Adler, PhD, Chair, CUREB-A

Bernadette Campbell, PhD, Vice-Chair, CUREB-A

Appendix B: Email Request to ESLA Teachers

Dear [ESLA instructor],

My name is Ciaran McCullough, and I am a 2nd year MA student in Applied Linguistics and Discourse Studies in the School of Linguistics and Language Studies at Carleton. My area of study is language testing and assessment, and my MA thesis supervisor is Dr. Janna Fox

I am currently in the preliminary stages of researching my MA thesis, entitled “Gatekeepers of Aspiration: The Effects of Test Preparation Techniques on Test-Taker Performance and Confidence on High-Stakes Language Proficiency Tests”. I am interested in how test-takers prepare for the TOEFL iBT and IELTS tests, and how these preparation techniques influence test-takers’ reported confidence on the test, as well as their objective test score outcomes.

The first stage of my research will entail asking individuals who have sat either the TOEFL iBT or IELTS in the previous two years to complete a short questionnaire about the preparation techniques they employed prior to taking the test. Respondents will be rewarded for their participation with a small (< \$5 value) gift.

If possible, would I be able to visit your ESLA class to recruit participants for my study? My suggestion is that I visit in the last 5-10 minutes of a lesson, briefly explain the study, and distribute questionnaires to students who agree to participate. To avoid rushing respondents they will be allowed to take these questionnaires home to complete, and I will visit at the end of the same class the following week / day to collect responses.

I am also looking for five volunteers to participate in semi-structured interviews to ask follow-up questions about their test-preparation techniques for the qualitative element of the study. Students who are willing to participate in a short (~30 minute) interview will be asked to email me to communicate their availability. I will accept the first three respondents who have sat the TOEFL iBT, and the first three who have sat the IELTS, and these volunteers will be rewarded with a \$25 Amazon Canada voucher.

I plan to begin collecting data in March 2018. If you are willing to allow me to visit your class, please let me know by emailing me at ciaran.mccullough@cmail.carleton.ca before March 1st, 2018.

Please find attached a copy of the script I intend to read during the class visit.

Best regards,
Ciaran McCullough

Appendix C: ESLA Class Visit Script

TOEFL iBT / IELTS STUDY

ESLA Class Visit Script for questionnaire distribution (Ciaran McCullough)

Good [morning/afternoon], ESLA students,

My name is Ciaran McCullough, and I am an MA student in the School of Linguistics and Applied Language Studies. I would like to thank [ESLA instructor's name] for kindly allowing me to visit your class today.

I have been an ESL / EFL teacher for around eight years, and I have taught in Japan, France, and here in Canada. My background as a language teacher led me to the MA program in Applied Linguistics at Carleton, and I am currently in the process of conducting research for my MA thesis.

The topic I have chosen to research is large-scale, high-stakes language tests, specifically the TOEFL iBT and the IELTS tests. I am sure most of you will know about these tests; indeed, some of you will have sat one of them as part of your application process to Carleton.

As you probably know, the TOEFL iBT and IELTS are critical tests in many people's lives, used as they are for academic selection and immigration, to name just two. Therefore, I believe that it is necessary to research these tests to gain a deeper understanding into how students prepare for them, and the usefulness of these preparation techniques.

My specific research questions are as follows:

1. What techniques / methods do test-takers use to prepare for the TOEFL iBT / IELTS tests?
2. What effect do specific test preparation techniques have on test performance on the TOEFL iBT and IELTS tests?
3. What effect do specific test preparation techniques have on students' perceived confidence during the TOEFL iBT and IELTS tests?

I am visiting your class today because I have prepared a short questionnaire to collect information that I hope will help answer these questions. If you:

- have sat the TOEFL iBT or IELTS within the last two years, and
- are willing to complete the questionnaire,

please take a questionnaire on your way out. I will collect these tests at the beginning of [Wednesday's] class next week.

If you are willing to participate in this study, please read and sign the consent form attached to the front of the questionnaire. You are not obliged to complete this questionnaire, and if you do decide to participate you will be able to withdraw at any time with no questions asked. Students who return a completed questionnaire will receive a small gift for their participation.

I am also looking for five volunteers to interview about how they prepared to take the TOEFL iBT or IELTS (three who took the IELTS and two who took the TOEFL). If you are interested in being interviewed, please email me at my Carleton email address on the last page of the questionnaire, putting the test you took (IELTS / TOEFL) in the subject line. Interviews will take approximately 30 minutes, and will take place on campus at a convenient time for the participant. Individuals who volunteer to be interviewed will be rewarded with a \$20 Amazon Canada voucher for their participation, and the first three respondents for each test will be accepted.

Thank you for your attention!

Appendix D1: TOEFL Test Taker Questionnaire

TOEFL iBT TEST-TAKER FEEDBACK QUESTIONNAIRE

(Adapted from Fox & Cheng, 2015, Roberts, 2002, Doe, 2006).

Have you taken the TOEFL test in the previous two years? If so, we would like to hear your feedback!

You have the right to withdraw from this study at any time or skip any questions you do not wish to answer.

Name: _____

Age: _____

Gender: _____

Home country: _____

Languages spoken:

First / native language: _____

Second language: _____

Third language: _____

Degree program at Carleton: _____

Why did you choose to take the TOEFL test (and not another)?

1. How many years of English education did you have before taking the TOEFL?

Full-time education in English (e.g., international school, school in an English-speaking country): _____ years _____ months

Part-time English education (e.g., after-school English lessons):
_____ years _____ months

Total years of Education in English:
_____ years _____ months

2. Until now, how much time have you spent living in an English-speaking country (e.g., Canada, the UK)?

_____ years _____ months

3. Why did you take the TOEFL? **Check (✓) all that apply.**

___ To get into university

___ For my work

___ For practice

___ Other (Please specify)

-
4. How many times in total have you taken the TOEFL?

I have taken the TOEFL _____ times

5. How good do you think the TOEFL is as a measure of:

(a) academic English proficiency, and

(b) general English language proficiency?

(1 = not at all good, 6 = excellent)

Academic English proficiency 1 2 3 4 5 6

General English language proficiency 1 2 3 4 5 6

6. Where did you take the TOEFL test? Check (✓) all that apply.

☐ In my home country (Please specify)

☐ In Canada

☐ In neither my home country nor Canada: (Please specify)

7. Did you prepare in advance for the TOEFL? (Please circle) YES / NO

If **YES**, please answer questions **6** to **11**.

If **NO**, please answer questions **9** and **10**.

8. How did you prepare for the test? Check (✓) all that apply.

☐ I used official TOEFL preparation books / tests.

☐ I used an unofficial preparation book / test.

☐ I took a part-time instructed TOEFL preparation course.

☐ I took a full-time instructed TOEFL preparation course.

☐ I took an instructed general English course.

☐ I used free online practice materials (including videos / online courses).

☐ I used paid online practice materials (including videos / online courses).

☐ I searched online forums (e.g., Facebook groups, Reddit) for information on how to take the test.

☐ I talked to friends who knew about / had taken the test.

☐ I spent time with English-speaking friends.

☐ I watched TV / watched movies / listened to radio programs / listened to podcasts in English.

___ I used authentic materials such as newspapers, academic textbooks.

___ Other (please specify):

9. Please rate how much you think your preparation method(s) helped your performance on the test (1 = not at all, 6 = a lot).

Official TOEFL preparation books / tests 1 2 3 4 5 6

Unofficial test preparation books / tests 1 2 3 4 5 6

Part-time instructed TOEFL preparation course 1 2 3 4 5 6

Full-time instructed TOEFL preparation course 1 2 3 4 5 6

Instructed general English course. 1 2 3 4 5 6

Free online practice materials / courses (including videos)

1 2 3 4 5 6

Paid online practice materials / courses (including videos)

1 2 3 4 5 6

Online forums (e.g., Facebook, Reddit) 1 2 3 4 5 6

Talking to friends who knew about the test. 1 2 3 4 5 6

Spending time with English-speaking friends. 1 2 3 4 5 6

Watching TV / movies / listening to radio programs / podcasts in

English.

1 2 3 4 5 6

Using authentic materials such as newspapers, academic textbooks

Other (please specify) _____

1 2 3 4 5 6

10. Please rate how much you think your preparation method(s) helped improve your confidence when taking the test (1 = not at all, 6 = a lot)

Official TOEFL preparation books / tests 1 2 3 4 5 6

Unofficial preparation books / tests 1 2 3 4 5 6

Part-time instructed TOEFL preparation course	1	2	3	4	5	6
Full-time instructed TOEFL preparation course	1	2	3	4	5	6
Instructed general English course.	1	2	3	4	5	6
Free online practice materials / courses (including videos)	1	2	3	4	5	6
Paid online practice materials / courses (including videos)	1	2	3	4	5	6
Online forums (e.g., Facebook, Reddit)	1	2	3	4	5	6
Talking to friends who knew about the test.	1	2	3	4	5	6
Spending time with English-speaking friends.	1	2	3	4	5	6
Watching TV / movies / listening to radio programs / podcasts in English.	1	2	3	4	5	6
Using authentic materials such as newspapers, academic textbooks	1	2	3	4	5	6
Other (please specify) _____	1	2	3	4	5	6

11. How important do you think the following skills are for attaining a high TOEFL test score? Please **rank the categories 1 to 6**, with **1 being ‘most important’** and **6 being ‘least important’**.

Speaking: ____

Reading: ____

Listening: ____

Writing: ____

Grammar ____

Vocabulary ____

12. What TOEFL test score did you attain for each section? (**Only answer this question if you are comfortable doing so**)

Total TOEFL score: _____

If you remember your scores for each section of the test, please write them below.

Reading: _____

Listening: _____

Speaking: _____

Writing: _____

13. Overall, how much do you think taking the TOEFL test prepared you for your studies? (1 = **not at all**, 6 = **a lot**)

1 2 3 4 5 6

14. Please describe the **feelings** you experienced while taking the different sections of the TOEFL test (e.g., did you feel confident, nervous, prepared, anxious, confused, anything else? Why?)

Speaking section:

Listening section

Reading section:

Writing section:

If you are interested in participating in a short (30 minute) interview about your experience of preparing for the TOEFL, please contact me at ciaranmccullough@cmail.carleton.ca. The first 2 email responses for past TOEFL will be accepted for this phase of the study, and will be compensated for their participation with a \$20 Amazon Canada voucher. Thank you for your participation!

Appendix D2: IELTS Test Taker Questionnaire

Carleton
UNIVERSITY

IELTS TEST-TAKER FEEDBACK QUESTIONNAIRE

(Adapted from Fox & Cheng, 2015, Roberts, 2002, Doe, 2006).

Have you taken the IELTS test in the previous two years? If so, we would like to hear your feedback!

You have the right to withdraw from this study at any time or skip any questions you do not wish to answer.

Name: _____

Age: _____

Gender: _____

Home country: _____

Languages spoken:

First / native language: _____

Second language: _____

Third language: _____

Degree program at Carleton: _____

Why did you choose to take the IELTS test (and not another)?

1. How many years of English education did you have before taking the IELTS?

Full-time education in English (e.g., international school, school in an English-speaking country): _____ years _____ months

Part-time English education (e.g., after-school English lessons):
_____ years _____ months

Total years of Education in English:
_____ years _____ months

2. Until now, how much time have you spent living in an English-speaking country (e.g., Canada, the UK)?

_____ years _____ months

3. Why did you take the IELTS? **Check (✓) all that apply.**

___ To get into university

___ For my work

___ For practice

___ Other (Please specify)

4. How many times in total have you taken the IELTS?

I have taken the IELTS _____ times

5. How good do you think the IELTS is as a measure of:

(a) academic English proficiency, and

(b) general English language proficiency?

(1 = not at all good, 6 = excellent)

Academic English proficiency

1 2 3 4 5 6

6. Where did you take the IELTS test? **Check (✓) all that apply.**

☐ In my home country (Please specify)

☐ In Canada

☐ In neither my home country nor Canada: (Please specify)

7. Did you prepare in advance for the IELTS? (Please circle) **YES / NO**

If **YES**, please answer questions **6** to **11**.

If **NO**, please answer questions **9** and **10**.

8. How did you prepare for the test? **Check (✓) all that apply.**

☐ I used official IELTS preparation books / tests.

☐ I used an unofficial preparation book / test.

☐ I took a part-time instructed IELTS preparation course.

☐ I took a full-time instructed IELTS preparation course.

☐ I took an instructed general English course.

☐ I used free online practice materials (including videos / online courses).

☐ I used paid online practice materials (including videos / online courses).

☐ I searched online forums (e.g., Facebook groups, Reddit) for information on how to take the test.

☐ I talked to friends who knew about / had taken the test.

☐ I spent time with English-speaking friends.

☐ I watched TV / watched movies / listened to radio programs / listened to podcasts in English.

___ I used authentic materials such as newspapers, academic textbooks.

___ Other (please specify):

9. Please rate how much you think your preparation method(s) helped your performance on the test (1 = not at all, 6 = a lot).

Official IELTS preparation books / tests 1 2 3 4 5 6

Unofficial test preparation books / tests 1 2 3 4 5 6

Part-time instructed IELTS preparation course 1 2 3 4 5 6

Full-time instructed IELTS preparation course 1 2 3 4 5 6

Instructed general English course. 1 2 3 4 5 6

Free online practice materials / courses (including videos)

1 2 3 4 5 6

Paid online practice materials / courses (including videos)

1 2 3 4 5 6

Online forums (e.g., Facebook, Reddit) 1 2 3 4 5 6

Talking to friends who knew about the test. 1 2 3 4 5 6

Spending time with English-speaking friends. 1 2 3 4 5 6

Watching TV / movies / listening to radio programs / podcasts in

English.

1 2 3 4 5 6

Using authentic materials such as newspapers, academic textbooks

Other (please specify) _____

1 2 3 4 5 6

10. Please rate how much you think your preparation method(s) helped improve your confidence when taking the test (1 = not at all, 6 = a lot)

Official IELTS preparation books / tests 1 2 3 4 5 6

Unofficial preparation books / tests 1 2 3 4 5 6

Part-time instructed IELTS preparation course	1	2	3	4	5	6
Full-time instructed IELTS preparation course	1	2	3	4	5	6
Instructed general English course.	1	2	3	4	5	6
Free online practice materials / courses (including videos)	1	2	3	4	5	6
Paid online practice materials / courses (including videos)	1	2	3	4	5	6
Online forums (e.g., Facebook, Reddit)	1	2	3	4	5	6
Talking to friends who knew about the test.	1	2	3	4	5	6
Spending time with English-speaking friends.	1	2	3	4	5	6
Watching TV / movies / listening to radio programs / podcasts in English.	1	2	3	4	5	6
Using authentic materials such as newspapers, academic textbooks	1	2	3	4	5	6
Other (please specify) _____	1	2	3	4	5	6

11. How important do you think the following skills are for attaining a high IELTS test score? Please **rank the categories 1 to 6**, with **1** being ‘most important’ and **6** being ‘least important’.

Speaking: ____
 Reading: ____
 Listening: ____
 Writing: ____
 Grammar ____
 Vocabulary ____

12. What IELTS test score did you attain for each section? (**Only answer this question if you are comfortable doing so**)

Total IELTS score: _____

If you remember your scores for each section of the test, please write them below.

Reading: _____

Listening: _____

Speaking: _____

Writing: _____

13. Overall, how much do you think taking the IELTS test prepared you for your studies? (**1 = not at all, 6 = a lot**)

1 2 3 4 5 6

14. Please describe the **feelings** you experienced while taking the different sections of the IELTS test (e.g., did you feel confident, nervous, prepared, anxious, confused, anything else? Why?)

Speaking section:

Listening section

Reading section:

Writing section:

If you are interested in participating in a short (30 minute) interview about your experience of preparing for the TOEFL, please contact me at ciaranmccullough@cmail.carleton.ca. The first 2 email responses for past TOEFL will be accepted for this phase of the study, and will be compensated for their participation with a \$20 Amazon Canada voucher. Thank you for your participation!

Appendix E: Questionnaire Consent Form

Questionnaire Consent Form

Title: Gatekeepers of Aspiration: Researching the effects of test preparation techniques on test-taker performance and confidence in high-stakes language tests.

Date of ethics clearance: February 27, 2018

Ethics Clearance for the Collection of Data Expires: February 28, 2019

I _____, choose to participate in a study on test preparation techniques for high-stakes language tests. This study aims to investigate the effect of different test preparation techniques on test scores and test-taker confidence during the test. **The researcher for this study is Ciaran McCullough in the School of Linguistics and Language Studies.** He is working under the supervision of Dr. Janna Fox in the School of Linguistics and Language Studies.

This study involves completing a questionnaire which will ask you for some personal details (your name, age, home country, languages you speak), which test you took (TOEFL iBT or IELTS), why you chose to take this test, where you took the test, your background of English study, preparation techniques you employed to prepare for the test, your opinions of the importance of different skills in achieving a desirable test score, your actual test scores, and your personal experience of taking the test.

You have the right to end your participation in the study at any time, for any reason, up until April 15, 2018. You can withdraw by phoning or emailing the researcher or the research supervisor. If you withdraw from the study, all information you have provided will be immediately destroyed. You may decline to answer any questions on the questionnaire that you do not wish to answer.

I will take precautions to protect your identity and data. This will be done by keeping all questionnaires in a locked filing cabinet at Carleton, and anonymizing the data during digital tabulation. Research data will only be accessible by the researcher and the research supervisor. Once the project has been completed, the paper questionnaires will be destroyed, and all (anonymized) electronic research data will be kept for five years and potentially used for other research projects on this same topic. At the end of five years, all research data will be securely destroyed (electronic data will be erased).

As a token of appreciation, you will receive a small gift (value: \$5). This is yours to keep, even if you withdraw from the study.

If you would like a copy of the finished research project, you are invited to contact the researcher to request an electronic copy which will be provided to you.

The ethics protocol for this project was reviewed by the Carleton University Research Ethics Board, which provided clearance to carry out the research. If you have any ethical concerns with the study, please contact Dr. Andy Adler, Chair, Carleton University Research Ethics Board-A (by phone at 613-520-2600 ext. 2517 or via email at ethics@carleton.ca).

Researcher contact information:

Ciaran McCullough
School of Linguistics and
Language Studies

Carleton University
Tel: (613) 323-3013
Ext. 2046
Email: ciaranmccullough@cmail.carleton.ca
janna.fox@carleton.ca

Supervisor contact

Dr. Janna Fox

School of Linguistics
and Language
Studies

Carleton University
Tel: (613) 520-2600

Email:

Signature of participant

Date

Signature of researcher

Date

Appendix F: Email Response to ESLA Student Interview Volunteers

Email Response to Interview Volunteers (ESLA Students)

Dear [ESLA student],

Thank you for volunteering to be interviewed as part of my research into test-preparation techniques used by students preparing to take the TOEFL or IELTS tests. I am very interested in hearing what you have to say, and learning more about how you prepared to take the test. I hereby invite you for interview on any of the following dates:

[dd/mm/yy at: TIME]

[dd/mm/yy at: TIME]

Interviews will be conducted in the graduate lounge of the Applied Linguistics and Discourse Studies department in Paterson Hall. At the agreed time, I will meet you outside Paterson Hall and we will proceed to the lounge to commence the interview. It will take around 30 minutes, and refreshments will be provided. You will also receive an Amazon Canada voucher to the value of \$25 for your participation. I will ask you questions about your language study background and your test preparation practices, as well as a few other topics. Your voice will be recorded, and after the study is complete your data will be securely deleted and no identifying information will be released.

The ethics protocol for this project was reviewed by the Carleton University Research Ethics Board, which provided clearance to carry out the research. (Clearance expires on: insert date here.)

If you have any ethical concerns with the study, please contact Dr. Andy Adler, Chair, Carleton University Research Ethics Board-A (by phone at 613-520-2600 ext. 2517 or via email at ethics@carleton.ca).

Please let me know when you are available for interview. I look forward to meeting with you.

Best regards,

Ciaran McCullough

Appendix G: Participant Pseudonym Options

Male	Female	Neutral
Esteban	Pearl	Bailey
Brett	Regenia	Robyn
Forest	Anastasia	Drew
Blaine	Gisele	Dylan
Lucas	Britta	Alex
Jose	Emily	Aiden
Casey	Loree	Hillary
Edwin	Kimberley	Harper
Sung	Tisha	Jordan
Darrell	Sondra	Logan
Percy	Lacie	Riley
Jaime	Odette	Rowan
Rodrigo	Julissa	Mika
Josiah	Shirely	Ashley
Valentin	Tonette	Angel
Micheal	Sherrill	Darcy
Eddie	Maida	Mel
Paris	Melany	Hunter
Warner	Lupita	Jude
Damon	Chan	
Cyril	Cher	
Rodney	Elida	
Sergio	Aileen	
Nelson	Fay	
Eloy	Gidget	
Jarrett	Cathey	
Carmen	Vashti	
Jerrell	Dagny	
Maxwell	Lakeesha	
Harley	Nelly	

Appendix H: Semi-structured Interview Protocol

TOEFL iBT / IELTS STUDY (Researcher: Ciaran McCullough)

Semi-structured Interview Questions

1. What's your name?
2. How are you today?
3. Where are you from?
4. What language do you speak at home?
5. What subject do you study (or intend to study) at Carleton?
6. How long have you been studying English?
7. What level are you in the ESLA program? Are you enjoying it?
8. Did you take the TOEFL iBT or IELTS? Why did you take this specific test?
9. How did you prepare to take the TOEFL iBT / IELTS test?
10. Why did you choose to use these preparation techniques?
11. How do you think your preparation techniques helped you to succeed on the test?
12. How did you feel during the different sections of the test?
 - Speaking
 - Reading
 - Listening
 - Writing
13. What test-preparation techniques do you feel helped improve your confidence on the different sections of the test?
 - Speaking
 - Reading

- Listening
- Writing

14. Which of your skills do you believe improved the most as a result of your test preparation?

- Speaking
- Reading
- Listening
- Writing
- Grammar
- Vocabulary

15. What are your general opinions of the test you took? What did you think of the different sections?

16. Is there anything you would like to ask me?

Appendix I: Interview Consent Form

Interview Consent Form

Title: Gatekeepers of Aspiration: Researching the effects of test preparation techniques on test-taker performance and confidence in high-stakes language tests.

Date of ethics clearance: TBD

Ethics Clearance for the Collection of Data Expires: TBD

I _____, choose to participate in a study on test preparation techniques for high-stakes language tests. This study aims to investigate the effect of different test preparation techniques on test scores and test-taker confidence during the test. **The researcher for this study is Ciaran McCullough in the School of Linguistics and Language Studies.** He is working under the supervision of Dr. Janna Fox in the School of Linguistics and Language Studies.

This procedure takes the form of an interview, and will take around 30 minutes to complete. This interview involves answering questions about your language study background, your experience taking the TOEFL iBT / IELTS, and the preparation techniques you used when preparing for the TOEFL iBT / IELTS test. Specifically, you will be asked to expand on responses to the TOEFL iBT / IELTS questionnaire you completed previously. You are encouraged to expand upon your answers to this questionnaire in as much detail as you can. Your responses will be audio recorded and transcribed by the researcher for analysis. You personally are not the subject of the study, rather, your test preparation techniques are of interest in this research.

You have the right to end your participation in the study at any

time, for any reason, both during the interview and up until April 15, 2018. You can withdraw by phoning or emailing the researcher or the research supervisor. If you withdraw from the study, all information you have provided will be immediately destroyed. You may decline to answer any questions on the questionnaire that you do not wish to answer.

I will take precautions to protect your identity and data. This will be done by keeping all audio files in a password-protected recording device and password-protected computer files, and anonymizing the data during digital tabulation. Your response will be assigned a pseudonym (false name), and you will not be in any way identifiable in any published research. Research data will only be accessible by the researcher and the research supervisor. Once the project has been completed, all audio files will be deleted, and all (anonymized) transcribed data will be kept for five years and potentially used for other research projects on this same topic. At the end of five years, all research data will be irreversibly erased.

As a token of appreciation, you will receive an Amazon Canada gift voucher to the value of \$20. This is yours to keep, even if you withdraw from the study.

If you would like a copy of the finished research project, you are invited to contact the researcher to request an electronic copy which will be provided to you.

The ethics protocol for this project was reviewed by the Carleton University Research Ethics Board, which provided clearance to carry out the research. If you have any ethical concerns with the study, please contact Dr. Andy Adler, Chair, Carleton University Research Ethics Board-A (by phone at 613-520-2600 ext. 2517 or via email at ethics@carleton.ca).

Researcher contact information:

Ciaran McCullough
School of Linguistics and Language Studies

Carleton University
Tel: (613) 323-3013
Ext. 2046

Supervisor contact

Dr. Janna Fox
School of Linguistics
and Language
Studies

Carleton University
Tel: (613) 520-2600

Email: CiaranMccullough@cmail.carleton.ca
JannaFox@cunet.carleton.ca

Email:

Signature of participant

Date

Signature of researcher

Date

Appendix J: Qualitative Coding Scheme

Opinions of Test

General

Negative

Neutral

Positive

Listening

Negative

Neutral

Positive

Reading

Negative

Neutral

Positive

Speaking

Negative

Neutral

Positive

Writing

Negative

Neutral

Positive

Performance on Test

General

Negative

Neutral

Positive

Listening

Negative

Neutral

Positive

Reading

Negative

Neutral

Positive

Speaking

Negative

Neutral

Positive

Writing

Negative

Neutral

Positive

Preparation Practices

Authentic texts
Instructed general English or EAP courses
Instructed test preparation courses
 Full-time
 Part-time
 Online
 In-person
Media
Online materials
 Online forums and social media
 Online published content (e-books, online text resources)
 Online sample tests
 Webinars and videos
Skills practice
Spending time with English speakers
Talking to friends who had taken the test
Test preparation books and sample tests

Reasons for Choosing Test

Available
Better
Common
Content
Convenient
Easier
Familiar
Format
 Computer
 Human examiner
 Paper
Free to take
Popular
Recommended
Required

Strategies

Subjective experience during test

General
 Negative
 Neutral
 Positive
Listening
 Negative
 Neutral
 Positive
Reading
 Negative
 Neutral

Positive
Speaking
Negative
Neutral
Positive
Writing
Neutral
Positive

Speededness

Repeated test taking

Appendix K: Primary Researcher Initial Coding Sample

I: Right, okay. So, did you prepare to take test? What kind of things did you do, if you did anything?

S: Well, at the time I was working really hard, I was teaching, and so I didn't have much time to get prepared properly, unfortunately, but I wanted [laughs], so the only thing I could do was to do some tests online, to have a look at the previous tests and the question types, and... I watched some webinars... about the content of the test, and I, uh did some listening tests which were online again, but I made a mistake, I only did those listening tasks which were prepared for the general, uh... IELTS...

question types / task

- Preparation Practices ✓
- Online sample tests ✓
- Webinars & Videos ✓
- Online sample tests ✓
- Online published content
- ★ Add to tree: Test Format
- Skills Practice (Listening)

I: Not the academic?

S: Not the academic, unfortunately I realized the difference during the test! [laughs]

I: Okay!

S: It was much more difficult than I expected, so I cannot say I prepared really well, but I did my best at least. For reading I did not do any preparation because I thought that was going to be the easiest part of it, which turned out to be the opposite [laughs]! But that was to my surprise that I got an 8.5 from reading, because it was really hard, maybe you have another question for that, maybe I'm too early to give you this explanation...

Stress + Anxiety + Motivation:

- Performance on test
 - General / neutral
 - Listening = Negative ✓
- Performance on test ✓
 - reading / positive ✓

I: No, no, go ahead...

S: The... the part that I was most... uh, let's say scared [laughs] was listening... and writing, because I was

RIG IMPORTANT CONCEPTS

Test Anxiety / Stress

Subjective experience during test:

- listening / negative ✓
- writing / negative X

= Stress = long-term Anxiety - sharp, sudden.

Appendix L: Second Coder Initial Coding Sample

Coder 2

I: Right, okay. So, did you prepare to take test? What kind of things did you do, if you did anything?

S: Well, at the time I was working really hard, I was teaching, and so I didn't have much time to get prepared properly, unfortunately, but I wanted [laughs], so the only thing I could do was to do some tests online, to have a look at the previous tests and the question types, and... I watched some webinars... about the content of the test, and I, uh did some listening tests which were online again, but I made a mistake, I only did those listening tasks which were prepared for the general, uh, IELTS.

I: Not the academic?

S: Not the academic, unfortunately I realized the difference during the test! [laughs]

I: Okay!

S: It was much more difficult than I expected, so I cannot say I prepared really well, but I did my best at least. For reading I did not do any preparation because I thought that was going to be the easiest part of it, which turned out to be the opposite [laughs]! But that was to my surprise that I got an 8.5 from reading, because it was really hard, maybe you have another question for that, maybe I'm too early to give you this explanation...

I: No, no, go ahead...

S: The... the part that I was most... uh, let's say scared [laughs] was listening... and writing, because I was

Handwritten Notes:

- 3 preparation ✓
- personal prep
- online questions webinars ✓
- online do sample test ✓
- review sample test question types ✓
- mistake in prep for general IELTS
- 3 skills practice ✓
- content X
- Impact of test prep on performance
- Performance 2 ✓
- Diff (Listening) 2 ✓
- Listening Neg. ✓
- Reading
- most diff highest score
- Factor
- anxiety - ~~bad~~
- stress
- Listening ✓
- help ✓
- 2 Perform Reading + ✓