UNDERSTANDING TEST PREPARATION PHENOMENON THROUGH CHINESE STUDENTS’ JOURNEY TOWARDS SUCCESS ON HIGH-STAKES ENGLISH LANGUAGE TESTS

by

Jia Ma

A thesis submitted to the Graduate Program in Education in conformity with the requirements for the Degree of Doctor of Philosophy

Queen’s University
Kingston, Ontario, Canada
September, 2017

Copyright © Jia Ma, 2017
Abstract

Many Chinese students take test preparation courses with an aim to achieve success on English language tests because of the high-stakes consequences associated with test scores. However, contrary to the massive test preparation industry in China and the huge claim of achieving high scores, there is very limited empirical evidence to help us understand how students are prepared and if the claim can be substantiated. Therefore, the purpose of my doctoral dissertation was to understand Chinese students’ test preparation for high-stakes English language tests, through examining the nature, the perceived effects, and the perceived value of test preparation practices for the Test of English as a Foreign Language Internet-based Test (TOEFL iBT) and the College English Test Band 4 (CET4).

This study employed a multiple-stage, multiple-method design to collect data from three participant groups, including students, teachers, and administrators. Three research sites reflecting the stages of test preparation experience were included: ‘currently’ taking test preparation courses, recently completing test preparation courses, and achieving success on tests. Data were collected from classroom observations and interviews with 18 students, eight teachers and three administrators, questionnaires with 534 students, and documents related to test preparation courses and test preparation centre’s administration.

Test familiarization, English skill improvement, and stress management were three major practices that helped Chinese students achieve success; however, the factors related to participants’ characteristics, perceptions, and their situated context interacted with each other to shape these practices. Although no immediate substantial test score gains were found, the practices that targeted at English skills were perceived to gradually improve test scores. The trustworthiness, the efficacy and the reorientation of the test preparation practices were valued to
help with achieving success, although the extent of these aspects varied to individual students. Drawing on Green’s (2007) washback model, the findings contributed to the understanding of test preparation phenomenon that the variability in the nature of the test preparation practices influence the direction that test preparation practices work on test scores and English learning, and the perceived value are related to what students experienced (nature) and what they achieved (perceived effects).
Acknowledgements

“You raise me up, so I can stand on mountains  
You raise me up to walk on stormy seas  
I am strong when I am on your shoulders  
You raise me up to more than I can be”  
   Brendan Graham

I would like to use the lyrics of You raise me up to start the acknowledgements because this song speaks to the supports that many of you gave me throughout my journey towards completing this dissertation.

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Liying Cheng, for her mentorship, support, and care for both my academic study and personal life. Everything she taught me through academic lessons and life lessons raised me up to keep confident and carry on. Most of all, she helps me to understand, “Everything comes to us that belongs to us if we create the capacity to receive it.”

I would also like to thank my committee members, Dr. Don Klinger and Dr. Lyn Shulha, for teaching me that the key to academic research is to think critically and approach precisely. I will not forget the conversations with them that pushed me to go beyond my comfort zone to reach the ‘truth’ I wanted to find out in my study. I understand what they wanted to tell me through these conversations is, “If you shut your door to all errors, truth will be shut out.”

Thank you to my examining committees, Dr. Janna Fox, Dr. Pamela Beach, Dr. Jan Mennell, and Dr. Stanka Fitneva, for the thought-provoking questions and comments during the defense and in the dissertation that helps me to bring new perspectives and research methodology for further inquiries on understanding test preparation phenomena.

I would like to thank my friends, who held my hands all the way through, Allison, Yi, Wei, King, Xiaomei, and Judy. At the moments when I needed them, they were always there to give
me encouragement, support, and company. They helped me to believe, “The real friendship is like fluorescence; it shines better when everything has darken.”

Thanks also go to my research participants for their generosity in sharing their experiences in test preparation. More importantly, I learned from them that the path to success is to dream, believe, and achieve.

I want to thank the Social Sciences and Humanities Research Council of Canada, the Educational Testing Service, and The International Research Foundation for English Language Education for funding my research.

Last, my heartfelt thanks go to my loving family: my parents and parents-in-law, my husband, Yu, my two lovely kids, Vanessa and Ryan, for their unfailing love and unconditional support that keep me strong, and for their never doubting belief that I can be “more than I can be.”
Table of Contents

Abstract........................................................................................................................................... ii

Acknowledgements......................................................................................................................... iv

Table of Contents ............................................................................................................................. vi

List of Tables ...................................................................................................................................... xi

List of Figures ..................................................................................................................................... xii

Chapter 1 Introduction ..................................................................................................................... 1

  Context: Test Preparation in China................................................................................................. 2

    Development of the Test Preparation Phenomenon................................................................. 2

    One Huge Claim and A Central Dilemma............................................................................. 4

  Research Purpose ......................................................................................................................... 5

  Research Questions ....................................................................................................................... 5

  Rationale ...................................................................................................................................... 7

    Current Test Preparation Research on the TOEFL and the CET .......................................... 7

    Insights from Existing Test Preparation Research ............................................................. 10

    Dissertation Overview ............................................................................................................. 13

Chapter 2 Literature Review .......................................................................................................... 15

  Test Preparation: Definitions and Scope..................................................................................... 15

    Three Key Terms that Define Test Preparation................................................................. 16

    Scope of Test Preparation in the Current Study ................................................................. 19

    Empirical Evidence of Test Preparation Research ............................................................ 21
Participants .......................................................... 22
Process ................................................................. 30
Products ............................................................... 36
Section Summary ......................................................... 40
Theoretical Frameworks for Test Preparation Research .................. 41
Theoretical Frameworks of Existing Test Preparation Research .......... 41
Theoretical Framework for Current Study: Green’s Washback Model ..... 47
Chapter Summary .......................................................... 50
Chapter 3 Methodology ..................................................... 52
Overall Research Design .................................................... 52
Research Design ........................................................... 54
Research Sites ............................................................ 55
Site 1: A Commercial Test Preparation Centre in Beijing—Centre L .......... 56
Site 2: Eight Universities and Colleges in China ................................ 57
Site 3: A Canadian University ................................................. 58
Participants ............................................................... 59
Students .................................................................. 60
Teachers .................................................................. 63
Administrators ............................................................ 64
Section Summary ............................................................ 65
Instruments ................................................................. 65
Instruments for Site 1 ...................................................... 66
Instruments for Site 2 ...................................................... 68
Instrument for Site 3 ....................................................... 72
Section Summary ............................................................ 72
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Collection</td>
<td>73</td>
</tr>
<tr>
<td>Documents</td>
<td>74</td>
</tr>
<tr>
<td>Classroom Observation</td>
<td>75</td>
</tr>
<tr>
<td>Interviews</td>
<td>76</td>
</tr>
<tr>
<td>Questionnaires and Test Scores</td>
<td>78</td>
</tr>
<tr>
<td>Primary Data and Supplementary Data</td>
<td>79</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>81</td>
</tr>
<tr>
<td>First Cycle Analysis</td>
<td>81</td>
</tr>
<tr>
<td>Second Cycle Analysis</td>
<td>84</td>
</tr>
<tr>
<td>Chapter Summary</td>
<td>86</td>
</tr>
<tr>
<td>Chapter 4 Results</td>
<td>88</td>
</tr>
<tr>
<td>Nature of the Test Preparation Practices</td>
<td>89</td>
</tr>
<tr>
<td>Characteristics: Who were Engaged in Preparing for the Test?</td>
<td>90</td>
</tr>
<tr>
<td>Perceptions: Why did the Participants Engage in Test Preparation?</td>
<td>93</td>
</tr>
<tr>
<td>Context: How did the Participants Choose Their Test Preparation Practices?</td>
<td>99</td>
</tr>
<tr>
<td>Practices: What did the Participants Do to Prepare for the Test?</td>
<td>106</td>
</tr>
<tr>
<td>Section Summary</td>
<td>117</td>
</tr>
<tr>
<td>Perceived Effects of the Test Preparation Practices</td>
<td>118</td>
</tr>
<tr>
<td>Perceived Effects on the TOEFL iBT Test Scores</td>
<td>119</td>
</tr>
<tr>
<td>Perceived Effects on the CET4 Test Scores</td>
<td>120</td>
</tr>
<tr>
<td>Value of the Test Preparation Practices</td>
<td>131</td>
</tr>
<tr>
<td>Trustworthiness</td>
<td>131</td>
</tr>
<tr>
<td>Efficacy</td>
<td>134</td>
</tr>
<tr>
<td>Reorientation</td>
<td>137</td>
</tr>
<tr>
<td>Chapter Summary</td>
<td>140</td>
</tr>
</tbody>
</table>
Chapter 5 Discussion .................................................................................................................. 142

Discussion of the Three Research Questions ........................................................................ 142

Research Question 1: Nature of the Test Preparation Practices ............................................. 142
Research Question 2: Perceived Effects of the Test Preparation Practices ............................. 149
Research Question 3: Value of the Test Preparation Practices .............................................. 153

Discussion by Research Purpose and Green’s Washback Model ......................................... 156

Three Dimensions in Green’s Washback Model .................................................................. 157
Direction .................................................................................................................................. 158
Variability ................................................................................................................................. 160
Intensity .................................................................................................................................... 161
Contribution to Green’s Washback Model ............................................................................. 162

Implications ............................................................................................................................. 164
Implications for practice ......................................................................................................... 164
Implications for research ......................................................................................................... 165

Limitations ............................................................................................................................... 167

Recommendations for Future Research ................................................................................ 169

Conclusion .................................................................................................................................. 170

References ................................................................................................................................. 172

Appendix A Letter of Approval from the General Research Ethics Board .............................. 186

Appendix B. 1 Letter of Information/Consent Form—Administrators and Teachers at Test Preparation Centres ........................................................................................................... 187

Appendix B. 2 Letter of Information/Consent Form—Students at Test Preparation Centres in China .................................................................................................................................. 191
Appendix B. 3 Letter of Information/Consent Form—Students at Universities in China....... 194
Appendix B. 4 Letter of Information/Consent Form—Students at Universities in Canada ...... 197
Appendix C. Document Review Form............................................................................. 200
Appendix D. Green’s (2007b) Observation Schedule ....................................................... 201
Appendix E. Interview Guide for Student Participants at Centre L ................................. 202
Appendix F. Interview Guide for Teacher Participants at Centre L................................. 203
Appendix G. Interview Guide for Administrator Participants at Centre L....................... 204
Appendix H. 1 Questionnaire of CET4 Test Preparation Experience (English) ............... 205
Appendix H. 2 Questionnaire of CET4 Test Preparation Experience (Chinese) ............... 209
Appendix I. Interview Guide for Student Participants at a Canadian University ............. 213
Appendix J. Example of Descriptive Field Notes of Classroom Observation.................. 215
Appendix K. Example of Reflective Field Notes of Classroom Observation .................... 218
Appendix L. Example of Integrating Interview Data and Questionnaire Data ................. 219
Appendix M. Initial Versions of Figure 2.......................................................................... 220
List of Tables

Table 1. Overview of Multiple-Stage Multiple-Method Research Design ........................................ 55
Table 2. Background Information of the Student Participants at Centre L ....................................... 61
Table 3. Profile of Twelve Students at a Canadian University ...................................................... 62
Table 4. Course Structure by the Teacher Participants at Centre L .................................................. 63
Table 5. Profile of Eight Teacher Participants at Centre L ............................................................ 64
Table 6. Profile of the Administrator Participants at Centre L ....................................................... 64
Table 7. Test Description of the CET4 ............................................................................................ 71
Table 8. Summary of the Instruments by Three Research Sites ...................................................... 73
Table 9. Summary of the Collected Documents at Centre L .......................................................... 75
Table 10. Summary of Classroom Observation Data Collection .................................................... 76
Table 11. Summary of Primary and Supplementary Data Collected from Three Sites ................. 79
Table 12. Frequency of CET4 Test Preparation Course Types by Status of Test Repeaters ........ 102
Table 13. Frequency of Timing of Taking CET4 Test Preparation Courses by Status of Test Repeaters ......................................................................................................................... 104
Table 14. Descriptives of the Test-Specific Practices by CET4 Test Section (n=534) ............... 110
Table 15. Descriptives of General Practices for English Skills in CET4 Test Preparation Courses ........................................................................................................................................ 112
Table 16. Steps of Factor Analyses ................................................................................................. 122
Table 17. Factor Loadings of the CET4 Test Preparation Practices (n=534) ................................. 123
Table 18. Correlations between Independent Variables and Dependent Variables (n=534) .... 126
Table 19. Summary of Five Multiple Regression Analyses (n=534) .............................................. 129
Table 20. Thematic Findings of Students’ Perceived Value of the Test Preparation Practices .. 131
List of Figures

Figure 1. Green’s washback model (2007)........................................................................................................... 48

Figure 2. Themes, categories, and codes reflecting the nature of the test preparation practices 117
Chapter 1 Introduction

The offices and classrooms of test preparation Centre L are in the Central Business District (CBD) in Beijing. It is a dream world for youth. Stepping out of the subway, I spot many commercial posters presenting the latest fashion trends to attract young people. Taking test preparation courses is clearly also a trend, because among those posters is a distinctly striking one for test preparation at Centre L. Its message encourages youth to come together, and to achieve success together, through test preparation.

I found myself wondering: Is test preparation trendy because success on the TOEFL is associated with chasing dreams? The students I observed at Centre L felt that this was the case. They were among the thousands of the students who were preparing for overseas study, and test preparation for the TOEFL was the first step on their journey to chase dreams. (Observation notes, June 25, 2013)

In May, 2013, a movie called American Dreams in China (“中国合伙人” in Chinese) attracted overwhelming public attention, and its box office sales hit 77.44 million US dollars in China (Lei, 2013). The plot of this movie was based on the true story of the New Oriental School (New Oriental)—a widely recognized private educational institution in China. The movie tells a dream-chasing story of how the New Oriental, with its primary focus on test preparation, became the first Chinese educational institution to trade on the overseas stock market. Viewers applauded the youth, the dreams, and the powerful motto of the New Oriental School: “Hew a stone of hope out of a mountain of despair and you can make your life a splendid one.” The successful, dream-chasing story of the New Oriental School highlights the prominent phenomenon of test preparation and the rapid development of this profitable industry in China.

---

1 Centre L is a pseudonym given to a popular commercial test preparation centre in Beijing. Centre L was one of the research sites of this study.

2 “TOEFL” is the acronym of the Test of English as a Foreign Language.
**Context: Test Preparation in China**

Test preparation is a historically rooted and educationally accepted phenomenon in the Chinese context, which has paved the way for test preparation for English language tests in particular to become a resounding business success in China. In the following section, I trace the historical roots of test preparation practices, describe the massive scale of the test preparation industry and the factors contributing to the industry’s success, and highlight the gap between the magnitude of the test preparation industry and the relative dearth of test preparation research in China. This information sets the background necessary to understand both the features of the test preparation context in China, and the necessity of researching the test preparation phenomenon in this particular context.

**Development of the Test Preparation Phenomenon**

Test preparation is rooted in Chinese society’s long history of using testing as a selection mechanism. This phenomenon dates back to the Chinese imperial examinations, which were used in China for over a thousand years to make high-stakes decisions around the selection of government officials (Cheng, 2008). The result is a culture wherein “people read simply for the sake of the civil service examinations” (Jin & Cortazzi, 2006, p. 12), while the inherent value of what is to be learned is not always considered.

Test preparation for English language tests became industrialized in China in the late 1980s and early 1990s, as part of the larger English language training industry (Wang, 2007; Xu, 2007). Today, the continuing presence of the high-stakes consequences of tests in Chinese society—such as those administered around English language—drives Chinese students to commercial test preparation centres that promise better test performance. Consequently, test preparation centres are thriving (Matoush & Fu, 2012). Indeed, the estimated value of this industry is currently about
4.5 billion US dollars and is expected to grow at a rate of 12–15% over the next few years, with
test preparation for English language tests contributing a significant share (Gamlam, 2016; Ji,
2011).

Test preparation centres are established countrywide within English language training
schools, providing dedicated courses and services to prepare students for various English
language tests. By 2013, English language training schools accounted for approximately one-
third of the 141,100 commercial educational training schools in China—or around 50,000
English training schools in roughly 700 cities—ranging from small independent businesses to
large chains such as the New Oriental School (Gamlam, 2016; Ji, 2011; Hu, 2013). It is no
exaggeration, therefore, to say that thousands of commercial test preparation centres have
sprouted across China over the last three decades.

The rapid expansion and profitability of the test preparation industry can be attributed to
two factors: the tremendous importance of English language education, and the high-stakes
consequences of English language tests in China. Since China carried out its Economic Reform
in 1978, Chinese students have been required to take English as one of three compulsory
subjects—along with mathematics and Chinese—from primary school through postgraduate
programs. More importantly, achieving high scores on the many English language tests students
are required to take throughout their educational experience is considered critical to their future
success (Cheng, 2008). As such, while the intended goal of English tests is to measure students’
English language proficiency, test performance has high-stakes consequences for students with
respect to determining future opportunities such as entrance into academic programs and the
completion of academic degrees (e.g., Cheng, 2008; Mei, 2016; Sun, 2016; Zheng & Cheng,
2008). The need to succeed in English tests, to access further academic and life opportunities, is
therefore expected to be a major driver for Chinese students who choose to enroll in commercial
test preparation courses.

One Huge Claim and A Central Dilemma

To tap into this need to succeed, the potential benefit of achieving high scores on English
tests has been strongly promoted by test preparation centres: “Our test preparation courses focus
on quality instruction and test-taking techniques designed to help [Chinese] students achieve
high scores on the most widely used admissions tests [in China]” (New Oriental Education,
2015, italics added for emphasis). The huge, powerful claim of achieving high scores through
test preparation has been repeatedly sold to and believed by Chinese students; therefore, the
students go to test preparation centres after school, over weekends, and during holidays to
achieve the benefits these centres have promised. Test preparation centres have become a
popular destination choice for Chinese students driven to achieve high scores, and commercial
test preparation courses remain popular even though they are costly and require a significant
commitment of time and effort to complete.

Despite this widespread phenomenon of Chinese students seeking success on high-stakes
English language tests through taking test preparation courses, empirical studies focusing on this
phenomenon—in the Chinese context—are surprisingly limited (e.g., Chau, 2008; Liu, 2014; Yu,
2012; Xie, 2010). Consequently, we are facing the central dilemma of insufficient empirical
evidence to reach a thorough understanding of the impact of this phenomenon. Specifically, we
do not know how Chinese students are prepared during test preparation courses and at these
centres, nor whether the claim of test score gains these centres have promised can be
substantiated.
Hamp-Lyons (1998) argues that test preparation practices for high-stakes tests can be better understood through research. Many researchers studying language testing in countries other than China have addressed the understanding of this phenomenon through examining the aims, methodology, and effects of test preparation practices (e.g., Hayes & Read, 2004; Green, 2007; Wall & Horák, 2006; 2008). The results of such studies can be used to contribute to understanding the test preparation phenomenon in the Chinese context.

**Research Purpose**

The purpose of this study is to understand the phenomenon of Chinese students taking the test preparation courses for two high-stakes English language tests—the Test of English as a Foreign Language Internet-based Test (TOEFL iBT) and the College English Test Band 4 (CET4)—at commercial test preparation centres in China. Specifically, the understanding of this phenomenon focuses on three interrelated aspects of test preparation practices claimed by test preparation centres to help Chinese students succeed: the nature of test preparation practices, the effects of these practices, and the value of these practices.

**Research Questions**

This study focuses on the test preparation practices of the TOEFL iBT and the CET4 test preparation courses in China and asks the following three research questions:

1. What is the nature of test preparation practices employed to help Chinese students achieve success in the two tests?
2. What are the perceived effects of these test preparation practices on Chinese students’ scores on the two tests?
3. How do Chinese students perceive the value of these practices with respect to helping them achieve success?
To better understand the nature of test preparation practices (RQ1) this study sought to identify the characteristics of the participants’ involved in test preparation practices, as well as how these participants employed specific practices. To shed light on the perceived effects of these practices (RQ2) the focus was on the relationship between these practices and the students’ test scores. Finally, to understand the perceived value (RQ3), this research examined how students judge the value with respect to these practices both for, and beyond their success on the tests. The results of these three questions work together to create a clearer picture of test preparation phenomenon in China. Specifically, the understanding of the nature of test preparation practices (RQ1) contributes to explaining the relationship between these practices and Chinese students’ test scores (RQ2), and the students’ value judgments about these practices (RQ3) are interpreted in relation to the evidence regarding the nature (RQ1) and the perceived effects of test preparation practices (RQ2).

Central to understanding the phenomenon of Chinese students’ taking test preparation courses is the fact that test preparation is a direct consequence of the influence of high-stakes English language tests on teaching and learning. This influence that a test can have on teachers’ teaching and students’ learning is referred to as “washback” by researchers in applied linguistics and language testing. This study is thus mapped onto Green’s (2007) washback model, which conceptualizes washback in three dimensions: washback direction (positive or negative washback), washback variability (differences among participants with respect to how they are affected), and washback intensity (the degree to which participants are affected). The findings based on the combined evidence of the three research questions provide a thorough understanding of Chinese students’ test preparation for high-stakes English language tests from these dimensions.
Rationale

The rationale for conducting this study is twofold: (1) the limited understanding of test preparation research on two high-stakes English language tests—the TOEFL and the CET in the Chinese context; and (2) the insights from previous test preparation research that direct attention onto the specific aspects of test preparation practices to achieve a clearer understanding of the test preparation phenomenon. The rationale that is articulated from these aspects contributes to contextualizing this study and justifying why the research questions are formulated as they are.

Current Test Preparation Research on the TOEFL and the CET

Among the many high-stakes English language tests that Chinese students undergo, the TOEFL iBT and the CET4 are the two most popular and important. The TOEFL iBT measures the ability of non-native speakers of English to use English language skills in an academic environment (Alderson, 2009). The TOEFL iBT scores are used as indicators of test-takers’ English proficiency to combine reading, listening, speaking and writing skills to perform academic tasks at university level; therefore, many education institutions across world use the TOEFL iBT scores to grant admissions to non-native speakers of English (Alderson, 2009). The TOEFL iBT includes four sections that measures sub-skills of reading, listening, speaking and writing; and in particular the speaking and writing sections include integrated tasks that required test takers to use reading and listening skills in order to complete writing and speaking tasks. Five scores are reported, including a total score and four subscores (i.e., reading, listening, speaking, and writing). Because the TOEFL iBT scores are valid for two years, Chinese university students who plan to study in English-speaking countries usually take the TOEFL iBT in the upper years of undergraduate programs (e.g., third- or fourth-year undergraduate) to ensure
their test scores are with valid periods when they make academic applications to education institutions.

Since its first administration (TOEFL Paper-based test, TOEFL pBT) in China in 1981 (Fingar & Reed, 1982), the TOEFL (either pBT or iBT) has become a critical test for Chinese students because its score is used to determine entry into universities in English speaking countries. Not surprisingly, therefore, TOEFL test preparation courses are extremely popular. Indeed, 70 percent of Chinese students studying at North American universities have taken test preparation courses (e.g., TOEFL test preparation courses) at the New Oriental School alone (Tang, 2010).

At the national test level, all Chinese university/college students are required to take the CET4. The CET4 is to examine the English proficiency of undergraduate students to ensure that Chinese undergraduates reach the required English levels specified in the National College English Teaching Syllabuses (Zheng & Cheng, 2008). The Chinese undergraduates need to complete the required College English Courses Band 1 to 4 in first two years of undergraduate programs before they take the CET4. Therefore, the CET4 test takers are usually upper years of undergraduates in China. The CET4 includes five sections, listening, reading, cloze, writing, and translation; and five scores are reported on the CET4 report card, including a total score and four subscores (i.e., listening, reading, integrative, and writing). Although the CET4 intends to measure the undergraduates’ achievement in English learning after completing the required College English Courses, the CET4 scores are used for many purposes that are not intended by the CET4 test developers, such as one of the requirements for obtaining academic degrees, and for seeking employment opportunities (e.g., Cheng, 2008; Sun, 2016). In 2006, 13 million Chinese students took the CET4 (Jin & Yang, 2006). While exact numbers are not available, the
number of students taking CET4 test preparation courses was a very large portion of the approximately one million students enrolled at the New Oriental School that year (Xu, 2007).

Despite the large number of Chinese students annually enrolling in test preparation courses for the two tests, empirical studies are very limited. The few studies that do exist on the TOEFL (either pBT or iBT) test preparation practices reveal that some practices focus on helping students improve specific English language skills (e.g., speaking skills; Yu, 2012), while others target training students to manage specific test tasks (e.g., writing tasks; Chau, 2008). The subsequent effects of these two types of practices were either positively correlated to test scores or detrimental to English proficiency development, respectively (e.g., Chau, 2008; Liu, 2014; Yu, 2012).

With respect to the CET4, Xie (2008, 2010) found that Chinese students’ perceptions of test-taking skills influenced the test preparation practices they used, and such test preparation practices subsequently produced different effects on resulting test scores. However, Xie did not explicitly state whether the test preparation practices reported by her participants (Chinese university students) were those in the context of commercial test preparation courses.

With these inconsistent findings (e.g., the studies on the TOEFL pBT or iBT) and limited findings (e.g., the studies on the CET4), it is difficult to piece the results together to understand the unprecedented test preparation phenomenon in China. Given the insufficient evidence in the Chinese context, therefore, I turn to empirical test preparation research in the other contexts (e.g., on different testing programs or in different language speaking contexts) to gain insight into what the possible “pieces” related to test preparation practices are, and how these pieces should be addressed and integrated into further research.
Insights from Existing Test Preparation Research

Since its widespread of coaching/test preparation programs beginning from the late 1980s, test preparation has gradually drawn attention of researchers in language testing (e.g., Alderson & Hamp-Lyons, 1996; Hayes & Read, 2004; Tsagari, 2012) and educational studies (e.g., Lai & Waltman, 2008; Messick, 1982; Montgomery & Lilly, 2012; Powers, 1985, 2012). The existing empirical studies in these fields usually focus on two aspects: what the actual test preparation practices are (e.g., Alderson & Hamp-Lyons, 1996; Green, 2007), and what the effects of these practices are (e.g., Elder & O’Loughlin, 2003; Gan, 2009; Powers, 1985).

The major test preparation practices identified in the empirical studies include improving academic skills, training test-taking strategies, drilling for test familiarization, maximizing motivation, and optimizing test anxiety (e.g., Cole, 1982; Ma & Cheng, in press; Messick, 1982). As well, different test preparation courses have been shown to have different combinations of these practices (e.g., Green, 2007; Wall & Horák, 2006, 2008), and researchers from their particular perspective argue for various contributing factor(s) that influence how the participants in the test preparation courses decide and/or use specific test preparation practice(s). These factors mainly include: (1) participants’ characteristics, (2) participants’ perceptions, and (3) context in which participants are situated.

There is the evidence that teachers’ characteristics—including the amount of teaching experience, sociolinguistic background, and teaching beliefs—significantly influences decision-making regarding what test preparation practices are used (e.g., Cheng, 2014; Green, 2007; Mickan & Motteram, 2009; Watanabe, 2004). There is also evidence that students’ characteristics—such as experience of English language learning, levels of English proficiency, and motivation for taking tests and test preparation courses—influences the uptake and
usefulness of test preparation practices that teachers chose (e.g., Doe & Fox, 2011; Ferman, 2004; Gan, 2009). Further, teachers adjust their pedagogy toward “teaching to the test” because of some contextual factors, such as when they face the pressure from education administrators and school policy to have students ready for taking tests, and when they understand the stakes of tests (e.g., Deng & Carless, 2010; Firestone, Monfils, & Schorr, 2004). The possible influences of these multiple factors indicate that test preparation practices could be better understood as the interactions between participants and contextual characteristics (Green, 2006a).

Empirical studies that have tried to determine the effects of test preparation practices on test score gains have provided inconsistent results. Some studies have found that students had significantly better test performance after taking a test preparation course (e.g., Brown, 1998; Elder & O’Loughlin, 2003), while others found no significant advantages when comparing test preparation course takers to their non-course taking peers (e.g., Gan, 2009; Green, 2007; Powers, 1985). These conflicting findings yet again highlight the need for further studies. More importantly, the various characteristics that participants and context bring into test preparation practices may contribute to the different effects reported in these studies. Hence the interrelated relationship between the participants and context of test preparation practices and the subsequent effects of these practices suggest the need for further empirical studies.

In addition to the necessity of further examining the effects of test preparation practices, the scope of test preparation studies needs to be considered. Messick (1982) argues for the necessity of expanding the scope of investigating coaching, which refers to the commercial programs preparing students for a particular test. Messick stresses that the investigation into coaching needs to address its implications for both validity and value. However, the existing
evidence mostly contributes to understanding if and how test preparation practices influence test validity; few empirical studies investigate the value of coaching/test preparation practices.

In addition, most existing studies examine test practices from the teachers’ perspectives, through interviews or observations (e.g., Alderson & Hamp-Lyons, 1996; Wall & Horák, 2006, 2008). Despite being an important stakeholder group, students have rarely had their voices heard (Wall, 2000), in particular with respect to how they perceive the value of the test preparation practices that are claimed to help them succeed. Yu’s study (2012) addressed this research gap and examined Chinese students’ perceived value and usefulness of test preparation practices for TOEFL speaking tasks. He found that students judge usefulness according to the frequency with which teachers use the practices. Further empirical studies are needed to make value judgments regarding test preparation practices, especially with respect to which those students attend to and what the reward is reward for their attention (e.g., Green, 2006b). As Yu (2012) states, studies along this line can contribute to enhancing the understanding of the relationship between testing (e.g., test performance), learning (e.g., language skills development), and teaching (e.g., test preparation practices), and can also provide supporting empirical evidence to the test validity from test takers’ perspective (e.g., Fox & Cheng, 2007; Moss, Girard, & Haniford, 2006).

The findings of these existing studies—including the contributing factors that shape test preparation practices, the multiple effects of the practices on students’ test scores, and the unexplored value implications of the practices—suggest that further studies need to focus on these pieces and their interconnected relationship to reach a thorough understanding of test preparation practices in a particular context. Therefore, these insights provide clear empirical foundations for this study to investigate the under-explored test preparation phenomenon in the
Chinese context from three interrelated aspects: the nature, the effects, and the value of the test preparation practices that test preparation centres promise will help Chinese students succeed.

Dissertation Overview

This dissertation is written in five chapters. In Chapter 1, I introduce the context of test preparation phenomenon in China, the research purpose, the research questions, and the rationale for conducting this study. Chapter 2 begins with a review of test preparation definitions and sets the scope of the term test preparation used in this study. The chapter continues with a review of existing empirical studies on test preparation to establish the current understanding of test preparation research and the research gaps that need to be addressed. The chapter then reviews the two theoretical foundations upon which the major previous test preparation studies were grounded on—validity and washback—to understand the researchers’ theoretical perspectives. Situating the current study within the washback framework, this chapter closes by illustration the use of Green’s (2007) washback model as the overarching theoretical framework to address the research purpose and research questions.

Chapter 3 introduces the overall research design and describes the research settings, participants, instruments, sources of data, procedures of data collection, and methods of data analysis and integration. Chapter 4 reports the results of the three research questions—the nature of test preparation practices, the perceived effects of these practices, and the perceived value of these practices from Chinese students’ perspective.

Drawing on Green’s washback model, Chapter 5 presents a discussion of how the particular test preparation practices are shaped by the various characteristics related to the participants and the context (washback variability), how the perceived effects on students’ test scores are interpreted in relation to the specific practices (washback direction), and how the
value is associated with the various characteristics, practices, and the perceived effects (washback variability and washback intensity). The chapter concludes with the implications of my research for both the language testing research community, and the language teaching community.
Chapter 2 Literature Review

This chapter reviews the literature on test preparation research to provide the empirical and theoretical foundations that guide this study. I begin by reviewing the key terms that have been used to define ‘test preparation’ across educational research fields (including curriculum studies, educational measurement, and language testing) to establish the scope of this term in this study. I then review the existing test preparation studies in these research fields from three main dimensions: the participants involved in test preparation practices, the process of test preparation practices that are taken and used by various participants, and the products (outcomes) of these practices. The empirical evidence produced by these studies contributes to understanding the aspects related to test preparation practices that have been researched, and also to identifying the gaps in this research that led to the current study.

Next, I review the two theoretical foundations upon which the majority of test preparation researchers have grounded their studies—broadly, ‘validity’ for educational researchers; and specifically, the ‘consequential aspect of validity’ or ‘washback’ for language testing researchers—and discuss the contributions and limitations of each. This discussion leads to the last section of this chapter, a discussion of Green’s (2007) washback model, which I used as the overarching theoretical framework for this study. I illustrate why and how the washback model guided this study to reach an understanding of the test preparation phenomenon in China.

Test Preparation: Definitions and Scope

In the following section, I first review the three key terms most commonly used to defined test preparation, focusing on the particular settings and research perspectives in which these terms are used. Understanding the relevant contexts and perspectives helps to clearly define both
the scope of the term test preparation and the related operational terms used to frame the research purpose and research questions of this study.

**Three Key Terms that Define Test Preparation**

The following three key terms have been commonly used to define the practices that prepare students to take tests: coaching, teaching to the test, and, more commonly test preparation.

**Coaching.** The term coaching is often applied to the commercial programs in which students participate outside of the school setting to prepare for taking tests. Coaching programs usually target the preparation for the standardized tests widely used for screening students for admission purposes, such as the Scholastic Aptitude Test (SAT) (Messick & Jungeblut, 1981), the Graduate Record Examinations (GRE) (e.g., Powers, 1985), the Test of English as a Foreign Language Internet-based test (TOEFL iBT) (e.g., Liu, 2014).

Although the use of this term tends to be limited to refer to short-term instruction targeted at improving students’ test performance on specific tests, without necessarily aiming for the improvement of students’ abilities more generally (e.g., Pike, 1978), coaching has been accepted by many researchers as representing a wide variety of test preparation practices (e.g., Bond, 1989; Cole, 1982; Messick, 1982). These range from short-term practice on sample test items to long-term instruction on skill development (Messick, 1982), and include test familiarization, practice drills, motivational enhancement, test-taking strategies (both general and test-specific), and subject content knowledge and skills development.

The claim by the test preparation centres that they can produce students’ score gains on high-stakes tests raises concern over test validity among researchers within both testing agencies and the educational measurement field. These researchers have extensively investigated if, and to
what extent, coaching might influence students’ test scores (e.g., Bangert-Drowns, Kulik, & Kulik, 1983; Koretz, 2005; Messick, 1982; Montgomery & Lilly, 2012; Powers, 1985; 2012). The collective evidence from these studies contributes to understanding the relationship between coaching and test validity (e.g., Haladyna & Downing, 2004; Liu, 2014; Messick & Jungeblut, 1981).

**Teaching to the test.** The expression ‘teaching to the test’ is often used within the K-12 school settings in general, and describes classroom teachers’ practices of focusing their instruction on test content to improve students’ scores on provincial or national standardized tests (e.g., Au, 2007). These tests are usually mandatory for students to take in order to measure if they have reached the required levels of subject knowledge specified in curriculum or syllabus, such as the literacy and numeracy tests for Grade 3, 5, 7, and 9 developed by the National Assessment Program in Australia (e.g., Wilson & Hornsby, 2014).

Depending on the focus of teachers’ instruction, teaching to the test is defined in two ways: item teaching or curriculum teaching (Popham, 2001). Item teaching refers to teachers focusing on actual test items to increase students’ test scores (such as by using retired test papers). In contrast, curriculum teaching refers to teachers building their instruction around curriculum objectives sampled by the tests to enhance students’ test performance. The criterion for defining teaching to the test is the alignment between teachers’ practices with curriculum or with a narrowed focus on tested content. Teaching to test, therefore, is predominantly discussed within the curriculum research field with a focus on curriculum alignment, in particular regarding if the teachers’ pedagogy has been narrowed to test-related content (e.g. Au, 2009; Popham, 2001) and is appropriate or ethical (e.g., Miyasaka, 2000; Moore, 1994; Popham, 1991).
**Test preparation.** The term test preparation is applied to broader settings and is defined as a variety of activities used in an attempt to improve students’ test scores (e.g., Crocker, 2005; Moore, 1994), including reviewing content sampled by tests and practicing test-taking skills. Test preparation has been addressed across research fields, such as curriculum studies (e.g., Smith, 1991; Turner, 2009), educational measurement (Crocker, 2005; Lai & Waltman, 2008; Powers, 2012), and language testing (e.g., Green, 2007; Hamp-Lyons, 1998; Tsagari, 2012).

Compared to the specific settings and perspectives in which the previous two terms are used, the term test preparation is more inclusive. It has been used to refer to practices that help students prepare for a variety of standardized tests, including provincial or national tests (e.g., Firestone, Monfils, & Schorr, 2004; Lai & Waltman, 2008), and tests used for admission purposes (e.g., Green, 2007; Hayes & Read, 2004; Tsagari, 2012). This term has also been used to refer to preparation practices both inside school settings (e.g., Deng & Carless, 2010) and outside school settings, such as commercial test preparation centres (e.g., Montgomery & Lilly, 2012). Compared to the negative connotations often associated with coaching and teaching to the test (e.g., item-teaching), the expression test preparation is also more inclusive and can be applied to any practice that helps to prepare students for taking tests from drills and training for managing specific test items to instruction for improving academic skills (e.g., Chau, 2008; Yu, 2012). Therefore, test preparation is more inclusive among the three terms.

**Discussion.** As illustrated above, while the ultimate goal of the practices included in these terms is the same—to improve students’ test scores—these terms differ with respect to the settings, the perspectives involved, and the extent of inclusiveness implied. Coaching and teaching to the test are distinguished from one another by three aspects: (1) the setting, e.g., outside versus inside schools; (2) the testing programs, such as standardized tests developed by
testing agencies and used for admission purposes, or provincial versus national standardized tests based on curriculum; and (3) the research fields and research focus in question, such as concern regarding validity in the educational measurement field versus curriculum alignment in the curriculum research field. Despite such differences, the practices defined by these two terms are similar, and include a wide range of test preparation practices.

These definitions have implications for understanding test preparation, in particular the effects of test preparation on test scores. When discussing the inconsistent findings of the empirical studies on SAT coaching programs, Messick (1982) argued that the yes-no question “does coaching work?” is too simple to provide sufficient evidence with which to interpret these findings. The narrow extent to which the term coaching is used—to represent drills or practice on test items—limits researchers investigating the variety of practices and experience in coaching programs, such as students’ characteristics (e.g., time invested by students) and program characteristics (e.g., number of hours, detailed procedures). In contrast, the inclusive perspective that the term test preparation provides the foundation for examining “what degree coaching is effective and by what means”, or, more specifically, “how much students’ time and efforts devoted to what kinds of coaching experiences yield what level of score improvements” (p. 71, Messick, 1982, emphasis in the original). Evidence of this kind can contribute to better understanding what test preparation looks like in specific contexts, and also how such test preparation influences test scores and the validity of test score interpretation.

**Scope of Test Preparation in the Current Study**

The review of key terms in the previous section provides my rationale for using the term test preparation, given the following research context for the current study. First, the study focuses on two testing programs: The TOEFL iBT and the CET4. The TOEFL iBT is developed
by the ETS and is used for admissions purposes; the special preparation programs for the TOEFL iBT are thus commonly called coaching. The CET4 is based on the national English syllabus and is used to measure Chinese undergraduates’ achievements in English learning; the preparation practices for such tests are usually referred to as teaching to the test. The term test preparation, which is not bound to any particular type of test, is thus used herein to refer to the preparation practices across both tests.

Second, there is limited understanding regarding which preparation practices Chinese students experience. For this reason, the term test preparation, which is more inclusive than either the term coaching or teaching to the test, was chosen; it allows me to explore any possible practices that are used to help Chinese students achieve success on these tests. The scope of the term test preparation in the current study therefore includes the wide variety of practices that Chinese students experience at commercial test preparation centres.

Given such scope, I propose the following operational definitions of the key terms that are used consistently in this study to help frame the research purpose and questions. Those used to state the research purpose are defined first, followed by the key terms used in the three research questions.

**Test preparation** is used as a general term for the extracurricular experience that Chinese students have in order to achieve successful test performance on high-stakes English language tests.

**Test preparation centres** are English language training institutes that offer courses and services to students who take high-stakes English language tests administered in China.

**Test preparation courses** refer to fee-paying courses with structured teacher-led instruction that Chinese students take at test preparation centres as an extra-curricular activity.
**Test preparation practices** refer to the activities and strategies that prepare Chinese students to take high-stakes English tests.

**Nature** focuses on the *participants* involved in test preparation practices, specifically, the *characteristics* and *perceptions* they bring into test preparation practices, the *context* in which these participants were situated, and the *specific practices* they employed.

**Perceived effects** refer to the relationship between Chinese students’ test preparation practices and their scores of the two high-stakes English tests.

**Value** refers to Chinese students’ perceptions of the value about the test preparation practices that claimed would help them achieve success on the two English tests, specifically the benefits they received from test preparation practices on their test taking and beyond test taking, if any.

**Empirical Evidence of Test Preparation Research**

Students’ test performance has been increasingly used to make high-stakes decisions in relation to accountability, policymaking, and gatekeeping in the educational systems of many countries (e.g., Cumming, 2009; Klinger, DeLuca, & Miller, 2008; Ungerleider, 2003; Wang, 2012). These high-stakes decisions greatly impact multiple participants involved in testing and test preparation, such as students working to obtain an academic diploma or admissions/retention into educational programs (e.g., Cheng, 2008; Cheng, Klinger, & Zheng, 2007); teachers wishing to attain professional promotions and positive student evaluations (e.g., Cumming, 2009, Shi, 2006); and school boards allocating funds (e.g., Madaus, 1988). Many researchers have demonstrated the tremendous influence these high-stakes tests have on teaching (e.g., Alderson & Hamp-Lyons, 1996; Nichols & Berliner, 2007; Popham, 2001). As Madaus (1988) stated, “teaching has been inordinately skewed toward test preparation” (p. 36).
Test preparation has thus been addressed using different terms from different research perspectives. Coaching is studied in the field of educational measurement to explore its influence on validity (e.g., Bangert-Drowns, Kulik, & Kulik, 1983; Koretz, 2005; Messick, 1982; Montgomery & Lilly, 2012; Powers, 1985; 2012); teaching to the test is used in curriculum studies to examine the extent that test preparation aligns with curriculum (e.g., Au, 2009; Moore, 1994; Popham, 1991, 2001); and in the field of language testing, test preparation has been explored using the concept of “washback”—that is, the influence of testing on teaching and learning (e.g., Alderson & Hamp-Lyons, 1996; Elder & O’ Loughlin, 2003; Green, 2007; Hamp-Lyons, 1998; Tsagari, 2012).

Studies across these fields cover three major aspects of test preparation: who engages it, (e.g., teachers, students, people of test preparation/coaching programs), what and how actions are taken towards it (e.g., to teach test-related contents, to practice specific skills to be measured in tests), and what the subsequent outcomes are of these actions (e.g., changes in test scores, learning outcomes). In turn, these aspects align with three key components of how washback works: participants, process, and products in language testing (e.g., Bailey, 1996; Hughes, 1993; Green, 2007).

In keeping with the above, I use the logical structure of participants, process, and products to review test preparation studies. This review is organized into three sections: (1) the participants who engage in test preparation, (2) the process by which they prepare for their tests, and (3) the products that they achieve through this process.

**Participants**

Following the definition used in washback research literature (Bailey, 1996; Hughes, 1993), participants of test preparation are those who are influenced by test preparation.
Participants may include those who are influenced to directly perform practices aimed at test preparation, such as teachers and students (e.g., Green, 2007; Mickan & Motteram, 2009; Moore, 1994; Lai & Waltman, 2008; Wall & Horák, 2006, 2008); and also those who are influenced to contribute to the context of test preparation, such as administrators and course designers/providers who may emphasize test preparation (e.g., Deng & Carless, 2010; Firestone, Monfils, & Schorr, 2004; Green, 2007). These participants may hold varying perceptions of test preparation, and bring a variety of characteristics into test preparation that ultimately determine which specific practices are used or whether such practices can lead to success on tests (e.g., Green, 2007).

In the next sections, I will outline the research findings of these participants’ perceptions and characteristics, and then discuss what further empirical studies on are needed.

**Participants’ perceptions.** The perceptions of test preparation have been studied from the perspectives of various groups of participants, including teachers, students, and administrators. The perceptions cover the following aspects: reasons of employing test preparation practices (e.g., Deng & Carless, 2010; Firestone, Monfils, & Schorr, 2004; Green, 2006a; Smith, 1991), perceived appropriateness of test preparation (Lai & Waltman, 2008; Moore, 1994), and expectations for test preparation (e.g., Green, 2006a; Rahimi & Nazhand, 2010; Wall & Horák, 2006). Many studies have examined teachers’ perceptions of why they engage in test preparation. The teachers explained that they were providing their students with the kind/type of preparatory instructions based on various considerations, including affective concerns (e.g., anxiety), students’ demands, and contextual influences (e.g. policy). Some teachers said the anxiety to improve students’ test performance and the shame related to students’ poor test performance drove them to teach test-related content (e.g., Hamp-Lyons, 1997; Smith, 1991).
Others thought their students demanded teaching activities to help with test preparation (e.g., Alderson & Hamp-Lyons, 1996; Green, 2006a). Additionally, the pressure from education administrators and school policies was found to contribute to teachers’ increased didactic instruction on test content. Teachers experienced pressure to increase students’ test scores (e.g., Firestone, Monfils, & Schorr, 2004; Swanson & Stevenson, 2002) and had to follow the mandatory policy of increasing dedicated time on test preparation (e.g., Deng & Carless, 2010). Despite the different accounts of why teachers participate in test preparation, however, all reasons identified in these studies were centred on the same key consideration: increasing student test scores.

Consideration of test scores may also influence teachers’ perceptions of whether their test preparation practices are appropriate (e.g., Crocker, 2006; Popham, 1991). For example, Moore (1994) compared the appropriateness judgment of commonly used test preparation practices between teachers and test specialists (e.g., test developers, educational measurement researchers) and found that teachers rated significantly higher on the appropriateness of some practices than test specialists, such as extensive drills and practice on items similar to those on tests. Such differences suggest that teachers and test specialists use different lenses with which to see test preparation, with teachers viewing it as a way to increase students test scores (e.g., Popham, 1991), while test specialists view it as a source that can influence test score validity (e.g., Haladyna & Downing, 2004). Lai and Waltman (2008), by surveying and interviewing teachers, further probed the factors contributing to teachers’ justifications for their test preparation practices. They found that, when teachers judged appropriateness, they tended to evaluate whether scores demonstrated students’ ability levels (score meaning) and whether their practices contributed to learning (educational value). The inconsistent findings regarding how teachers
judge and justify their test preparation practices indicate the necessity for further research on how teachers make instructional decisions in test preparation. Such evidence can subsequently contribute to understanding how teachers’ test preparation practices influence the validity of students’ test scores.

In contrast to teachers, students reported reasons for engaging in test preparation beyond improving their test performance. When asked why they participated in test preparation courses, Green (2006a) found that an equal percentage of students chose between achieving good scores and improving their English language skills, and also that a comparable percentage of students chose learning academic skills. Similarly, Yu (2012) found that the major reasons students reported for taking test preparation courses included test performance improvement, English proficiency development, affective management (e.g., increase confidence, reduce anxiety), and social requirements (e.g., to meet parents’ requirement of preparing for tests). These student perspectives were more diverse than the teachers’ predominant perspective of focusing on increasing test scores. This difference indicates that teachers and students have varying opinions on what drives their engagement in test preparation.

Subsequently, the perceptions that teachers and students hold as to why they participate in test preparation tend to influence what teachers expect to teach and what students expect to learn (e.g., Green, 2007). Some studies found that teachers thought it was the students who prioritized test-related content (e.g., Rahimi & Nazhand, 2010; Wall & Horák, 2006). In contrast, Alderson and Hamp-Lyons (1996) found that students in TOEFL test preparation courses expected to learn more than just the test-taking related content that was emphasized by the teachers. Similarly, Green (2006a) found that students reported the prioritized expectations of developing English writing skills over some test taking strategies, although the teachers’ activities were identified as
having a focus on test format (e.g., class activities similar to test tasks). The reasons that teachers and students gave for participating in test preparation can offer a possible explanation for the different expectations reported in these studies; however, there is still a lack of evidence as to what contribute to such differences. Further inquiries along this line are needed to probe the contributing factors that bring about differences in what teachers expect to teach and what students expect to learn during test preparation.

Although teachers and students have been studied as the participant groups involved in test preparation practices, administrators as participants have rarely been researched. Their work is closely related to teachers’ and students’ test preparation practices, and may significantly influence how teachers teach for test preparation. Evidence with respect to administrators was collected from teachers’ comments—that some teachers felt various levels of administrators (e.g., schools, school boards) pressured them to dedicate time to test preparation instruction (e.g., Deng & Carless, 2010; Firestone, Monfils, & Schorr, 2004; Popham, 1991). However, such perceived pressure was not supported by further evidence, such as how administrators respond to teachers’ perceived pressure and how administrators understand test preparation practices. Since the research findings of these studies have indicated that administrators can be a source that influences test preparation practices, further evidence directly from administrators’ perspectives, such as their perceptions and attitudes towards test preparation, are greatly needed to understand the role administrators play in test preparation.

**Participants’ characteristics.** The characteristics of participants who engage in test preparation, such as teaching and learning background (e.g., Alderson & Hamp-Lyons, 1996; Ferman, 2004; Gan, 2009) and sociolinguistic background (e.g., Gibson & Swan, 2008), have been shown to influence participants’ perceptions and test preparation practices. Teachers’
characteristics, including teaching background, teaching beliefs, and linguistic and cultural backgrounds, were found to be related to their test preparation practice (Lai & Waltman, 2008; Gibson & Swan, 2008). Lai and Waltman (2008) found that grade level teaching background (e.g., elementary vs. middle/high schools) was a discriminative factor in teachers’ use and ethical judgment of specific test preparation practices. For example, elementary teachers tended to rate some test preparation practices as being more ethical than did secondary teachers, such as structuring all classroom assessments like the standardized tests students were going to take (Lai & Waltman, 2008).

The finding of grade level as an influencing factor raises the question of whether and what underlying reasons related to grade level influence teachers’ test preparation practices, such as teaching beliefs (e.g., applying pedagogy suitable to students) or student characteristics (e.g., learning needs) specific to the grade levels. More recently, Chappell, Bodis, and Jackson (2015) found diverse differences in teachers’ beliefs regarding English language knowledge that they considered appropriate for test preparation classes. Gibson and Swan (2008) focused on the sociolinguistic background of non-native English speaking teachers in test preparation classes in Malaysia, and found that the varied linguistic (e.g., English and native languages) and cultural backgrounds (e.g., cultural experiences in English-speaking countries and home country) helped these teachers understand the use of the English tests and thus enabled them to teach their students relevant skills. However, some background characteristics that may influence teachers’ test preparation—such as teaching experience, professional qualifications, and educational background (e.g., He, 2010)—have seldom been studied.

Students’ characteristics were also examined in relation to why they performed test preparation practices. The students who had not had regular courses in the subjects to be tested
tended to turn to coaching programs for possible test score gains (Cole, 1982), while students with lower language proficiency levels had strong incentives to take test preparation courses (Ferman, 2004; Gan, 2009). Focusing on test-takers’ accounts of preparing for an English language test in the Canadian context, Doe and Fox (2011) found that students with similar levels of English proficiency had different motivations for taking the test, and subsequently their acquisition of strategies in test preparation classes were different. More recently, Yu (2012) found that the majority of the students (90%) in test preparation courses did not have TOEFL iBT test taking experience (i.e., were taking the TOEFL iBT for first time) and took the TOEFL iBT preparation courses quite close to their test dates to familiarize themselves with the test. Similarly, Liu (2014) found that 60% of the students surveyed were taking the TOEFL iBT for the first time and these students spent two to eight weeks on preparation before their test dates. Although the characteristics revealed in these studies—learning background (timing and length of taking courses), proficiency levels, and test-taking experience—contributed to a description of students in test preparation, the empirical evidence linking the characteristics students bring into test preparation to what and how they do in it remains limited.

Discussion. There are many studies examining the perceptions and characteristics of participants in test preparation, and their findings draw attention to some important aspects to be further explored. In particular: (1) the participant group of administrators, (2) systematic investigation of participants’ characteristics, and (3) interactions among multiple groups of participants.

While a number of studies have examined teachers and students in test preparation, little attention has been given to administrators (both in school setting and in commercial program setting). Some studies have demonstrated that the pressure from administrators and school policy
may influence teachers’ test preparation practices, making it clear that it is necessary to put
administrators on the agenda for further research (e.g., Deng & Carless, 2010; Firestone,
Monfils, & Schorr, 2004). Nevertheless, this group of participants involved in test preparation
practices remains understudied. One possible reason is that the research context is typically
classrooms where teachers and students are directly involved in test preparation. Teachers and
students have thus become the key studied groups of participants. One feasible method to rectify
this gap is to expand the boundary of the research context from the classroom level to the school
or institution level (e.g., test preparation centres), and also to include the voices of administrators
in the research design.

Although some studies revealed the characteristics of their participants through
demographic descriptions (e.g., linguistic background, teaching beliefs), participant
characteristics were not the direct focus of the research or research question. As a result, the
characteristics that participants bring into their test preparation practices have not been
systematically researched. Many researchers have indicated the necessity of linking participants’
characteristics to interpret the practices and outcomes of their test preparation (e.g., Messick,
1980; Gibson & Swan, 2008; Green, 2007). For example, identifying teachers’ characteristics—
such as teaching experience, teaching beliefs, and professional qualifications—can provide
relevant evidence with which to understand why teachers conduct specific preparation practices
(e.g., Gibson & Swan, 2008; Chappell, Bodis & Jackson, 2015).

Furthermore, Messick (1980) suggested that effects estimates of coaching/test preparation
programs on test scores should be carefully interpreted because of the combined effects of
coeaching and students’ characteristics. For example, students with low English proficiency have
a strong incentive to take test preparation courses (e.g., Gan, 2009), suggesting that language
proficiency needs to be taken into account when interpreting the effects of test preparation practices on test scores. The examination of participants’ characteristics, therefore, can provide empirical evidence first to understand who these participants are, and second, to link their characteristics to what/how they engage in test preparation and what outcomes they achieve as a result.

This review also highlights the inconsistent results that have emerged from studies of participants’ perceptions of test preparation, such as teachers’ justifications for their practices (e.g., pressure from administrators vs. educational value), and the reasons teachers and students reported for taking test preparation courses (e.g., increasing test scores versus. developing particular skills). These inconsistent results might be attributed to the research design of studies, many of which only included a single group of participants in test preparation studies (e.g., Firestone, Monfils, & Schorr, 2004; Rahimi & Nazhand, 2010). The results based on data collected from one particular perspective (e.g., either from only the teachers’ or the students’ perspective) thus produced inconsistent results when trying to explain test preparation practices that involved the interactions of multiple participants. Green (2006a) compared the perspectives of both teachers and students and found that the narrowed instruction on test-taking strategies was not driven by the students but by the teachers’ focus on test features. The inclusive approach that Green (2006a) took is needed to further examine how these participants interact (e.g., the perceptions they hold towards test preparation) and how the characteristics of these multiple groups of participants interact.

**Process**

The process of test preparation includes any actions that participants take to prepare for the test (Bailey, 1996; Hughes, 1993). As discussed in the previous section, given a great number of
studies have focused on teachers and students, the test preparation process has mostly been examined from the perspective of the teaching and learning practices of these two groups (e.g., Firestone, Monfils & Schorr, 2004; Green, 2007; Mickan & Motteram, 2009; Xie, 2008). In the following sections, I will review this research—that is, on the actions and practices teachers and students undertake—to understand how these two groups engage in the process of test preparation.

**Teachers’ practices.** The practices that teachers use in test preparation have attracted much research attention (e.g., Cole, 1982; Gibson & Swan, 2008; Moore, 1994; Popham, 1991; Smith, 1991; Lai & Waltman, 2008; Yu, 2012). These studies have identified three main focuses in the content of teaching practices: teaching subject content, emphasizing test content, and instructing test-taking strategies. Specifically, some teachers were found to direct their instruction towards the knowledge, skills, or content represented by the test, such as improving English skills (e.g., Yu, 2012; Gibson & Swan, 2008; Green, 2007) or reviewing academic knowledge and skills in the curriculum objectives (e.g., Firestone, Monfils, & Schorr, 2004; Lai & Waltman, 2008), that is, teaching subject content. Others emphasized test content, building their instruction around test features and demands, such as providing students with templates to use for writing tasks on English proficiency tests (e.g., Chau, 2008; He, 2010) and practicing test-like items (e.g., Alderson and Hamp-Lyons, 1996; Firestone, Monfils, & Schorr, 2004; Wall & Dorák, 2006, 2008). Finally, instructing test-taking strategies was a practice commonly used by many teachers to help students familiarize themselves with test formats and manage affective influence (e.g., maximizing motivation, optimizing test anxiety) (e.g., Cole, 1982; Green, 2007; Moore, 1994; Yu, 2012).

These research findings highlight the blurred lines among the three foci of teachers’
practices, and draw attention onto the factors driving teachers switch their focus. This attention may help move us beyond simply describing each focus to understanding why and how teachers shift the focus of their practices. For example, Firestone, Monfils, and Schorr (2004) examined teachers’ preparation practices and how they changed their teaching in response to statewide mathematics and science testing of Grade 3 students in New Jersey. Firestone et. al (2004) observed teachers using two types of practices: “didactic instruction,” with extensive drills on test items; and “inquiry-oriented instruction,” which included activities on curriculum subjects. They attributed short-term didactic instructional strategies to the pressure from school board administrators, and the inquiry-oriented approach to the test information and support available to the teachers (Firestone et. al, 2004).

These findings revealed the external factors (e.g., pressure, resources of test information) that made the surveyed teachers change their practices for test preparation (Deng & Carless, 2010; Firestone et. al, 2004; Lai & Waltman, 2008). However, there was no evidence identifying what internal factors—such as teachers’ pedagogical knowledge or sociolinguistic background—may have played a role in the teachers changing their practices. Since these internal factors have been shown to help shape such practices (Gibson & Swan, 2008), further studies are needed to examine how teachers shift the focus of their test preparation practices, especially in relation to both the internal and external factors.

**Students’ practices.** In the literature, predominantly two major types of student test preparation practices have been examined: self-test preparation practices (Mickan & Motteram, 2009; Powers, 1985; Xie, 2008, 2010) and use of test-taking strategies (Doe & Fox, 2011; Liu, 2014; Tian, 2000; Yu, 2012). The first major type—self-test preparation—is the most commonly reported student self-study practices both in and out of test preparation courses (e.g., Mickan &
Motteram, 2009; Liu, 2014; Xie, 2008). Some students practiced test items to focus on test familiarization, such as practicing writing using released TOEFL writing tasks (Liu, 2014) and using the writing models in test preparation textbooks (Mickan & Motteram, 2009). Some students reported that they assessed their strengths and weaknesses related to specific English skills by practicing test items (e.g., Xie, 2008) and used some targeted practices to improve specific skills (e.g., Liu, 2014). Practices focused on improving language proficiency were also noted in Mickan and Motteram’s (2009) study, and many test candidates engaged in the preparation activities that included various discourses, such as listening to the radio, reading newspapers, or talking with local people.

Although the aforementioned studies reported students’ self-study practices for test preparation, including what the different focuses of their practices were (e.g., test items vs. English language skills), there was no evidence to explain how the students determined their specific focus. As the effects of student self-study test preparation were examined in relation to the focus of their practices (e.g., Powers & Swinton, 1982; Powers, 1985), understanding how students determine their focus can highlight the factors that impacts their practices (e.g., circumstances where students are situated) (Mickan & Motteram, 2009) and further can contribute to the interpretation of the effects of students’ test preparation practices.

The other major type of student test preparation practice is the use of test-taking strategies learned from test preparation courses (e.g., Doe & Fox, 2011; Tian, 2000; Yu, 2012). These strategies include those specific to test tasks, such as using argument structure and memorizing sentence structure for speaking tasks (e.g., Yu, 2012); those generalized across tasks and tests, such as taking notes and identifying main ideas (e.g., Doe & Fox, 2011; Yu, 2012), and those
related to the test-taking process, such as time management and practicing using a microphone (e.g., Tian, 2000; Yu, 2012).

The findings of these studies also highlight that students’ strategy use is related to their characteristics, i.e., who they are and what they bring into test preparation. For example, in his study examining Chinese students’ test preparation for TOEFL speaking tasks, Yu (2012) found that the students without TOEFL test-taking experience mostly expected their teachers to offer test-taking strategies. Further, Tian (2000) examined how Taiwanese students with high and low scores on TOEFL reading tasks used the strategies. She found that high-score students personalized the use of various strategies to help them focus on comprehension, while low-score students followed the strategies mechanically and stressed the use of strategies rather than comprehension. In contrast, in their study on student strategy use on an English proficiency test in Canada, Doe and Fox (2011) found that students with similar English proficiency levels approached strategy use differently, and that their motivation for taking the test and test preparation influenced their strategy use. Specifically, the students who viewed the test as a learning opportunity used the strategies (e.g., strategies related self-evaluation) to improve their English skills, while the students who viewed the test as either a barrier or a test-taking experience used the strategies (e.g., selective attention, planning) mechanically just for passing or experiencing the test (Doe & Fox, 2011).

Understanding how students use test-taking strategies is challenging because it involves students’ internal decision-making processes (Doe & Fox, 2011). The different strategy use highlighted by these research findings direct attention to potential factors that may contribute to why students select and use test-taking strategies differently, such as English proficiency, test-taking experience, or motivation for taking tests and test preparation (Doe & Fox, 2011; Yu,
Attention placed on these factors can enable further studies to go beyond simply identifying the specific strategies students use to exploring the decision-making process involved in choosing which test-taking strategies are used. Since test-taking strategies are an important aspect of students’ test preparation practices, examining their decision-making process in relation to the aforementioned factors should be on the research agenda.

**Discussion.** The understanding of test preparation practices, including their various foci (e.g., test contents, academic skills, test-taking strategy) and forms (e.g., self-study preparation, structured preparation with teachers’ instruction), can only be fully reached if we take into account the factors that influence these practices. The factors highlighted in the literature are related to the participants’ characteristics, such as English language proficiency, test-taking experience, teaching experience, and motivation for test-taking (e.g., Doe & Fox, 2011; Tian, 2000; Yu, 2012; Chappell, Bodis, & Jackson 2015; Lai & Waltman, 2008). The discussion in the previous section (Participants) has identified the significance of examining these characteristics to further understand test preparation practices. The other factors that may influence test preparation practices were related to the context in which these practices occur, such as the larger sociolinguistic context (e.g., English language environment) (e.g., Alderson & Hamp-Lyons, 1996; Gibson & Swan, 2008) and the immediate local context (e.g., classrooms, schools, coaching schools, test preparation courses) (e.g., Firestone, Monfils, & Schorr, 2004; He, 2010).

The context of the language environment may influence which specific test preparation practices are chosen. For example, students in English-speaking environments have opportunities to use English beyond their test preparation courses (e.g., talking with local people, listening to the radio) (e.g., Doe & Fox, 2011; Mickan & Motteram, 2009), while such opportunities are often inaccessible to students in non-English speaking environments. Therefore, because their
research was done on students living in an English-speaking environment, Alderson and Hamp-Lyons (1996) concluded that their findings regarding test preparation practices for the TOEFL were not applicable to contexts where English is not the medium of communication outside the classroom. Further, test preparation practices in school contexts are greatly different from those in commercial-test preparation centre contexts (Ma & Cheng, in press). In schools, test preparation practices can be oriented to both subject content and test content, while practices in test preparation centres tend to centre only on test content (e.g., Cole, 1982; He, 2010; Rahimi & Nazhand, 2010). Although many researchers have stressed the need to consider the specific context when interpreting test preparation practices (e.g., Alderson & Hamp-Lyons, 1996; Mickan & Motteram, 2009; Gibson & Swan, 2008), it is necessary to admit that test preparation practices do not occur in single or individual context. Therefore, test preparation practices need to be understood in the multiple contexts (i.e., sociolinguistic context and local context), specifically, how multiple contexts altogether influence test preparation practices.

**Products**

The final component with respect to the empirical evidences from existing test preparation research is the products related to test preparation practices. The products of washback are defined as what is learned and how the quality of learning is (Bailey, 1999; Hughes, 1993). Following this definition, the products of test preparation may include the learning and the quality of learning through test preparation practices. Because educational researchers are most concerned with the influence of test preparation practices on test score validity (e.g., Messick, 1982; Powers, 2012), and because many test preparation practices have been oriented towards test score improvement, the research on test preparation products has been mainly on the effects of test preparation on test score improvement.
Test score improvement. In the field of educational measurement, the effects of coaching programs on test scores is attracting many researchers, especially with respect to major large-scale assessments developed by the Educational Testing Services (ETS), such as the Scholastic Aptitude Test (SAT) and the Graduate Record Examinations (GRE). These researchers have studied coaching effects through either empirical research (Powers, 1985; Sesnowitz, Bernhardt, & Knain, 1982) or via synthesized summary or meta-analysis of research findings (Bangert-Drowns, Kulik, & Kulik, 1983; Becker, 1990; Messick, 1982; Montgomery & Lilly, 2012; Powers, 1993, 2012). However, the findings across these studies came to different conclusions regarding coaching effects. Some found that coaching programs could help increase test scores significantly (e.g., Montgomery & Lilly, 2012), while others concluded that the effects on test score gains were not as significant as what was claimed by coaching companies (e.g., Powers, 1993).

The difference in magnitude of coaching effects has been attributed to many factors related to both the characteristics of the students attending coaching programs and the programs themselves (e.g., Bangert-Drowns, Kulik, & Kulik, 1983; Becker, 1990). For example, Sesnowitz, Bernhardt, and Knain (1982) examined the impact of coaching on SAT scores by studying high-school students with different levels of coaching experience. They used multiple regression analysis to examine the coaching effects and took into consideration the factors that might impact them, such as previous coaching experience (e.g., attending different coaching schools), academic performance at school (e.g., grades in English and Math courses), and background characteristics (e.g., ethnic groups). Sesnowitz et. al found that the coaching at one coaching school helped students with previously poor test performance increase their SAT scores “appreciably” above the scores predicted by their demographic characteristics (p. 439), while the
SAT scores of students at the other coaching school did not show any increase. They argued that such effects needed to be interpreted carefully because there was no data collected regarding the curriculum administered in the two coaching schools other than the duration of coaching programs (Sesnowitz et. al, 1982). Sesnowitz et. al recommended doing additional research to examine the curriculum at coaching schools to understand their effectiveness (or potentially lack thereof).

Similarly, based on her meta-analysis that revealed considerable variability of coaching effects, Becker (1990) also strongly urged for further studies to attend to “what constitute coaching” (p. 408), including student characteristics, detailed descriptions of instructional activities, materials, and time, to reach a clear understanding of how these various facets contribute to coaching.

In language testing field, some researchers used score gains to demonstrate test preparation effects (Brown, 1998; Elder & O’Loughlin, 2003), and others examined if the students who took test preparation courses performed significantly better on English language tests in comparison to the students who did not (e.g., Gan, 2009; Hayes & Read, 2004; Robb & Ercanbrack, 1999). Brown (1998) found that the test preparation for the writing tasks of the International English Language Testing System (IELTS), which has a combined emphasis on both test tasks and English writing skill development, helped students move up one IELTS bandscore (IELTS bandscores are 0-9) over a ten-week test preparation course. Similarly, Elder and O’Loughlin (2003) found that students with non-English speaking backgrounds had an overall half-bandscore gain after three-months of intensive instruction on test preparation. The score gains in both studies were at the lower end of the IELTS bandscores, indicating that test preparation instruction worked well for the students with lower English proficiency.
In contrast, studies that compared test scores of students with and without test preparation experience showed either a significant difference in test performance between groups (e.g., Hayes & Read, 2004; Robb & Ercanbrack, 1999) or no difference between groups (e.g., Gan, 2009). These conflicting findings reveal that the understanding of the effects of test preparation for language tests remains unclear.

**Discussion.** The different approach to test preparation studies and the different results of test preparation effects on tests scores indicate a rather muddy literature on how test preparation works (Becker, 1990). Although research in the field of educational measurement highlighted the importance of both program and student characteristics in relation to interpreting test preparation effects, the majority of studies examining language testing did not include these facets. The omission of characteristics is the result of either limitations in their study design (e.g., small numbers of participants) or the use of statistical analysis methods (e.g., t-test) that were not able to consider multiple variables (with a few exceptions e.g., Green, 2007 and Liu, 2014).

Further, compared to studies examining coaching effects in the educational measurement field, research on test preparation effects in language testing were limited. One reason is that many language testing researchers take the conceptually challenging perspective of washback to study test preparation effects. This approach guides language testing researchers to primarily focus on how a test influences language teachers and leaners to do things they would not otherwise necessarily do. In other words, using washback leads to studying the process of teaching, and there is lack of understanding of “whether students have learned more or better because they have studied for a particular test” (Wall, 2000, p. 502).

The other reason for limited studies on test preparation in language testing is the methodological challenge of how to measure test preparation effects. Bailey (1996) has pointed
out that the experimental design with controlled variables would be unhelpful to measure washback effects because the natural setting in which washback occurs makes it difficult to guarantee a high degree of control over variables. This challenge also applies to the setting where test preparation practices happen.

However, the desire for higher test scores is one of the main reasons that students take preparation courses (e.g., Hamp-Lyons, 1998; Wall and Horák, 2006), and many language testing researchers have emphasized test preparation effects as a necessary focus for language testing research (Green, 2007a; Wall, 2000). Some researchers have used the statistical analysis that is effective to examine relationships among variables (e.g., correlation, multiple regression, structural equation modeling) to examine how test preparation practices are related to test performance (e.g., Green, 2007; Liu, 2014; Xie, 2010). These methods allow researchers to take into account many variables in test preparation processes (e.g., the characteristics of teachers and students, characteristics of test preparation courses), link these variables to students’ language test scores, and contribute to a clearer understanding of the effects of test preparation on test scores (Green, 2007).

Section Summary

Many educational researchers have made empirical contributions to understanding test preparation in relation to the participants involved, the process, and the products of these practices. The empirical evidence demonstrates that test preparation practices are influenced by an array of factors related to the inner decision making process of multiple participants coupled with the effect of external contexts. The test preparation practices influenced by such factors thus produce various effects on students’ test scores. These findings highlight the need for further research on participants to better understand who they are (e.g., their characteristics and their
perceptions). This evidence can subsequently contribute to conceptualizing how these participants perform test preparation practices (e.g., focus of test preparation practices, contextual influence on practices), and to what extent test scores are improved and learning outcomes achieved.

I review two major theoretical frameworks that provide the foundations for the existing test preparation research and also discuss the rationale of using Green’s washback model (2007) for research purpose of the current study.

**Theoretical Frameworks for Test Preparation Research**

In this section, I start by reviewing the two major theoretical frameworks upon which previous empirical test preparation research has been grounded: broadly, validity in educational research field; and specifically, the consequential aspect of validity or washback in language testing field. When reviewing each of these theoretical foundations, I also discuss the scope of the validity framework used to investigate test preparation, and the perspective that washback framework offers for understanding test preparation. The discussion then leads into the section on the theoretical framework for the current study. I introduce Green’s (2007) washback model and further illustrate the rationale for using Green’s model, specifically, how this model helps to frame my three research questions.

**Theoretical Frameworks of Existing Test Preparation Research**

Test preparation research in the educational studies field is mapped onto validity frameworks in order to understand how test preparation practices influence interpretations and uses of test scores (Haladyna & Downing, 2004; Messick, 1982; Powers, 2012). The phenomenon of test preparation has received increasing attention amongst language testing researchers who are interested in washback (Alderson & Hamp-Lyons, 1996; Cheng, 2014;
Green, 2007; Tsagari, 2012), because test preparation activities provide cases with which to examine washback—the influence of a test on what teachers teach and students learn—and how it works. Evidence from test preparation research in the field of language testing is thus examined from the washback perspective, or, as Messick (1996) argued, the consequential aspect of validity.

Validity. Test preparation has been studied within the validity framework because of its potential threat to the validity of score interpretation. The threat that test preparation poses to validity has been categorized into two main aspects: score inflation (e.g., Koretz, 2005) and construct-irrelevant variance (e.g., Haladyna & Downing, 2004; Messick, 1989). Test preparation may lead to score inflation when the increase in test scores does not reflect the equivalent improvement in students’ academic abilities, or the gain in test scores substantially overstates the improvement in students’ learning. Alternatively, if some students received test preparation and others do not, the differences in students test scores could be mingled up with both the unequal access to test preparation and the variance in the construct that the test intends to measure.

Whichever term is used, the focus is the contribution of test preparation to test scores. Framing it in validity language, the focus is the variance in test scores attributed to the effects of test preparation. With Messick’s (1982) insightful conceptualizations of coaching effects, the focus of coaching/test preparation research has been shifted from the yes-no question of “Does coaching work?” to the categorical question, “To what degree coaching is effective?” Messick further distinguished three types of coaching effects and illustrated these effects centreing on how to interpret test score gains in relation to improvement in students’ abilities.
There has been an increasing interest among the educational research community and testing agencies (educational measurement in particular) to address the concern of coaching/test preparation on validity. Researchers collect evidence to understand the variance in test scores that can be attributed to test preparation, or whether a test score was inflated or produced variance that was irrelevant to students’ abilities. This research aimed to contribute to the appropriate interpretations and plausible use of test scores.

Messick (1989) argued that test score interpretation and use should be demonstrated on both an evidential and a consequential basis. Mapping onto Messick’s validity framework, previous studies conducted to address validity concerns focused on two facets of the evidential basis: construct validity (e.g., the variance in test scores interpretation) (e.g., Powers, 1985; Messick & Jungeblut, 1981) and the relevance and utility of test scores (e.g., the use of students’ test scores for admission purpose) (e.g., Messick, 1982). The data collected from these studies provided the evidential basis with which to understand how test preparation was related to test score interpretation and test score use. However, with regard to understanding test preparation in relation to value implications and social consequences, these previous studies did not collect explicit evidence on these facets, although they were grounded on the theoretical framework of validity. The argument that Cole and Moss (1989) made regarding the scope of the validity concept may offer the rationale why these studies focused on evidential basis rather than consequential basis, or in concise words, many existing studies have examined test preparation practices from the “technical” perspective:

There are a host of other types of values held by those with an interest in testing situations…. Messick (1980, 1981) includes such values and evidence about the consequences of testing in his broadly defined validation concept…. We share these writers’ and others’ sense of the tremendous importance and legitimacy of considering values in the evaluation of a particular use of a test. However, we have limited the concept of validity to only a portion of all the information relevant to the
decision to use a test, namely, evidence about the appropriateness of the scores to accomplish the immediate purpose of test use. One’s values about those immediate purposes are relevant and important and relevant to the decision of test use, but they are outside our technical concept of validity. (p. 204)

Kane (2012) argued that the evolving concept of validity and validation addresses the ranges of consequences associated with test score interpretations and uses that expand over time. Therefore, including consequences (e.g., the consequences of score-based decisions) is important to validate the interpretations and uses of test scores. Perhaps due to the influence of the technical scope of the validity concept, as Cole and Moss (1989) argued, recent studies on coaching still focused on score interpretation and the relationship between coaching program practices and test scores (e.g., Liu, 2014), as measured by the size of the effect on test score improvement. Although the validity concept is evolving, recent test preparation research conducted through the lens of validity did not collect relevant evidence (the evidence of consequences in particular) to reflect the current conceptualization of validity.

When tracing the evolving role of consequences in validation, Kane (2012) commented that the language testing community has attended to “the myriad ways” that testing programs can have an impact, or “washback”, on teaching and learning (p. 59). Therefore, the following section on reviewing washback frameworks in the language testing field could provide insight for further test preparation research to collect evidence related to consequence.

**Washback.** Washback research has been a dominant area of investigation in language testing. It has great implication on test validation because washback research contributes to understanding the consequences, intended or unintended, that language tests bring to teaching and learning (Cheng, 2014; Cheng, Sun, & Ma, 2015). The understanding of the washback phenomenon is developing and evolving within the current educational climate because language tests are administrated to an increasing number of students with a variety of characteristics, and
are also increasingly used in a range of contexts and for a range of purposes (Cheng, Sun, & Ma, 2015). Language testing researchers make ongoing contributions to the development of the conceptualization of the phenomenon of washback. Building on the question “does washback exist?” posed by Alderson and Wall (1993), language testing researchers have attempted to conceptualize “what does washback look like?” (Cheng, Wanatabe, & Curtis, 2004, p. ix), which highlights that washback depends on the factors within the test and test setting (including the test, the testing contexts, and the stakeholders) (e.g., Cheng, 2014).

In the field of language testing, test preparation is studied within the washback framework because it is regarded as a case in which testing (high-stakes testing in particular) influences teaching and learning and subsequently brings consequences to test score interpretations and uses (Alderson & Hamp-Lyons, 1996; Messick, 1996). There has been a growing interest among language testing researchers in examining how teachers’ instruction and students’ learning are influenced when preparing for language tests (e.g., Alderson & Hamp-Lyons, 1996; Green, 2007; Tsagari, 2012; Wall & Horák, 2006, 2008). The findings of these studies are thus interpreted from the washback perspective to understand the phenomenon of test preparation, such as washback hypothesis (Alderson & Wall, 1993) and washback models (Bailey, 1996; Green, 2007).

Alderson and Wall (1993) proposed the following washback hypotheses to understand the influence of tests on teaching and learning. They hypothesized the influence on many aspects, including content (e.g. a test will influence what teachers teach and learners learn), methods (e.g., a test will influence how teachers teach and learners learn), beliefs (e.g., a test will influence attitudes to content, method of teaching, and learning), extents (e.g., a test will influence the degree and depth of teaching and learning), the importance of consequences (e.g., a
test with important consequences will have washback), and the scope (e.g., a test will have washback effects on *some* teachers and learners, but *not* for others).

Building on Alderson and Wall’s (1993) seminal paper, Alderson and Hamp-Lyons (1996) explored “what was going on in the TOEFL preparation classrooms” from the washback perspective. They found that the washback effects of the TOEFL presented different patterns with different teachers, and the multiple participants involved in the test preparation process (such as administrators and material writers) contributed to the washback effects they observed. Alderson and Hamp-Lyons further argued that washback effects are far more complex than simple forms of the washback hypotheses, and suggested the need for more complex conceptualizations. Language testing researchers have further proposed some models to conceptualize what washback looks like and how washback works (Cheng, Wanatabe, & Curtis, 2004). Bailey’s (1996) model and Green’s (2007) model are the two most influential conceptual models available to research washback (Rea-Dickins & Scott, 2007; Weir & Milanovic, 2007).

Bailey’s (1996) model develops Hughes’ (1993) trichotomy mechanism of participants, process, and products by which washback works, and illustrates a relationship between the trichotomy and tests (Rea-Dickins & Scott, 2007). Bailey’s model visualizes that a test influences products (e.g., learning, teaching, materials, and research findings) through both participants and process, and these may in turn provide feedback to the test. Based on this model, Bailey further distinguished between washback to learner and washback to the program, and identified different aspects of students’ learning and the various participants in washback research (Cheng, Sun, & Ma, 2015). Some researchers examining washback have reported and discussed their findings from the perspectives of participants, process, products and its
relationship, explicitly or implicitly using Bailey’s model as guiding theoretical framework (e.g., Ferman, 2004; Manjarrés, 2005; Muñoz & Álvarez, 2010).

The ongoing contributions that language testing researchers have made significantly add to the knowledge of the complexity of washback mechanism, including the understanding of an array of variables that influence how washback works (Cheng, 2008b, 2014; Spratt, 2005). These variables are not limited to the participants (e.g., background characteristics of teachers and students), and the process (e.g., characteristics of language courses, such as focus and length), but also to the test design features (e.g., language skills required in test tasks); all of these variables together further complicate the relationship between tests, learning outcomes, and test score gains. Therefore, such an evolving understanding of washback calls for the development of theoretical models of washback that “recognize a wide variety of moderating variables interacting with the influence of test” (Weir & Milanovic, 2007, p. xi). Green (2007) takes this one step further by developing a washback framework and proposes a model that allows for investigating and understanding washback from the three dimensions of direction, variability, and intensity through the interrelated multiple variables. In the following section, I review Green’s washback model and discuss the rationale for using it to guide the current study.

**Theoretical Framework for Current Study: Green’s Washback Model**

Green (2007) investigated the influence of the IELTS Academic Writing Module on test preparation for academic study and the equivalence between IELTS test preparation and other forms of language learning targeted at English for Academic Purposes. In this study, Green proposed a washback model and investigated whether this model is supported in relation to the role of the IELTS test in the context of preparation for academic study.
**Review of Green’s washback model.** Green’s washback model (Figure 1) conceptualizes washback from three dimensions: washback direction, washback variability, and washback intensity.

![Green's washback model](image)

*Figure 1. Green’s washback model (2007)*

Green argues that washback is most intense when participants value success on the test beyond developing language skills for target use, consider success is attainable but challenging, and work in a context where these perceptions can be shared among participants. These conditions are actually related to variability demonstrated among participants (e.g., participants’ characteristics and perceptions), and indicate that the dimension of intensity is influenced by variability in addition to test importance and test difficulty.

Green’s model allows for the collection of detailed evidence from the three dimensions of washback. Evidence from the participants’ focus, the focal construct (e.g., language skills
required for academic studies) or test characteristics (e.g., test design features) help to distinguish whether positive or negative washback occurs. Participants’ characteristics, including their knowledge, beliefs, values, and resources of test demands, help to understand how the variability of washback is related to a variety of characteristics. Studying how the participants perceive test importance and test difficulty allows for the interpretation of the degree of influence on the participants (e.g., students’ tests score gains). All of this evidence together contributes to making specific interpretations of the washback mechanism and to reach a clear understanding of washback phenomenon.

**Rationale of using Green’s washback model.** Green’s washback model is used as the guiding framework for the current study, and the rationale is based on the consideration of the research context, the research purpose, and the research questions of this study. With respect to the research context, there is insufficient empirical data on the phenomenon of Chinese students’ test preparation for high-stakes English tests and more evidence is greatly needed. Green’s model offers a guide for the current study to collect new data to explore multiple potential variables that influence Chinese students’ test preparation practices.

As discussed in the review of previous test preparation research, test preparation practices are influenced by many internal and external factors in relation to the participants who are involved and the context in which they are situated. This analysis highlights the nuanced layers necessary for researching and understanding test preparation. For the research purpose of the current study, Green’s model provides the guidance to collect such layered and interrelated evidence to shed new light on the phenomenon of Chinese students’ test preparation for high-stakes English tests.
Green’s model covers three aspects of test preparation practices to be studied in the research questions of the current study. First, the nature of test preparation practices in RQ1 can be explored through collecting data on the participants’ characteristics (washback variability). Second, the effects of test preparation practices in RQ2 can be interpreted by linking the evidence of the participants’ characteristics and the participants’ focus (washback direction). Finally, the value of test preparation practices in RQ3 can contribute to understanding the degree of washback (washback intensity) by connecting the data collected on the participants’ characteristics to the outcomes they achieved; as Green (2007) argues, the most intense washback occurs when learners value success beyond developing language skills for target use.

Chapter Summary

As Hamp-Lyons (1998) argued, “test preparation practices cannot be solved by research, though it can be much better understood” (p. 336). The review of the existing test preparation research in this chapter has identified the gaps that need addressing in further studies for a better understanding of test preparation phenomenon. These gaps are: examining how administrators and also the operation of test preparation centres influence test preparation practices, understanding how and why participants decide and/or shift the focus of test preparation practices, understanding the interactions of participants and their situated context, and examining what specific test preparation practices work to increase test scores and to what degree, and also understanding in what way students value test preparation practices. To address these gaps, the examination of test preparation practices needs to pay attention to the potential factors related to participants’ characteristics (e.g., educational background, training experience), perceptions (e.g., motivation of participating in test preparation, teaching beliefs) and situated contexts (e.g., the operation of test preparation centres, timeline for test preparation). The three research questions
of this study address these gaps by collecting detailed evidence in relation to the nature, the perceived effects and the value of test preparation practices; and Green’s (2007) washback model that conceptualizes washback from the dimensions of direction, variability, and intensity, provides the theoretical guidance to examine the collected evidences to understand the phenomenon of Chinese students preparing for high-stakes English language tests.

In the next chapter, I will present the methodology of conducting this study, including the overall research design, participant recruitment, research instruments, data collection, and data analysis for each of the three research questions.
Chapter 3 Methodology

In this chapter, I first provide the research design I developed to address the overall research purpose and research questions for my dissertation study. I then describe in detail how I conducted the research, including details about the research sites, the participants, the instruments, the procedures of data collection, and the methods of data analysis to address the three research questions.

Overall Research Design

The purpose of this study is to understand the phenomenon of Chinese students’ test preparation for two high-stakes English language tests—the TOEFL iBT and the CET4. To reach this purpose, three inter-related aspects of the Chinese test preparation phenomenon were examined: the nature of test preparation practices, the effects of these practices, and the value of these practices. I explain the overall research design in relation to the three research questions below.

Research Question 1 (RQ1) asked: what is the nature of test preparation practices to help Chinese students achieve success? This question was developed to explore the nature of test preparation practices from three aspects: (1) who the participants are that engage in these practices; (2) why these participants chose to engage in test preparation; and (3) what it is that these participants actually did to prepare for the test. The aspect of who explored participants’ characteristics; the aspect of why examined their perceptions and influencing contexts; and the aspect of what sought to determine the specific test preparation practices actually used.

To best capture the nature of test preparation practices using these three aspects, I chose a popular commercial test preparation centre in Beijing—Centre L—as a research site (Site 1) because there were the multiple groups of people who participated in the TOEFL iBT test
preparation courses at this centre (including the students, teachers, and administrators). To answer RQ1, I chose to observe what these participants were experiencing (through classroom observation and document analysis) and listened to how they justified their engagement in such experiences (interviews).

Research Question 2 (RQ2) asked: what were the perceived effects of test preparation practices on Chinese students’ test scores? This question explored the relationship between test preparation practices and students’ test scores. To answer this research question, I chose eight universities and colleges in five cities (Harbin, Taiyuan, Beijing, Changzhou, Guangzhou from north to south) in China as research sites, and these universities and colleges had the students who recently completed their test preparation courses before they took the same live CET4 test administration. To be consistent, I name this collection of research sites as Site 2 to indicate these were the sites for RQ2. From these universities and colleges, I selected students who recently completed their test preparation courses and were going to take the CET4 that same year. I gathered data on the students’ reported test preparation practices through a questionnaire and collected the test scores they achieved in the live CET4 test in order to understand their test preparation practices and the perceived relationship between test preparation practices and the students’ test performance and answer RQ2.

Research Question 3 (RQ3) asked: how do Chinese students perceive the value of these practices with respect to helping them achieve success? This research question explores the value that the Chinese students gave to their test preparation experiences. To answer RQ3, I chose a Canadian university with a stable enrolment of international students and that requires a successful TOEFL iBT test scores for admission as the third research site (Site 3). I then identified the Chinese students currently studying at this university who had taken test
preparation courses. Focusing on their previous test preparation experience and their current academic studies, I conducted semi-structured interviews to explore their perceptions about how the test preparation practices they engaged in related to the success they experienced both on the test and beyond the test.

Research Design

To address the above three research questions, I used the multiple-stage and multiple-method approach as the research design. The multiple-stage approach, which looked at the various stages of Chinese students’ test preparation experiences, was used to address the respective focus of each research question. The multiple-method approach was used to collect multiple sources of data from the multiple participants involved in Chinese students’ test preparation for the two high-stakes English language tests. The multiple methods with the multiple participants included classroom observations and interviews with students, the teachers, and the administrators; questionnaires with the students; and an examination of the documents (e.g., descriptions of test preparation courses, videos of public lectures Centre L offered) with regards to the test preparation courses and the test preparation centre’s administration.

Table 1 presents the overview of the multiple-stage, multiple-method design for this study and the samples used to answer each question. I describe in detail how I executed this research design in the following sections of research sites, participants, instruments, data collection and analyses.
Table 1

Overview of Multiple-Stage Multiple-method Research Design

| Purpose: To understand the phenomenon of Chinese students’ taking test preparation courses for two high-stakes English language tests—TOEFL iBT and CET4 |
|---|---|---|
| **Stage/Site** | ‘Currently’ taking test preparation courses | Recently completing test preparation courses | Achieving success in tests |
| Focus of RQs | Site 1: Centre L in Beijing, China | Site 2: Eight universities and colleges in five Chinese cities | Site 3: A Canadian university |
| RQ 1: Nature of test preparation practices | Qualitative evidence of the test preparation practices that helped the students achieve success in the TOEFL iBT |
| | • Documents |
| | • Classroom observations |
| | • Interviews |
| RQ 2: Effects of test preparation practices | Quantitative evidence of the relationships between the test preparation practices and the students’ CET4 test scores |
| | • Questionnaire |
| | • Test scores |
| RQ 3: Value of test preparation practices | Qualitative evidence of the students’ perceptions of the value the test preparation practices that helped them succeed in the TOEFL |
| | • Interviews |

Research Sites

In the following sections, I provide the detailed descriptions for each research site.
Site 1: A Commercial Test Preparation Centre in Beijing—Centre L

Site 1 was a commercial test preparation program in Beijing: Centre L. This site was purposefully chosen to address the focus of RQ1 (the nature of the test preparation practices). I chose this centre because its continuous enrollments of students allowed me to recruit participants who were experienced with the test preparation courses. This site allowed me to explore the participants’ characteristics and perceptions, as well as the particular context of Centre L, and also to observe the specific practices the participants involved in test preparation practices performed.

Centre L, founded in 2008, is unique as its main business is test preparation courses for global English languages tests, including the TOEFL iBT and the IELTS. I came to know about Centre L through a friend, Calvin (pseudonym), who obtained a master’s degree in engineering from a Canadian university. He joined this centre in 2011 as an educational consultant when Centre L first started the business of overseas educational consultation.

Centre L offers TOEFL iBT test preparation courses year round at two sites. One is the central site, located in an office building in the Central Business District in Beijing. Centre L rents two adjunct levels for administration/teacher offices and classrooms. TOEFL iBT test preparation courses for commuting local students (thus called the commuting TOEFL iBT course) are offered at this central site. Their other site is residential and is similar to a boarding school or study camp as students stay on-site when taking test preparation courses (it is thus usually called the residential TOEFL iBT course). The residential site is only available for summer and winter vacations.

By the summer of 2012, a total of 40 cohorts of students had taken TOEFL iBT test preparation courses at Centre L, and a total of 32 cohorts of the students had taken the IELTS
test preparation courses. Although no exact total was available, one cohort in the summer 2013 TOEFL iBT test preparation courses comprised roughly 200 students. The availability of the TOEFL iBT test preparation courses (e.g., offered year round) and the stable registration of students created opportunities to make multiple observations of what the participants (e.g., students, teachers, and administrators) were experiencing at this centre.

As a relatively newly established test preparation centre, Centre L was fairly small scale with regards to their personnel structure, which included: (1) a supervisory board, (2) a supporting team, (3) teachers, and (4) educational consultants. The major role of the supervisory board was strategic planning and quality control. The supporting team provided logistical supports to ensure the courses offered ran as scheduled (e.g., registration, classroom arrangements). The teachers were responsible for the test preparation courses and other general English language training. The educational consultants provided consultation services to students who inquired about academic applications to overseas universities. The small scale of the personnel structure allowed me to approach all four levels of personnel, and to make close observations of how each level contributed to the students’ test preparation experience at this centre.

Site 2: Eight Universities and Colleges in China

Site 2 included eight universities and colleges in five Chinese cities, including Harbin, Taiyuan, Beijing, Changzhou, and Guangzhou, from north to south. These universities and colleges were chosen to address RQ2 (the perceived effects of the test preparation practices on test scores). The rationale for focusing on these universities/colleges to address RQ2 was twofold. First, the popularity of the CET4 test preparation phenomenon and the large number of the CET4 test takers, as discussed in the earlier chapters, allowed me to recruit a large number of
participants with which to perform effective statistical analyses. Examination of the statistical relationship between test preparation practices and students’ test scores, as suggested by previous test preparation research, was better supported by a large sample size. The number of participants available through Site 2 offered greater understanding of the relationship between these practices and students’ test performance.

Second, the fact that the CET4 is administrated on university/college campus provided additional opportunity to examine the perceived effects of test preparation practices on CET4 test scores. When undergraduate students registered for the CET4, they obtained an individual CET4 registration number at their own university or college. With the help of the CET4 test developer—the National College English testing Committee—I was able to anonymously collect students’ actual test scores using the registration numbers the participants entered on their questionnaires. Subsequently, I was able to link students’ reported test preparation practices to their test scores.

The criterion of selecting these universities and colleges was CET4 test administration. Ten universities and colleges in seven Chinese cities met this criterion and were approached through my contacts—my supervisor, and colleagues in my PhD program—who had experience teaching and/or researching the CET4. Of the original ten, eight universities/colleges in five cities were chosen, based on student willingness of providing their CET4 registration numbers.

Site 3: A Canadian University

Site 3 was a Canadian university with a steady enrollment of international students (including Chinese students) on campus. This site was specifically chosen to address RQ3 (the value of the test preparation practices). English is the medium of instruction at this university. Therefore, international students whose first language is not English, or who have not recently
studied at a post-secondary institution where English is the medium of instruction, are required to obtain satisfactory results on an English language test before being admitted. The TOEFL iBT is one English language test that is accepted by this university, with minimum scores (both total and sub-scores) set for international students to meet.

International students attending this university have met or surpassed the required TOEFL iBT scores. Among these students it is believed that many are Chinese, although there are no official statistics. And it is probable that some of these Chinese students may have taken test preparation courses prior to their taking the TOEFL iBT; such possibility would allow them to retrospectively reflect on their test preparation experience to judge the usefulness of the test preparation practices, in particular in relation to their successful test performance. It is also possible that some of these Chinese students have encountered situations in their current academic studies where they have evaluated their English proficiency. Such situations would allow them to judge the worth and/or importance of their test preparation practices. All these understandings of the usefulness, worth, and importance were collected, and contributed to interpreting how these Chinese students perceived the value of test preparation practices.

**Participants**

The participants in this study included three groups who involved in the test preparation practices for the TOEFL iBT and the CET4: students, teachers, and administrators. The student participants were from three research sites: six students at Centre L in Beijing, 534 students at eight universities/colleges in China, and 12 students at a Canadian university. Teacher participants included eight teachers at Centre L. Administrator participants included two supervisors and one educational consultant at Centre L. In the following section, I describe these participants by each group.
Students

The first group of six student participants was recruited from Centre L. Two parallel TOEFL iBT courses—Course A and Course B—were offered at Centre L when I started this study in the summer of 2013. Both courses were intensive and were completed in two weeks; however, they were different in class size. Course A was a medium-sized class and had 42 students, while Course B was a small-sized class and had 10 students. I approached the students in each course when I did classroom observations and invited the students to participate in this study. The educational consultant at Centre L, Calvin, also helped me recruit participants among the students who came to him for consulting. In the end, two students in Course A and four in Course B agreed to be the participants.

Each student participant was given a pseudonym and an ID with the combination of the letters of CL-S (Centre L-Student) and one digit (e.g., from 1 to 6). Each of these six students hoped to study abroad, although they had different plans with regards to programs (e.g., undergraduate, graduate) and timing (e.g. within a year, or the next year). Three students—Lindsay, Liam, and Lucas—were in a master’s program at the time they participated in the study; they planned to apply to PhD programs or a second master’s program in the United States or Canada. The other three students—Young, Rossie, and Nicole—were in undergraduate programs at the time of this study; they either planned to be exchange students in their upper years or to apply for graduate programs in the United States. Of the six, only Young, had taken the TOEFL iBT at the time of this study. Table 2 provides a summary of the basic background information of these student participants, including gender, current academic program, major, and whether taking the TOEFL iBT.
Table 2

Background Information of the Student Participants at Centre L

<table>
<thead>
<tr>
<th>ID</th>
<th>Pseudonym</th>
<th>Gender</th>
<th>Academic program</th>
<th>Major</th>
<th>Taken the TOEFL iBT</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL-S1</td>
<td>Lindsay</td>
<td>Female</td>
<td>Master’s</td>
<td>Commerce</td>
<td>Not yet</td>
</tr>
<tr>
<td>CL-S2</td>
<td>Liam</td>
<td>Male</td>
<td>Master’s</td>
<td>Mechanical manufacturing</td>
<td>Not yet</td>
</tr>
<tr>
<td>CL-S3</td>
<td>Young</td>
<td>Male</td>
<td>Bachelor</td>
<td>Accounting</td>
<td>✓</td>
</tr>
<tr>
<td>CL-S4</td>
<td>Rosie</td>
<td>Female</td>
<td>Bachelor</td>
<td>Event management</td>
<td>Not yet</td>
</tr>
<tr>
<td>CL-S5</td>
<td>Lucas</td>
<td>Male</td>
<td>Master’s</td>
<td>Mass media</td>
<td>Not yet</td>
</tr>
<tr>
<td>CL-S6</td>
<td>Nicole</td>
<td>Female</td>
<td>Bachelor</td>
<td>Finance</td>
<td>Not yet</td>
</tr>
</tbody>
</table>

The second group of student participants included 534 students from eight universities and colleges in five cities in China, and all of these students aimed to take the live CET4 in December 2012. Usually second-year undergraduates take the CET4 (Zheng & Cheng, 2008), however, some universities and colleges allow first-year students to take it as well. Therefore, among these students, all the first-year undergraduate and some of the second-year undergraduate students were first-time CET4 test takers, and some of the second-year students were repeaters of the CET4.

Between November and December 2012, 11 teachers that I established contact with at these universities/colleges helped me to invite the students who had taken the CET4 test preparation courses to participate in this study by answering the questionnaire on their test preparation course experience before taking the CET4. A total of 534 students returned a completed questionnaire, a return rate of 63.6% among the 840 distributed questionnaires. There were slightly more female students (n=274, 51.3%) than male students (n=260, 48.7%).

The third group of student participants included twelve Chinese students who did their postsecondary studies at a Canadian university. Table 3 provides a profile of these students.
Table 3

Profile of Twelve Students at a Canadian University

<table>
<thead>
<tr>
<th>ID</th>
<th>Pseudonym</th>
<th>Gender</th>
<th>Program</th>
<th>Major</th>
<th>Taken the TOEFL iBT more than once</th>
</tr>
</thead>
<tbody>
<tr>
<td>CU-01</td>
<td>Rachel</td>
<td>Female</td>
<td>Bachelor</td>
<td>Math &amp; Economics</td>
<td>✓</td>
</tr>
<tr>
<td>CU-02</td>
<td>Wilson</td>
<td>Male</td>
<td>PhD</td>
<td>Material engineer</td>
<td>✗</td>
</tr>
<tr>
<td>CU-03</td>
<td>Tony</td>
<td>Male</td>
<td>Master’s</td>
<td>Electrical engineer</td>
<td>✓</td>
</tr>
<tr>
<td>CU-04</td>
<td>Clark</td>
<td>Male</td>
<td>PhD</td>
<td>Electrical engineer</td>
<td>✓</td>
</tr>
<tr>
<td>CU-05</td>
<td>Mary</td>
<td>Female</td>
<td>PhD</td>
<td>Geography</td>
<td>✓</td>
</tr>
<tr>
<td>CU-06</td>
<td>Sean</td>
<td>Male</td>
<td>Bachelor</td>
<td>Business</td>
<td>✓</td>
</tr>
<tr>
<td>CU-07</td>
<td>Justin</td>
<td>Male</td>
<td>Master’s</td>
<td>Business</td>
<td>✓</td>
</tr>
<tr>
<td>CU-08</td>
<td>Jenny</td>
<td>Female</td>
<td>Bachelor</td>
<td>Nursing</td>
<td>✓</td>
</tr>
<tr>
<td>CU-09</td>
<td>Jack</td>
<td>Male</td>
<td>Master’s</td>
<td>Physics</td>
<td>✓</td>
</tr>
<tr>
<td>CU-10</td>
<td>Chris</td>
<td>Male</td>
<td>Master’s</td>
<td>Electrical engineer</td>
<td>✓</td>
</tr>
<tr>
<td>CU-11</td>
<td>Maggie</td>
<td>Female</td>
<td>Master’s</td>
<td>Chemistry</td>
<td>Not available</td>
</tr>
<tr>
<td>CU-12</td>
<td>William</td>
<td>Male</td>
<td>Master’s</td>
<td>Education</td>
<td>✗</td>
</tr>
</tbody>
</table>

The criterion for recruiting participants at this research site was that the students had taken test preparation courses in China prior to taking the TOEFL iBT. The recruitment of these student participants was conducted in two steps. I first sent invitation emails through the list-serve to the Chinese Students and Scholars Association at this university. Three Chinese students replied to indicate their willingness to participate. I then used the snowball technique of asking these three Chinese students to nominate potential additional participants and subsequently found nine more student participants. Of these 12 students, four were female and eight were male. They were enrolled in either undergraduate or graduate programs, in different majors. Each student was given a pseudonym and an ID that was a combination of the letters CU (Canadian University) followed by two digits (e.g., 01, 02, 03, … 10, 11, 12). It was interesting to note that nine of these students had taken the TOEFL iBT more than once to achieve the required TOEFL score set by this Canadian university, or to achieve a better score in hopes of making their academic application more competitive.
Teachers

The teacher participants in this study were eight teachers of the two parallel TOEFL iBT courses—Course A and Course B—that were offered at the central site of Centre L. Four teachers were from Course A, and four were from Course B. These two courses had the same structure: four modules (listening, reading, writing, and speaking) and 32 sessions (eight sessions for each module). Each course had four teachers, and each teacher taught one module. Table 4 summarizes the structure of these two courses.

Table 4

Course Structure by the Teacher Participants at Centre L

<table>
<thead>
<tr>
<th>Course</th>
<th>Teacher (Pseudonym)</th>
<th>Module</th>
<th>Sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course A</td>
<td>Kerry</td>
<td>Listening</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Leo</td>
<td>Reading</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Henry</td>
<td>Writing</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Megan</td>
<td>Speaking</td>
<td>8</td>
</tr>
<tr>
<td>Course B</td>
<td>Ethan</td>
<td>Listening</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Yolanda</td>
<td>Reading</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Jimmy</td>
<td>Writing</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Will</td>
<td>Speaking</td>
<td>8</td>
</tr>
</tbody>
</table>

Of these teachers, six were male and two were female, and their ages ranged from 28 to 45. Only two teachers majored in English language or language education programs (e.g., literacy education), while the other six were fine arts (e.g., visual arts, film directing), business, science, and engineering majors. All of these teachers had to meet at least one of the key requirements to be selected as the candidates for teaching positions at Centre L: the experience of teaching test preparation courses, the demonstration of high English language proficiency (with TOEFL iBT or IELTS test scores as proof), and the overseas educational experiences (preferred as an asset). Five out of the eight teachers had overseas educational experience in either English-speaking countries (e.g., the United States) or countries with English as a medium of instruction (e.g., Malaysia, Sweden). Each teacher was given a pseudonym and an ID that combined the letters of
CL-T (Centre L-Teacher) with one digit (from 1 to 8). Table 5 provides a profile of the eight teacher participants and summarizes their educational background.

Table 5  
Profile of Eight Teacher Participants at Centre L

<table>
<thead>
<tr>
<th>ID</th>
<th>Pseudonym</th>
<th>Major</th>
<th>Degree</th>
<th>Country of degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL-T1</td>
<td>Kerry</td>
<td>Film production</td>
<td>Master’s</td>
<td>U.S.</td>
</tr>
<tr>
<td>CL-T2</td>
<td>Leo</td>
<td>Engineer</td>
<td>Bachelor</td>
<td>China</td>
</tr>
<tr>
<td>CL-T3</td>
<td>Henry</td>
<td>English language and literature</td>
<td>Diploma</td>
<td>China</td>
</tr>
<tr>
<td>CL-T4</td>
<td>Megan</td>
<td>Visual arts</td>
<td>Bachelor</td>
<td>U.S.</td>
</tr>
<tr>
<td>CL-T5</td>
<td>Ethan</td>
<td>Business</td>
<td>Master’s</td>
<td>Sweden</td>
</tr>
<tr>
<td>CL-T6</td>
<td>Yolanda</td>
<td>Physics</td>
<td>Bachelor</td>
<td>China</td>
</tr>
<tr>
<td>CL-T7</td>
<td>Jimmy</td>
<td>Literacy education</td>
<td>Master’s</td>
<td>U.S.</td>
</tr>
<tr>
<td>CL-T8</td>
<td>Will</td>
<td>Business management</td>
<td>Master’s</td>
<td>Malaysia</td>
</tr>
</tbody>
</table>

Administrators

The administrator participants included two supervisors from the supervisory board and one educational consultant from the educational consulting team. These staff were involved in administrative tasks at Centre L, such as creating proposals for strategic development or in-house professional training. This stakeholder group was included to offer their perspectives on the contextual conditions that supported the operation of the TOEFL iBT preparation courses at this centre. Each administrator was given a pseudonym and an ID that combined the letters CL-A (Centre L Administrators) and one digit (from 1 to 3). As summarized in Table 6, they all had the experience of studying at universities in Canada or the United States, although they each studied different majors.

Table 6  
Profile of the Administrator Participants at Centre L

<table>
<thead>
<tr>
<th>ID</th>
<th>Pseudonym</th>
<th>Major</th>
<th>Degree/Country</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL-A1</td>
<td>Calvin</td>
<td>Engineering</td>
<td>Master’s/Canada</td>
<td>Educational consultant</td>
</tr>
<tr>
<td>CL-A2</td>
<td>Hayden</td>
<td>Management</td>
<td>Bachelor/U.S.</td>
<td>Supervisor in chief</td>
</tr>
<tr>
<td>CL-A3</td>
<td>James</td>
<td>Education</td>
<td>Master’s/U.S.</td>
<td>Supervisor of test preparation courses</td>
</tr>
</tbody>
</table>
With a Master’s degree in engineering, Calvin was responsible for consulting services with regards to engineering training at universities in the United States and Canada, in particular graduate programs. The additional goal of Calvin’s work was to reach out to potential students and interest them in coming to Centre L. Hayden, as a member on the supervisory board, was assigned as the supervisor-in-chief to coordinate each team and ensure the routine operation of Centre L. James was the supervisor of all test preparation courses at the centre, and his primary duties included recruiting teachers, in-house teacher trainings, and evaluations.

Section Summary

In summary, there were three groups involved in test preparation courses that participated in this study: students, teachers, and administrators. At Centre L, six students, eight teachers, and three administrators who were currently involved in the TOEFL iBT test preparation courses were invited to participate. Each provided their particular perspectives to help understand the nature of test preparation practices. At eight universities and colleges in China, 534 undergraduate students who recently completed CET4 test preparation courses were invited to participate; their responses to a questionnaire on their test preparation experience and their CET4 test scores contributed to exploring the perceived effects of test preparation practices. At a Canadian university, 12 Chinese students who succeeded on the TOEFL iBT and started their academic studies abroad were invited to reflect on their experience of taking TOEFL iBT test preparation courses, and provided their perspectives on the value of test preparation practices.

Instruments

In this study, a total of eight instruments were used for data collection at three research sites. In the following, I describe each of these instruments by each research site.
Instruments for Site 1

Five data collection instruments were used for the participants at Site 1, including a document review form, a classroom observation schedule, and three interview guides (one for each participant group i.e., students, teachers, and administrators).

**Document review form.** A document review form (Appendix C) was designed to collect contextual information about the TOEFL iBT test preparation courses and about the operation of Centre L. Such information was used to not only generate interview questions that were related to specific test preparation practices but also to provide a contextual basis with which to interpret data collected from classroom observations and interviews. The document review form included five sections—date, type, purpose(s), content(s), comment(s)—and the information collected from the documents were organized into these sections. Date recorded the time when the document was published or distributed (e.g., the course schedules posted on the Centre L website). Type recorded the way by which the documents were delivered or distributed, such as brochures, public lectures, or posts on the centre’s website. Purpose(s) collected information regarding the purpose these documents intended to effect (e.g., outreach, advertising, or recruiting). Content(s) recorded the brief description of the major content the documents covered. And, finally, comments collected the comments made on the documents, if any, such as the comments on a course description.

**Classroom observation schedule.** The classroom observation schedule (Appendix D) in Green’s (2007) study was employed to collect classroom observation data in order to understand the nature of test preparation practices. This schedule was based on an observation instrument that was widely used in washback studies—the Communicative Orientation to Language Teaching (COLT) observation schedule (Spada & Fröhlich, 1995). Green (2007) adopted the
COLT to include the features of direct test preparation, such as learning test-taking strategies, and used the adopted schedule to examine IELTS test preparation courses. Green’s observation schedule included seven categories: activities, participant organizations, content, test references, content control, student modality, and materials.

Descriptions of these categories are as follows. An activity was considered to be any single event happening during classroom instructions, such as a drill, a translation task, a discussion, or a game. The term participant organization was used to describe how teachers and students participated in activities, e.g., teacher to student, student to student, or pair/group work. Content referred to what the teachers and students were talking, reading, and writing about, or what they were listening to in activities, including: management (e.g., discipline), explicit focus on language (e.g., grammar, vocabulary, coherence, appropriateness), and specific topics (e.g., academic knowledge, public issues, personal reflections). Test reference addressed how teachers did the direct instructions on test preparation, and Green further added three sub-categories: test (e.g., the target test), other (other assessment methods), and test strategies (e.g., test-taking strategies). The category of content control described who selected the content for classroom activities—the teacher, the student, or both together. Student modality described the particular skill or combination of skills involved in a classroom activity. Finally, material referred to the materials used in connection with classroom activities. These categories helped to collect and organize the data of classroom observations at Centre L.

**Interview Guides.** Three interview guides were used, one for each of the three groups of participants: teachers, students, and administrators at Centre L (Appendix E, F, G, respectively). Each of the three interview guides included two sections. The first section focused on the participants’ characteristics, collecting detailed information about their background related to test
preparation (such as educational background, training background, test-taking background, teaching background). This information provided important links to understanding the specific practices they performed. The second section focused on their experiences with regards to their engagement at Centre L, including studying, teaching, and working.

Two sources were used to help design the interview questions for each of three interview guides. First, the research findings of previous studies with regards to the contributing factors that influenced test preparation practices (e.g., teachers’ rationale for their pedagogy, students’ motivation for test-taking, and administrators’ perception of test importance) were used to design questions related to the participants’ characteristics. Second, the evidence related to a particular participant’s engagement in test preparation courses collected via document analysis and classroom observations, such as the course description of an English Grammar course that Henry, the teacher of the writing module, offered at Centre L in addition to the TOEFL iBT test preparation courses. This evidence was used to develop interview questions that elicited the participants’ responses with regards to their engagement in the test preparation courses and at the Centre.

**Instruments for Site 2**

Two instruments were used to collect data from the undergraduate students at eight universities and colleges in China: (1) a questionnaire on their test preparation experience for the CET4 (Appendix H), and (2) a live CET test that was administrated in December, 2012. In the following sections, I first describe the questionnaire, including its components and contributing sources. I then introduce the CET4, in particular the test scores and sections.

**Questionnaire.** The first instrument used to collect data from student participants at Site 2 was a questionnaire (Appendix H). This questionnaire included three sections with a total of 21
items and was administered only in Chinese. The questionnaire was designed in English but was administered in Chinese to ensure that there was no language difficulty in completing the questionnaire. The first two parts were translated into Chinese, and then back translated to ensure the accuracy of the translation. For the third part were taken from the Chinese version of Xie’s (2010) questionnaire. Section 1 focused on the students’ background characteristics, including gender, major, years of English language learning, test scores from university entrance English language tests, and previous CET4 test taking experience. Section 2 focused on the students’ perceptions of the importance and usefulness of the CET4. Finally, Section 3 focused on the students’ experiences taking CET4 test preparation courses.

Section 1 included six items. Three were open-response items and collected: CET4 registration number and/or student number at the local university/college (Item 1), major (Item 4), and test scores on the National Matriculation English Test (NMET)\(^3\) (Item 5). The other three items were closed responses with specific answers from which the student participants were asked to choose: gender (Item 2), years of English language learning (Item 3), and previous CET4 test taking (Item 6).

Section 2 had two items. These asked the students how they perceived the importance of CET4 test scores with respect to various participants (e.g., self, teachers, parents) (Item 7) and various test uses (e.g., degree/diploma, job) (Item 8). These items were on a 7-point Likert scale, where 1 represented not important at all, 6 represented extremely important, and 0 represented not applicable. All eight items in Sections 1 and 2 were taken from Cheng et. al’s (2014) study that examined Chinese students’ motivation for and anxiety about taking the CET.

---

\(^3\) The National Matriculation English Test (NMET) is the English language entrance test that secondary school graduates who wish to gain entrance to Chinese universities and colleges must take (see Cheng & Qi, 2009, for a review).
Section 3 included 13 items. Six items (Item 9-14) asked the students about the characteristics of the CET4 test preparation courses they took, including course type (Item 9), hours of instruction (Item 10), hours spent on practicing (Item 11), number of practiced CET4 test tasks (Item 12), timing of taking test preparation course (Item 13), and instructional contents of test preparation courses. Four items (Item 9, 10, 13, 14) were designed according to the information about the CET4 test preparation courses in the five cities where the eight universities and colleges were located. From my contacts at each university/college, I learned about the major local test preparation centres that offered CET test preparation courses; I then searched online for information about course type, hours of instruction, times when test preparation courses were offered, and the major instructional contents in the course descriptions. Two items (Item12 and 13) were taken from Xie’s (2010) study examining the relationship between Chinese students’ CET4 test preparation and their test performance.

The other seven items in Section 3 (Item 15-21) asked the students about their reasons of taking test preparation courses (Item 15), the specific test preparation practices for each of the five CET4 test sections (i.e., listening, reading, writing, translation, and cloze) they learned from test preparation courses (Items 16-20), and the perceived usefulness of the test preparation practices for each test section (Item 21). All eight items were on 7-point Likert scale, where 1 represented totally disagree, 6 represented totally agrees, and 0 represented not applicable. Items 14, 15, and 21 were taken from Yu’s (2012) study that explored Chinese students’ expectations, reasons, and perceived value of test preparation practices. Items 16-20 were taken from Xie’s study (2010) and each item included the various sub-items listing the specific test preparation practices for each of the five CET4 test section, as follows: Item 16 included eight sub-items for the listening section; Item 17 had nine sub-items for the reading section; Item 18 contained ten
sub-items for the writing section; Item 19 had four sub-items for the translation section; and Item 20 contained five sub-items for the cloze section.

**Live CET4 Test.** The second instrument used to collect data from undergraduate students with experience taking CET test preparation courses was a live CET4 test. The CET4 includes five sections—listening, reading, cloze, writing, and translation—and is 125 minutes in length. The CET test papers are scored and rated at the CET4 official rating centres, and each CET4 test-taker receives a certificate that lists five scores—a total score and four sub-scores on listening, reading, integrative, and writing (for a comprehensive review of CET, see Zheng & Cheng, 2008). Table 7 is a brief description of the CET.

Table 7

<table>
<thead>
<tr>
<th>Test Description of the CET4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sections</strong></td>
</tr>
<tr>
<td>Listening</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Reading</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Cloze</td>
</tr>
<tr>
<td>Translation</td>
</tr>
<tr>
<td>Writing</td>
</tr>
</tbody>
</table>
Instrument for Site 3

One interview guide (see Appendix I) was used to collect data from 12 Chinese students enrolled at a Canadian university. The interview guide elicited the students’ responses to a central focus: how they perceived the value of the test preparation practices they engaged in prior to achieving success in their test performance. The interview guide used for this site included 18 interview questions. The first two questions focused on the students’ demographic background—including age, gender, programs, major, and time of taking the TOEFL iBT—and their test-taking experience. The remaining 14 interview questions focused on collecting two aspects: (1) the students’ reflections on their experience taking test preparation courses; and (2) their perceived value regarding how/if this experience in general, and the test preparation practices in particular, helped them achieve success in the TOEFL iBT. The students were asked to reflect on their experience taking a particular test preparation course, including their motivation for taking the course, their choice of course type, course duration, hours of instruction, and the timing of taking the course. The students were then prompted to comment on the value of such experience, especially on the test preparation practices, and how or whether these helped them in relation to their successful test performance and their current academic studies. To elicit the students’ perceptions on value, words that implied the meaning of value were used as prompts in the interview questions. These words included “usefulness”, “worth”, and “importance”, and were taken from previous test preparation studies (Messick, 1980; Yu, 2012) and the conceptualization of the value concept (Lincoln & Guba, 1980; Scriven, 1998, 2007).

Section Summary

As summarized in Table 8, a total of eight instruments were used in this study to collect data from three research sites. At Centre L, one document review form, one classroom
observation schedule (Green, 2007), and three interview guides were used to explore the nature of the test preparation practices. At eight universities and colleges, a questionnaire on the students’ CET4 test preparation experience and a live CET4 test were used to examine the effects of the test preparation practices. At a Canadian university, an interview guide was used to elicit students’ perceptions of the value of the test preparation practices that helped them achieve a successful test performance. In the following section, I explained the process of using these instruments to collect data from each participant group.

Table 8
Summary of the Instruments by Three Research Sites

<table>
<thead>
<tr>
<th>Research Sites</th>
<th>Instruments</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centre L</td>
<td>1. A document review form</td>
<td>Excerpts taken from the documents</td>
</tr>
<tr>
<td></td>
<td>2. Green’s (2007) classroom observation schedule</td>
<td>• Audio records of classroom observation</td>
</tr>
<tr>
<td></td>
<td>Three interview guides on the participants’ engagements in the TOEFL iBT test preparation courses at Centre L:</td>
<td>• Field notes</td>
</tr>
<tr>
<td></td>
<td>3. Interview guide – Teachers</td>
<td>Interview transcripts</td>
</tr>
<tr>
<td></td>
<td>4. Interview guide – Students</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5. Interview guide – Administrators</td>
<td></td>
</tr>
<tr>
<td>Eight universities &amp; colleges</td>
<td>6. A questionnaire on the students’ CET4 test preparation experience</td>
<td>Questionnaire data</td>
</tr>
<tr>
<td></td>
<td>7. A live CET test administrated in December 2012</td>
<td>Test score data</td>
</tr>
<tr>
<td>A Canadian university</td>
<td>8. An interview guide on the students’ TOEFL iBT test preparation experience</td>
<td>Interview transcripts</td>
</tr>
</tbody>
</table>

Data Collection

The data from this study were collected between November 2012 and February 2014, using the eight instruments described in the previous section. In the following, I describe in detail how
the data were collected via documents, classroom observation, interviews, questionnaires, and test scores.

**Documents**

I collected data at Centre L from May to July 2013, after receiving oral permission from the supervisor in chief, Hayden, to conduct this study. The data were collected in two rounds from documents related to both the TOEFL iBT test preparation courses and Centre L itself. The first round was carried out before the TOEFL iBT test preparation courses started, and the collection focused on documents that described how the courses were administered at Centre L. In this round, I collected documents from two major sources: Centre L’s official website and the administration office where the students inquire about TOEFL iBT courses. Altogether there were eight documents collected in this round: two videos of public promotion lectures that Centre L prepared to recruit local university students; two pamphlets that highlighted the features of the TOEFL iBT test preparation courses the Centre offered; two TOEFL iBT course schedules for summer 2013; a webpage that introduced the teachers who taught the TOEFL iBT courses; and registration information for TOEFL iBT courses at the Centre.

The second round focused on collecting documents that were directly related to the specific test preparation practices observed in class. A total of 15 documents were collected. Seven documents in total were collected in this round: an in-house textbook for TOEFL iBT test preparation, the descriptions of three courses targeting general English skills, an Official Guide to the TOEFL iBT (OG), and two sets of class notes that two student participants took. Table 9 summarizes the collected documents from the two rounds.
<table>
<thead>
<tr>
<th>Round /Time</th>
<th>Documents</th>
<th>Number</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Round</td>
<td>Videos of public lectures</td>
<td>2</td>
<td>Official website</td>
</tr>
<tr>
<td>May-June,</td>
<td>Pamphlets of the TOEFL iBT courses introduction</td>
<td>2</td>
<td>Administration office</td>
</tr>
<tr>
<td>2013</td>
<td>Schedules of the TOEFL iBT courses</td>
<td>2</td>
<td>Official website</td>
</tr>
<tr>
<td></td>
<td>Brief biography of the TOEFL iBT teachers</td>
<td>1</td>
<td>Official website</td>
</tr>
<tr>
<td></td>
<td>Registration information for the TOEFL iBT courses</td>
<td>1</td>
<td>Administration office</td>
</tr>
<tr>
<td>2nd Round</td>
<td>In-house textbook for TOEFL iBT test preparation</td>
<td>1</td>
<td>Administration office</td>
</tr>
<tr>
<td>July, 2013</td>
<td>Descriptions of three courses targeting general English skills</td>
<td>3</td>
<td>Three teacher participants</td>
</tr>
<tr>
<td></td>
<td>The Official Guide to the TOEFL iBT (OG)</td>
<td>1</td>
<td>Administration office</td>
</tr>
<tr>
<td></td>
<td>Class notes</td>
<td>2</td>
<td>Two student participants</td>
</tr>
</tbody>
</table>

**Classroom Observation**

Classroom observations were conducted during two TOEFL iBT preparation courses in June 2013 (Course A and Course B). All the eight teachers from the two courses agreed to be observed on the condition that interruptions caused by the process of observation be minimal. To meet this criterion, the classroom observations were audio-recorded only. Field notes were also taken to supplement the audio-recordings for two purposes: to record what was not easily captured by the audio-recordings, such as in-class handouts and teachers’ PowerPoint slides, and to record my reflections on some observed practices. Field notes included both descriptive notes and reflective notes. The descriptive field notes (see Appendix J) included information about classroom activities, such as time, settings (e.g., change of topic or turns between teachers and students), materials (e.g., teachers’ PowerPoint slides), and students’ reactions/responses to teachers’ instruction/talks. The reflective field notes (see Appendix K) mainly recorded my thoughts and questions while observing a particular teacher’s practices; the reflective notes helped me generate interview questions specific to each teacher and each practice.
Six teachers allowed me to conduct classroom observations during two sessions of his/her module, and the other two teachers allowed me to observe one session. Four teachers (denoted with an asterisk in Table 10) agreed to my chosen observation time points, which included one at the beginning (Session #1) and the other in the middle or near end of the module (e.g., Session #4 or #7). The reason for purposefully choosing these time slots was to observe the sessions at different stages of the courses. The other four teachers chose the sessions in which they preferred to be observed. Table 10 summarizes the classroom observation data collection for each teacher participant, the module they each taught, the number and sequence of the session(s) being observed, and the hours of observations.

Table 10

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Module</th>
<th>Course (A/B)</th>
<th>Number of session(s)</th>
<th>Sequence of session(s) #</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kerry*</td>
<td>Listening</td>
<td>A</td>
<td>2</td>
<td>Sessions 1 &amp; 5</td>
<td>5</td>
</tr>
<tr>
<td>Leo</td>
<td>Reading</td>
<td>A</td>
<td>2</td>
<td>Sessions 4 &amp; 7</td>
<td>2.5</td>
</tr>
<tr>
<td>Henry*</td>
<td>Writing</td>
<td>A</td>
<td>2</td>
<td>Sessions 3 &amp; 8</td>
<td>5</td>
</tr>
<tr>
<td>Megan</td>
<td>Speaking</td>
<td>A</td>
<td>1</td>
<td>Session 4</td>
<td>2.5</td>
</tr>
<tr>
<td>Ethan</td>
<td>Listening</td>
<td>B</td>
<td>1</td>
<td>Session 3</td>
<td>2.5</td>
</tr>
<tr>
<td>Yolanda*</td>
<td>Reading</td>
<td>B</td>
<td>2</td>
<td>Sessions 2 &amp; 4</td>
<td>5</td>
</tr>
<tr>
<td>Jimmy*</td>
<td>Writing</td>
<td>B</td>
<td>2</td>
<td>Sessions 5 &amp; 7</td>
<td>5</td>
</tr>
<tr>
<td>Will</td>
<td>Speaking</td>
<td>B</td>
<td>2</td>
<td>Sessions 4 &amp; 5</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>14</td>
<td></td>
<td>37.5</td>
</tr>
</tbody>
</table>

Interviews

A total of 27 participants were interviewed, including six students at Centre L, twelve students at a Canadian university, six teachers at Centre L, and three administrators at Centre L. The interviews were about 30—60 minutes in length, and all were conducted in Chinese. Among these participants, 25 were interviewed one-on-one in person, and two students from Course A at Centre L chose to be interviewed together. All participants were informed and consented to the audio recording of the interviews.
I began by interviewing the supervisor in chief and the educational consultant of Centre L before the two TOEFL iBT courses started in May 2013. These interviews collected information about the operations of the Centre L to obtain the contextual information. Upon completing the classroom observations, I then interviewed six teachers, six students, and the supervisor of the test preparation courses between June and July 2013. I purposely chose to conduct these interviews after I completed the classroom observations because I used the reflective class field notes to generate the interview questions. This strategy enabled me to further prompt the teachers and the supervisor to elaborate on their engagement in the TOEFL iBT courses, and in particular their test preparation practices.

From October 2013 to February 2014, I conducted one-on-one interviews with all 12 Chinese students at a Canadian university to explore their perceived value of the test preparation practices they undertook prior to their successful test performance. At the end of each interview, every student was asked to provide any available materials s/he obtained from the test preparation course. Five students provided the electronic documents they kept on laptops, such as Ji-jing (机经) (the test-takers’ reflections of the test-taking experience and test items), and saved posts with regards to TOEFL iBT test preparation on teachers’ blogs or online forums. I used these materials for triangulation with the interview data to link the students’ perceptions of the value with the concrete examples they provided. After transcribing all interview records and reviewing the transcripts, a fifteen-minute follow-up interview was conducted with each student (eight were conducted face-to-face and four were over the telephone). This follow-up protocol served as the member-checking mechanism and provided the opportunity review any accounts in the transcripts that needed to be clarified or explained.
Questionnaires and Test Scores

Both the questionnaire and test score data were collected to examine the perceived effects of the test preparation practices on students’ test scores. I contacted 11 teachers at eight universities and colleges in early November 2012, and asked their help to recruit participants among their students. I then sent the electronic version of the questionnaire to four teachers in Taiyuan, Beijing, Changzhou, and Guangzhou, and sent 600 hard copies of the questionnaire to 7 teachers at Harbin, where I conducted the data collection in person.

During the week before the CET4 test date (December 22, 2012), all teachers distributed the hard copies of the questionnaire to students who agreed to participate in this study. The questionnaires were purposely distributed close to the test date to ensure that the students had their CET4 registration number when they filled them out. The CET4 test registration cards were issued around 10 days before the test date. Therefore, during the week before the CET4 test date, the students were able to include the CET4 registration numbers on the questionnaire; these numbers subsequently ensured that collected test score data could be linked to the questionnaire data for the purpose of data analysis. The eleven teachers distributed a total of 840 questionnaires, and collected 534 returned questionnaires with the CET4 registration numbers, a return rate of 63.6%.

Collected CET4 registration numbers were sent to the CET4 test developer—the National College English Testing Committee—in March 2013 to collect these students’ CET4 test scores. Two weeks later, I received the test scores in a password-protected Excel file from the CET4 publisher that contained each student’s CET4 test scores, including a total score and the four sub-scores of listening, reading, integrative, and writing.
Primary Data and Supplementary Data

As illustrated in the overall research design, the evidence from each research site was collected to ensure the specific focus of each research question was addressed. These collected data became the primary data, and addressed the focus of each research question as follows. At each site, I also collected the data beyond the primary focus of the specific research question for that site, and these data became the supplementary data that contributed to addressing the other two research questions. Table 11 summarizes the primary data and supplementary data.

Table 11
Summary of Primary and Supplementary Data Collected from Three Sites

<table>
<thead>
<tr>
<th>RQ 1: Nature of test preparation practices</th>
<th>Centre L</th>
<th>Eight universities and colleges</th>
<th>A Canadian university</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The practices that prepared the students for the TOEFL iBT</td>
<td></td>
<td>Supplementary</td>
<td></td>
</tr>
<tr>
<td>• Documents</td>
<td></td>
<td>The practices that prepared the students for the CET4</td>
<td></td>
</tr>
<tr>
<td>• Classroom observation</td>
<td></td>
<td>• Questionnaire</td>
<td></td>
</tr>
<tr>
<td>• Interviews</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RQ 2: Perceived effects of test preparation practices</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supplementary</td>
<td></td>
<td>Primary</td>
<td></td>
</tr>
<tr>
<td>The comments on influences of the preparation practices on the TOEFL iBT scores</td>
<td></td>
<td>The relationship between the test preparation practices and the students’ CET4 test scores</td>
<td></td>
</tr>
<tr>
<td>• Interviews</td>
<td></td>
<td>• Questionnaire</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Test scores</td>
<td></td>
</tr>
<tr>
<td>RQ 3: Value of test preparation practices</td>
<td></td>
<td>Supplementary</td>
<td></td>
</tr>
<tr>
<td>Supplementary</td>
<td></td>
<td>Primary</td>
<td></td>
</tr>
<tr>
<td>The participants’ comments on the benefits of the test preparation practices for the TOEFL iBT</td>
<td></td>
<td>The students’ rating of the usefulness of the CET4 test preparation practices</td>
<td></td>
</tr>
<tr>
<td>• Interviews</td>
<td></td>
<td>• Questionnaire</td>
<td></td>
</tr>
</tbody>
</table>

79
With respect to RQ1, the primary data were collected from the teachers, students, and administrators who were currently involved in the TOEFL test preparation courses at Centre L, and contributed to exploring the nature of test preparation practices. The test scores of the students at the eight universities and colleges who recently completed their test preparation courses were the primary data to examine in the relationship between the test preparation practices and these students’ test performance, the focus of RQ2. Finally, the perceptions of the students who started their studies at a Canadian university after successful test performance were the primary data to shed light on RQ3 by helping to understand the value of the test preparation practices.

The supplementary data were collected along with the primary data for the particular research question. For example, when a student participant at Centre L was asked about his/her motivation for taking the TOEFL iBT course at this centre (nature), s/he commented on how the previous experience of taking test preparation courses had influenced the test scores on his/her previous three TOEFL iBT attempts (effects). When asked about the specific test preparation practices observed in class (nature), some student and teacher participants at Centre L commented on how they perceived the benefits of such practices (value). The student participants at the eight universities and colleges responded to the questionnaire items on the specific test preparation practices they learned from CET4 test preparation courses (nature), and also rated the usefulness of these practices (value). Finally, the Chinese students at the Canadian university reflected on the specific test preparation practices (nature) when they were asked to elaborate their perceived values of these practices to help them achieve success in the TOEFL iBT.
Data Analysis

Two cycles of data analysis were conducted on all collected data. The first was conducted within each dataset, grouped by the individual instrument used. This cycle of analysis thus included the following: document analysis of the 15 documents at Centre L, analysis of the 14 classroom observations conducted at Centre L, analysis of the interviews conducted with 15 participants at Centre L and 12 participants at a Canadian university, and the analysis of the questionnaire and test score data of 534 undergraduate students at eight universities and colleges. The second cycle of data analysis was conducted across all datasets; specifically, I integrated the primary and supplementary data (summarized in Table 11), to address the particular focus of each research question.

First Cycle Analysis

The collected documents were reviewed to (1) generate interview questions on the specific test preparation practices, and (2) link with the data from classroom observations and interviews as a cross-reference to enhance the validity of the data interpretations.

The classroom observation data collected at Centre L included the audio recordings and field notes of the classroom observations. These data were organized into the seven categories from Green’s (2007) observation schedule (Appendix D)—activities, participant organizations, content, test references, content control, student modality and materials—and contributed to understanding the specific classroom components of the test preparation courses. The process coding method was then used to analyze these components and capture the actions that happened in the classroom settings. Specifically, the “strategic, routine, random, novel, automatic, and/or thoughtful” actions were coded (Saldaña, 2009, p. 247) such that the generated codes of these actions served to describe what happened in the classroom.
The interview data included the verbatim transcripts of the interviews with 27 participants: six students at Centre L, six teachers and three administrators at Centre L, and 12 students at a Canadian university. The analysis of each transcript at this cycle was guided by one central question: “What did the participant tell me about…?” The three groups of participants at Centre L were prompted to talk about their engagement in the TOEFL iBT test preparation courses. I therefore analyzed the interview transcripts from these participants by asking the question: “What did the participant tell me about their engagement in the TOEFL iBT test preparation courses.” I then manually coded the transcript excerpts to find the meaningful segments that captured the process of their engagement, including why they taught/took/supervised the TOEFL iBT test preparation course(s), what they did specifically to prepare for the TOEFL iBT, and what achievements they made through the engagement. I then grouped the segments addressing the same issue into one code, and finally labeling the code to represent the core meaning (Saldana, 2009). This process was repeated until saturation was reached.

When analyzing the interview transcripts of the students at a Canadian university who achieved successful TOEFL iBT test performances, I was guided by the question: “What did the participant tell me about their success in relation to the test preparation practices that helped them prepared for the TOEFL iBT?” I then manually coded the transcript excerpts to find the meaningful segments on the various aspects of the test preparation practices that contributed to these students’ successful test performance. Both descriptive codes and InVivo codes were used when I did manual coding. The descriptive codes were used on the long excerpts to summarize the primary topic, and the InVivo codes were used to keep the data rooted in the participant’s own language (Saldaña, 2009). Similarly, the segments addressing the same ideas were grouped.
into one code, and the code was then labeled to represent the core meaning that these segments expressed.

I entered the responses to all 534 questionnaires into one Excel file. I then created a master Excel file that combined the questionnaire data and the test score data obtained from the CET4 test developer. The master Excel sheet was then imported into SPSS for a series of statistical analyses, including descriptive statistics, Chi-square tests, factor analysis, correlation, and multiple regressions. The descriptive statistics (i.e., frequency, mean) were calculated for seven items in Sections 1 and 2, and five items in Section 3, and used to enhance the understanding of the characteristics of undergraduates who took the CET4 test preparation courses, and also the characteristics of these courses. Chi-square tests of independence were then performed to examine the relationship between students’ characteristics and course characteristics, such as whether the status of test repeaters was related to the timing of taking CET4 test preparation courses.

To understand the perceived effects of the test preparation practices, statistical analyses were performed to examine the relationship between the CET4 test preparation practices and the students’ CET4 test scores; factor analysis, correlations, and multiple regressions were performed. First, exploratory factor analysis was conducted to examine the underlying factor structure of the items of CET4 test preparation practices (Items 16–20). Principal component analysis and direct oblimin rotation were used because principal component analysis was commonly used for factor extraction in language assessment research (Ockey, 2014) and oblimin rotation allowed the possibility of items of test preparation practices to be correlated. The Kaiser-Meyer-Olkin value and the Bartlett’s Test of Sphericity were used to decide the factorability. Eigen values greater than one, the levelling off of the scree plots, and the interpretability of the
factors identified, were used as the criteria to determine the number of factors. The internal consistencies of identified factors for each factor analysis was calculated.

Second, a correlation analysis was conducted between the factors of the CET4 test preparation practices identified in the previous factor analyses and the CET4 test scores (both total and sub-scores), to explore if there were correlational relationships.

Third, five individual stepwise multiple regression analyses were conducted to examine the significant predictors of the CET4 test scores, if any, and what they were. In each multiple regression analysis, either the CET4 total score or one of the four sub-scores was the dependent variable, and the independent variables included the characteristics of the CET4 students (e.g., gender, years of English learning, status of test repeaters, self-reported NMET score), the characteristics of test preparation courses they took (e.g., types, timing), and the CET4 test preparation practices identified in the previous factor analyses.

**Second Cycle Analysis**

The first cycle analysis contributed to establishing the initial codes (e.g., codes describing various actions that were generated from classroom observation data) or the initial findings (e.g., results of the statistical analyses of the questionnaire and test score data) within each individual dataset. By integrating the primary and supplementary data that addressed each research question, the second cycle analysis was then conducted across the datasets to discover the relationships among the array of initial codes/findings in first cycle analysis (Strauss & Corbin, 1998). The relationships woven together by the initial codes/findings were able to “explain why something happened or what something means” (Rubin & Rubin, 1995, p. 57). With respect to this study, these codes/findings were used to explain why the nature of test preparation practices were shaped in specific ways, why the test preparation practices had particular effects on the test
scores, and what the value of test preparation practices means.

In the second cycle analysis, the initial codes/findings were “regrouped and relinked in order to consolidate meaning and explanation,” and to further explore the relationships by the consolidated meaning and explanation (Saldana, 2009, p. 8). Strauss and Corbin (1998) suggested using the paradigm of conditions, actions, and consequences to explore the relationships. Conditions are the situations and circumstances in which a phenomenon is embedded, and can group the data by the questions: “why, where, how come, and when?” Actions are what happens under the conditions, and are represented by the questions: “by whom and how?” Finally, consequences are the outcomes of the actions, and are represented by the question: “what happens as a result of those actions?” For example, both the qualitative data and the quantitative data in relation to represent the reason why the students engaged in test preparation were grouped and integrated (see Appendix L).

Therefore, when mapping out the relationships among the initial codes/findings of the integrated datasets (both primary and secondary) that address the particular research question, I followed the paradigm that Strauss and Corbin (1998) suggested and kept asking questions with regards to the conditions, the actions, and the consequences. For example, several initial codes that addressed the nature of the test preparation practices were related to English skill building, such as the “importance of English skills for academic studies” (descriptive code from the transcripts of the interviews with four teachers), the “four-step method of practicing listening skills” (InVivo code from the transcripts of the interviews with two students), and “demonstrating the four-step method of practicing listening skill” (process codes from the data of the classroom observations with the teacher of listening module). I then asked the following questions:
• Which participant(s) emphasized improving English skills? What were the characteristics of the participant(s)?
• What reason(s) made them focus on English skills?
• When or under what conditions did the participant emphasize improving English skills (e.g., have sufficient time to prepare)?
• How and what specific actions did the participant take towards improving their English skills?
• What were the outcomes of their attention to improving English skills?

These questions about the conditions, actions, and consequences of improving English skills allowed me to find some subtle and implicit linkages among the initial codes, such as the codes related to the participants’ characteristics that influenced the focus on English skills, and contributed to adding dimensions to the theme of English skill Improvement.

I followed this paradigm (i.e., conditions, actions, and consequences) throughout the codes from the first cycle of data analyses to reach the major themes to answer research questions. In particular, considering the scope of data and the analysis for the first research question (i.e., the nature of test preparation practices), I used a figure to visually present the summary of the findings (Figure 2 in Chapter 4). I included the initial versions of this figure to ensure the credibility of the finding summary (Appendix M).

Chapter Summary

In this chapter, I described the overall research design that included the multiple stages during the complete process of Chinese students’ test preparation experience and the multiple methods to collect data relevant to the test preparation practices, and provided the discussion of how such research design contributed to reaching the overall research purpose of this study. I
then introduced three research sites, including one test preparation centre in Beijing, Centre L (Site 1), eight universities and colleges in China (Site 2), and one Canadian university (Site 3). I continued with four sections that presented in detail how the participants were recruited, what the research instruments were used and how data were collected in each research site, and how data were analyzed. In the next chapter, I will present the results of the data analyses.
Chapter 4 Results

The purpose of this study was to understand the phenomenon of Chinese students’ test preparation practices for two high-stakes English language tests—the TOEFL iBT and the CET4. In this chapter, I report the results corresponding to the focus of each research question in three sections: the nature of test preparation practices, the perceived effects of these practices, and the value of these practices.

In the first section, I present the results from RQ1 or “What is the nature of the test preparation practices employed to help Chinese students achieve success in the TOEFL iBT and the CET4?” These results were based on the data addressing the nature of the test preparation practices collected from classroom observations, interviews, and questionnaires. Three themes were identified reflecting three major test preparation practices: test familiarization, English skill improvement, and stress management. These themes and the related categories covering the participants’ characteristics, their perceptions, and their situated context, contribute to a thorough description of the nature of test preparation practices.

In the second section, I report the results from RQ2 or “What are the perceived effects of these test preparation practices on Chinese students’ scores on the two English tests?” These results were based on the following two sources of data: the qualitative data with regards to the influences of the TOEFL iBT test preparation practices on the students’ TOEFL iBT test scores collected from interviews and classroom observations, and the quantitative data from the CET4 test preparation practices and the CET4 test scores collected from questionnaires and the performance on a live CET4. I organize the results by test (i.e., the TOEFL iBT and the CET4) to clearly illustrate how the reported test scores (the TOEFL iBT) or the observed test scores (the CET4) are related to the test specific preparation practices.
In the third section, I present the results from RQ3 or “How do Chinese students perceive the value of these practices with respect to helping them achieve success?” Three themes emerge, reflecting the students’ perceived value: trustworthiness, efficacy, and reorientation. “Professional teachers” and “structured and reliable instruction” were the students’ perceptions on trustworthiness. Their perceptions on efficacy were related to the efficient investment in time and effort, the necessary knowledge of the test, and English skills improvement. Reorientation included strategy reorientation, where the students gained access to learning communities; and motivation reorientation, where the students gained strength to persist. Combined, trustworthiness, efficacy, and reorientation contributed to students’ successful test performances, although the extent to which this was true for each aspect varied from student to student.

**Nature of the Test Preparation Practices**

The results of the nature of the test preparation practices were organized by the descriptive structure of *who, why, and what*: *Who* were engaged in preparing for the test (characteristics)? *Why* did they prepare for the test (perceptions and context)? *What* did they do to prepare for the test (practices)? The descriptive structure was appropriate for identifying links among the data (Yin, 2009); therefore, the descriptive structure of *who, why, and what* was used to understand how the various factors (characteristics, perceptions, context) were linked. This structure also revealed how these factors, together, contributed to shaping three major practices: test familiarization, English skill improvement, and stress management. I combined the descriptive structure, the themes, and the related categories in a figure at the end of this section to both summarize the results and to highlight the links among the themes (*practices*) and the categories (*contributing factors of characteristics, perceptions and context*).
Characteristics: Who were Engaged in Preparing for the Test?

A total of 552 students (6 at Site 1, 534 at Site 2, and 12 at Site 3), eight teachers (at Site 1), and the three administrators (two supervisors and one consultant at Site 1) who participated in preparing for the two high-stakes tests were examined as the participants of test preparation practices in this study. The results of this study were as follows. The findings from both interviews and questionnaire revealed that, the 552 student participants had different academic backgrounds (i.e., major, program), language learning experiences (i.e., years of learning English), and test-taking experiences (i.e., repeating the test or taking the test for the first time). The interview data from the teachers and the administrators revealed that, they either had extensive teaching or personal learning experiences that directly related to English language and overseas study.

Among the students taking the test preparation courses for the TOEFL iBT or the CET4, approximately two-thirds majored in engineering and business management programs. The large number of the students in these two majors reflected the overall situation of undergraduate student enrollment by discipline/major in China, which engineering and business management/administrations were the top two disciplines/majors with the largest numbers of enrolled students (Ministry of Education (MoE) of China, 2014). The remaining third was from a variety of majors, including sciences, humanities and social science, and medical education. The majority of the students were undergraduates, with the exception of the 12 students with TOEFL iBT test preparation experience who were doing their master’s degrees.

All the students preparing for the TOEFL iBT reported that their English learning experience began in elementary school, which was equivalent to approximately 10 years of learning English by the time they took the TOEFL iBT. For the students preparing for the CET4,
293 (54.9%) had studied English for more than 10 years (started English learning from Kindergarten), 98 (18.4%) for 7–9 years (from Grade 3), and 142 (26.6%), studied for 4–6 years (from Grade 7). There were more students who took the test for the first time (non-repeaters) than those who repeated the test (test repeaters), with 357 non-repeaters and 194 test repeaters taking preparation courses for the TOEFL iBT or the CET4.

All eight teachers had different backgrounds with respect to English language professional training. Two had received professional training: Will (CL-T8) was accredited the TESOL (Teachers of English to Speakers of Other Language) certificate in Malaysia, and Jimmy (CL-T7) obtained a Master’s in Literacy Education from Harvard University in the United States. The other six teachers became “professional” by teaching at different English training schools for many years, or by completing in-house training at Centre L. However, the different professional training backgrounds did not limit them from teaching in commercial English training schools in China. All the eight teachers had the extensive experience teaching students of various ages (e.g., high-school students, undergraduates, adult learners) before being employed by Centre L. The teachers expressed in the interviews that their teaching experience with different groups of students enabled them to understand the students’ challenges (e.g., problems in English grammar, limits in vocabulary, struggles in logics) and to help them cope with these challenges. The essentiality of such relevant teaching experience, according to these teachers, helped them understand what their students wanted from their own test preparation experiences. They commented that there was “nothing wrong” with the fact that students came to test preparation centres hoping to “learn English better” and/or to “achieve high scores.” As Leo (CL-T2)—who had approximately 10 years of experience working at English training schools—said, attending

---

4 CL-T8 is the ID for the teacher participant Will, where CL stands for “Centre L”, T8 for “Teacher No. 8.”
test preparation centres “is like any patient who goes to clinic or hospital, and hopes his/her illness can be treated or cured.”

Five teachers (out of eight) and all three administrators held academic degrees from universities in English-speaking countries (e.g., the United States, Canada) or countries with English as a medium of communication (e.g., Malaysia, Sweden). While overseas education was not listed as a required qualification when Centre L recruited teachers and administrators, such experience was considered an asset because it was assumed such teachers and administrators could better understand the essentials of what is required in an English-speaking academic environment. This experience, in turn, would help transfer “what the students need” into their practices to guide students in preparation for the test, and subsequent overseas studies as well.

The interview data from the administrators and the teachers showed that the hiring policy at Centre L required its teachers to demonstrate a high level of English language proficiency. The appropriate evidence of this proficiency was determined to be high achievement on the same standardized tests (e.g. the TOEFL iBT, the IELTS) they were expected to prepare students to take. All of the teachers presented their TOEFL iBT scores when applying to Centre L, and all of them achieved close to full sub-scores on the module they taught (for example, Yolanda, who taught the reading module, scored 28 out of 30 on the sub-score of the TOEFL iBT reading). Only two of the teachers (Henry and Jimmy) had studied English language-related majors; the others had all taken a variety of other majors. The latter thought that their non-English major was a “bonus” because it enabled them to reflect on their own experience of English learning, especially the strategies contributing to specific English skill improvement. Subsequently, these teachers, according to the interview data with these teachers, felt capable of connecting the
challenges and problems in their students’ English learning with their own previous experiences, and thus better understand what the students needed to help them learn English better.

Perceptions: Why did the Participants Engage in Test Preparation?

A review of the interviews and questionnaire data showed that the different participants each had their respective reasons for engaging in test preparation. Students wanted to improve test scores because of the associated importance of high test scores, and/or were determined to “learn English better” after previously unsuccessful attempts at the test. The teachers reported that their primary focus was on convincing the students of the necessity of improving their English skills, and they were enthusiastic to share successful experiences in learning English and taking tests with students. The administrators, with the experience working in test preparation industry, emphasized the educational value of teaching and learning; in addition, these administrators, with their overseas educational experience, anticipated that students would “think carefully” before starting the test preparation journey. Both teachers and administrators also mentioned that receiving a decent salary was motivation for working in the test preparation industry.

All the students expressed the idea that the tests were very important because of the potential opportunities associated with the test scores. The students interpreted the TOEFL iBT as the gatekeeper for entrance into English-speaking academic programs, and thus attached great importance to their test score. As Liam (CL-S2) commented, “I have not taken this test before, but I know how important it is for my [academic program] application.” When asked to rate the importance of the CET4 to various participants and test uses on a scale from 0 (not applicable), 1 (not at all important) to 6 (extremely important), the students preparing for the CET4 rated the

---

5 CL-S2 is the ID for the student participant Liam, where CL stands for “Centre L”, S2 for “Student No. 2.”
importance “to myself” (M=5.41) and “to job applications” (M=5.39) as having the highest means among 10 items.

The top motivation students expressed for taking test preparation courses was to improve test scores because they understood the importance of higher test scores with respect to their future. The students in the TOEFL iBT test preparation courses admitted that they had very limited knowledge of the test except for the fact that it was an English language proficiency test and its score was a requirement for admission to academic programs in English-speaking countries. These students had heard many stories from their peers that high scores on the TOEFL iBT were attributed to the test preparation courses. When rating the reasons for taking CET4 test preparation courses—on a scale from 0 (totally disagree) to 6 (totally agree)—“to improve test scores” was rated with the highest mean of 4.87, followed by “to learn test-taking strategies” (M=4.62), which also was believed to contribute to test score improvement. It should be noted that the students in the TOEFL iBT test preparation courses who had previous test-taking experience also expressed in the interviews, that they believed the improved test scores would represent higher English proficiency. To these students, the motivation of “to improve test scores” is closely related to the motivation to improve English proficiency, or “to learn English better” in their words.

To “learn English better” was also a strong motivator for students. Those who repeated the test expressed this motivation more strongly than those who took the test for the first time or non-repeaters. When asked to comment on why they participated in the test preparation courses, the students repeating the TOEFL iBT used their failed attempts on the test to improve the specific English skills with which they struggled. Similarly, students repeating the CET4 gave higher ratings to the reasons “to improve English language proficiency” (M=4.70) and “to learn
strategies of English language learning” (M=4.62) than did students taking the CET4 for the first time did (M= 3.95 and M=3.92, respectively). The following comment made by Young (CL-S3), who had taken the TOEFL iBT three times, provides a clear explanation of these repeaters’ motivation for improving their English skills:

I mainly want to learn English better by taking this [TOEFL iBT test preparation] course. I have taken this test [the TOEFL iBT] too [so] many times, and I come to understand [from the previous test scores] that my English is not good. I also understand that there is no shortcut [to reach the expected score]. So I really want, from taking this course, to find English learning strategies that fit me and also to improve my English [skills] with these strategies.

With many years of teaching experience, the teachers at Centre L understood the students’ desires to achieve higher test scores by taking test preparation courses; and these teachers shared a belief that the TOEFL iBT test scores were an accurate reflection of the students’ English language proficiency. These teachers identified themselves as “professional” teachers with a strong desire to provide appropriate guidance to meet the students’ desires, rather than simply offering a “short-term remedy” to “crack” the test. However, the students commented—especially those at (CU-S6) Site 3 who attended test preparation centres other than Centre L—that whether or not teachers communicated such belief with students depended on the individual teacher. For the teachers at Centre L, the key to the appropriate guidance was to “tell the truth” of what contributed to test score increases to the students, as Leo (CL-T2) said, he communicated the idea with students that the increases in TOEFL iBT test scores were definitely related to the improvement on English skills.

All the teachers at Centre L expressed their enthusiasm for teaching, and they liked to help the students, as Kerry (CL-T1), the teacher of the TOEFL iBT listening module, said, “I like to help them [my students] with the test [the TOEFL iBT] just like the way that I was helped when

6 CU-S is the ID for all 12 student participants at Site 3, where CU stands for Canadian University, S for student.
I was preparing for the test.” The classroom observation and interview data showed that these teachers shared their experience in teaching, learning, test-taking, and overseas studies in their instruction to help students achieve their goals (e.g., improve test scores, improve English skills), by providing “appropriate guidance”. These teachers directly taught students the strategies of improving particular English skills that they summarized from their experiences of teaching different students and the personal experience of learning English. As Lindsay (CL-S1) recalled, each of the four teachers in TOFEL iBT course A introduced “the four-step strategy” to improve the respective English skills (listening, reading, writing, and speaking); these teachers expressed in the interviews that they had seen these strategies helped many students they previously taught. Kerry (CL-T1) told stories of success in learning English with the strategies to improve English skills, such as training a rock singer with zero background in English to become a teacher of the general English course at Centre L. In class, these teachers also told stories of their own learning as a way to encourage students not to be discouraged. For example, Henry (CL-T3) told the students his experience of English learning, from starting English learning at school in a small town with limited resources (e.g., non-certified teachers of English, no training in listening and speaking skills), to attending English training schools after moving to a large city after graduating from high school, until successfully entering into a college to study English language major. When Henry was asked in the interview why he told such personal experiences, Henry said he believed such experiences helped the students keep encouraged in English learning. Additionally, these teachers included their TOEFL test-taking experience in the instruction to provide the first-hand information related to the test-taking process. It should be noted in the interviews with these teachers that, they repeatedly emphasized the motivation of teaching test preparation courses because they wanted to help students through the provision of appropriate
guidance; however, these teachers further added that their guidance needed the concerted efforts from students to learn knowledge and strategies relevant to the test and English skills.

When asked what motivated them about their work related to test preparation courses, the three administrators at Centre L—two supervisors and one educational consultant—emphasized their understanding of the educational value of teaching and learning. They recalled their previous experiences of either working or applying for positions at various test preparation centres to explain how they understood the educational value. They explained that, in order to secure a financial advantage in the competitive market of the test preparation industry, some test preparation centres accept that some teachers frequently use the methods, such as singing, dancing and talking jokes, to give the students the impression that “it was an interesting class or teacher”. The administrators agreed that such methods were acceptable if used at an appropriate degree rather than used to predominate the instruction, and further added that knowledge, experience, and pedagogy is what should impress students, and the students’ sense of achievement should derive from the improvement in their English skills and the corresponding increase in test scores. The following comment by James (CL-A3), who obtained a Master’s degree in education in United States, spoke of his understanding of educational value, and the relationship between “educational” and “profitable”:

I think, for a healthy educational institution, the attribute of “educational” should be placed first. It should be, because of its excellence in education, the institution then could have its second attribute of “profitable.” The order of these two attributes should not be reversed.

Kerry’s (CL-T1) previous experience of working at a national-chained test preparation centre highlighted the different belief with respect to the relationship between “educational” and “profitable” among test preparation centres. Kerry said in the interview that the test preparation centre he used to work for continuously increased the scale for larger business profits, and many
teachers were recruited to fill the increasing teaching positions. As a consequence, the recruitment criteria were scarified for the demand of teachers, and Kerry said some teachers used some “tricks” that he felt “low taste” to interest students and to give students the impression that “it was quite a different experience compared to those at school.” The students who looked for an experience different from school thus contributed to the business profit of this centre; however, as Henry thought, what students learned at this centre were diluted by the tricks used throughout the instruction.

With personal experience of having studied in English-speaking countries, the administrators thought they were able to communicate to students “the big picture” behind the TOEFL iBT and overseas study. The administrators wanted to use their job responsibilities, such as supervising TOEFL iBT test preparation courses or providing educational consultation, to ask students to think about questions like “why do you want to take the TOEFL iBT?” and “why do you want to study abroad?”, or sometimes even “is studying abroad really what you want?” These questions highlight the administrators’ motivation for working at test preparation centres: to make the students “think carefully” before starting test preparation, specifically about their own goals, and about whether they were simply following others’ footsteps or genuinely taking steps of their own.

Every one of the teachers and administrators admitted that they understood the great potential of the test preparation industry in China; the income received for working in this industry offers teachers and administrators a decent means of “making a living.” For example, in a two-week intensive TOEFL iBT test preparation course, each teacher who taught eight sessions within one module of the course earned an average of 1,500 RMB (Chinese currency) per session for a total of 12,000 RMB. This sum is approximately equivalent to $1,700 US dollars (the
the exchange rate of US dollars to RMB was roughly 1:7 at the time of writing). In comparison, when I taught eight sessions per week as an English teacher at a university in northern China in 2010, and my monthly salary was around 3,000 RMB, or roughly $430 US dollars. In other words, these teachers’ bi-weekly payments were nearly four times higher than the monthly payment of a teacher at a university.

While no specific information was available on the supervisors’ and the consultant’s salary, it was shared during the interviews that the educational consultant was paid a base salary plus commission for his consulting service fees (that the student clients paid to Centre L). Considering the rapid growth of Centre L in recent years, it is reasonable to believe that jobs related to test preparation courses are able to not only meet these teachers and administrators’ needs of “making a living”, but surpass mere need and offer the opportunity to make a “great” living. Meanwhile, such considerable salaries put pressure on some teachers to be successful in raising students’ test scores, as Will (CL-T8) expressed that he needed to be responsible for what students paid for the test preparation course.

**Context: How did the Participants Choose Their Test Preparation Practices?**

The findings from the interviews and the questionnaire showed that the students in this study chose among different types of test preparation courses and different course timing based on their particular focus (e.g., the test or the specific English skills) and/or scheduling needs (e.g. timelines). The teachers and the administrators said in the interviews that, among the many test preparation centres on the market, they chose to work in the environment that best supported both their teaching beliefs and their financial needs. The students looked for the fit in types and timing for the choice of test preparation practices, and the teachers and the administrators looked for a practical and professional fit for their choices of test preparation practices.
In the market of test preparation industry, there are different test preparation options to meet students’ needs, such as duration (short-term, long-term), class size (VIP private class, small-size class—less than 10 students), large-size class—larger than 100 students), and location (local class, residential (camp-like) class) (Jiang, 2007). The TOEFL iBT students were asked in the interview how they made the choices among the options available to them; and the CET4 students were asked by the questionnaire about the prevalence of their choices.

**Types.** Test preparation courses for the TOEFL iBT and the CET4 are available in a number of different types: TOEFL iBT test preparation courses vary in duration, class size, and location while CET4 test preparation courses differ in instructional focus. The TOEFL iBT test preparation courses described by all those engaged in test preparation practices were either intensive (completed in two weeks), or extended (lasting two or three months). Fifteen of the student participants (out of 18) taking TOEFL iBT test preparation courses took the intensive type, and the other three took the extended type. The students explained that their choices were related to their entire plan for academic applications. The students choosing the intensive type intended to obtain the TOEFL iBT score first before moving onto preparing other requirements for their academic applications (e.g., personal statement, study plan, academic writing samples, or academic portfolios). The students choosing the extended type regarded the TOEFL iBT as one of many parallel steps in the academic application process.

The teachers and the administrators admitted that the options in duration did meet the students’ different scheduling plans. However, they recalled many students talking about stories of substantial test score increases after completing intensive test preparation courses. As such, the teachers and administrators interpreted the large number of students taking the intensive type
as a reflection of the students’ strong desires to increase their test scores “in a short period of time.”

Additionally, Hayden (CL-A2), the supervisor in chief at Centre L introduced, there were generally three class sizes for the TOEFL iBT test preparation courses: large classes, with over 100 students per class; medium classes, with 20–100 students per class; and small class, with fewer than 20 students per class. Hayden continued that, the students who expected to receive “more attention” and experience “closer contact” with the teachers chose the course with smaller student numbers, but they paid three or four times more than those in the larger classes. Some students chose to travel from other cities in order to take the test preparation courses they believed to have the “professional teachers and quality instruction” they sought, and enrolled in residential type courses. Even though no student participants from such type of course provided evidence, the expense inherent in choosing such type of courses probably made students have great expectations about the outcomes. Other students, however, chose local test preparation centres because they preferred the “camp-like” study experience in which to immerse themselves for test preparation. Although Centre L offered different types of TOEFL iBT test preparation courses by duration, size, and location, Hayden mentioned that the students experienced the similar instructional contents across them all: the integrated components of test-taking strategies, English skill improvements, practicing with retired test items, and engaging in mock test tasks (e.g., in-class exercises of timed speaking tasks). With his previous experience of working at other test preparation centres, Hayden added that teachers and teaching methodology really made a difference in the outcomes that students could achieve, although the instructional contents were similar among test preparation centres. This echoed to the point that students made in the
interviews that, they followed the particular teachers who established good reputation in teaching test preparation courses from the “word of mouth” of previous students.

Roughly 80% of the student participants (n=431, 80.7%) taking CET4 test preparation courses reported the course type they chose, which was as follows: 34.1% chose the CET4 test-taking strategies type, 26.2% took the English skills for the CET4 type, 23.2% registered for the retired CET4 test-items type, 12.8% opted for the CET4 mock test type (12.8%), and the remaining 3.7% selected the CET4 vocabulary type (3.7%). It can be seen in the breakdown by type that the CET4 test preparation courses were mainly classified by instructional focus. Although the Chi-square tests of independence did not find a statistical relationship between the students’ background characteristics (i.e., years of English learning, status of test repeaters) and their course type choices, some noticeable preferences were observed related to the status of test repeaters, which is presented in Table 12.

Table 12

<table>
<thead>
<tr>
<th>Types</th>
<th>Non-repeater</th>
<th>Repeater</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>English skills</td>
<td>64</td>
<td>49</td>
<td>113</td>
</tr>
<tr>
<td>Retired test items</td>
<td>55</td>
<td>45</td>
<td>100</td>
</tr>
<tr>
<td>Mock test</td>
<td>38</td>
<td>17</td>
<td>55</td>
</tr>
<tr>
<td>Test-taking strategies</td>
<td>100</td>
<td>47</td>
<td>147</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>10</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>Total</td>
<td>267</td>
<td>164</td>
<td>360</td>
</tr>
</tbody>
</table>

As shown in Table 13, the non-repeaters preferred the test-taking strategies type (n=100) to the English-skills type (n=64) and the retired test items type (n=55); however, the repeaters had roughly equal preferences to these three types, with 47 choosing the test-taking strategies type, 49 selecting the English-skills type, and 45 taking the retired test items type. These observations indicate that, among the students taking the CET4 for the first time, more cared about test-taking strategies, and learning how to be more effective at taking the test was a priority, because these
students might already have some confidence in their English proficiency, which indicated by their outperforming performance on the NMET than the CET4 test repeaters. Among the students repeating the CET4, however, attention was equally to English skills and test-taking strategies, and their previous test-taking experience (e.g., unsuccessful test scores), in addition to their English language proficiency indicate by their NMET scores, could be the contributing factor for this shift.

**Timing.** Timing refers to the time students chose to take test preparation courses. It was found that the timing was related to the test repeaters, with test repeaters planning a longer time, compared with non-repeaters, to start taking test preparation courses prior to their test dates.

The students without previous TOEFL iBT test-taking experience said that they planned to start taking the TOEFL iBT test preparation courses at least two or three months before their registered or tentative test dates because they needed such a timeline to become familiar with the test, and also to practice the specific skills that needed improving. All the teachers, based on their years of teaching experiences, agreed that two months was a reasonable minimum period of time to prepare for the TOEFL iBT, and each teacher concluded his/her instruction module by offering a 2-month preparation plan for the respective test section.

In contrast, some students with previously unsuccessful TOEFL iBT scores planned to start test preparation courses earlier so they could have more time prior to their next test attempt. These students wanted to learn some strategies to help them improve their English skills by taking test preparation courses. For example, Young (CL-S3) took the TOEFL iBT three times and none of the three TOEFL iBT scores passed 90, the required minimum score for the universities to which he intended to apply. Young said he planned half a year to exclusively concentrate on preparing
for the TOEFL iBT, and he wanted to “learn English better and find strategies that fits” his learning style, in order to prepare to take the TOEFL iBT for the fourth time.

A total of 465 students reported the timing of their taking CET4 test preparation courses. Roughly half (n=268, 57.6%) took their test preparation courses one or less than one month before the test date, the other half were divided almost equally between taking their course two to three months in advance (n=95, 20.4%), and more than four months (n=102, 21.9%) prior to the test date. The Chi-square tests of independence examining the relationship between the students’ background characteristics (i.e., years of English learning, status of test repeaters) and their decision around timing found a statistical significant level, $\chi^2 = 21.86, p < .000$, for the students who repeated the CET4. As shown in Table 13, among the students who took the CET4 for the first time, significantly more chose to take test preparation courses within the month prior to the test date (n=176, 58.7% of total non-repeaters). At the same time, among the students who repeated the CET4, significantly more students chose either within a month (n=92, 56.1%) or more than four months (n=53, 32.3% of total repeaters) for the timing of their test preparation courses.

Table 13

<table>
<thead>
<tr>
<th>Course types</th>
<th>Non-repeater</th>
<th>Repeater</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤ 1 month before test</td>
<td>176</td>
<td>92</td>
<td>268</td>
</tr>
<tr>
<td>2-3 months before test</td>
<td>75</td>
<td>19</td>
<td>94</td>
</tr>
<tr>
<td>≥4 months before test</td>
<td>49</td>
<td>53</td>
<td>102</td>
</tr>
<tr>
<td>Total</td>
<td>300</td>
<td>164</td>
<td>464</td>
</tr>
</tbody>
</table>

These significant relationships indicate that both the majority of students who took the CET4 for the first time and those who repeated the CET4 tended to take test preparation courses close to the test date. However, there were a number of CET4 test repeaters who started to take test preparation courses much sooner, a timing pattern similar to that observed among the student
participants who repeated the TOEFL iBT. A comparison of the CET4 test scores between non-repeaters and repeaters further suggested that the timing of taking a test preparation course was related to their level of English language proficiency, where the CET4 non-repeaters achieved higher test scores (M=453.2, n=349) than the repeaters (M=391.0, n=184). It is reasonable to believe, therefore, that the CET4 test repeaters interpreted their low test scores on previous attempts as corresponding to their level of English language proficiency; such an interpretation may have motivated them to start earlier or plan more preparation time to prepare for retaking the CET4, compared to the CET non-repeaters.

Working environment. In addition to the findings of how the students chose test preparation courses, the interview data with the teachers and the administrators at Centre L also revealed how these two groups of participants make their choices of test preparation centres where they liked to work. All eight teachers and three administrators mentioned that they chose to work at Centre L because the values the Centre holds are in alignment with their own teaching beliefs. Specifically, the findings from the interviews with the administrators showed that Centre L made efforts to keep both the educational value in teaching and learning and commercial benefit in its vision. The public lectures that Centre L offers at local universities attempt to debunk the selling points commonly used in the test preparation industry, including the myths of “guaranteed scores,” and “magical methods to crack test items.” Instead, the lectures emphasize that students themselves need to learn and that the Centre has the experienced and professional teachers and administrators to help and guide them to do just that. The findings from the interviews with the teachers were consistent with what was emphasized in the public lectures, that students were encouraged to make efforts to learn along with the guidance that teachers provided.
Centre L set a relatively high salary rate among local test preparation centres to recruit “professional” and “experienced” teachers, seeking those who demonstrated a thorough understanding of and excellent performance on the TOEFL iBT (e.g., test tasks, test scores, overseas study experience), and who had a strong profile of teaching TOEFL iBT test preparation courses (e.g., evaluation of previous teaching experience). Centre L offered a good salary in order to attract the best teachers to work here. Interviews with the teachers and administrators revealed that the highlighted educational value and competitive salary led them to believe that both “educational” and “profitable” could be reached at the same time, which, they pointed out, was not always the case at many other test preparation centres in China.

Practices: What did the Participants Do to Prepare for the Test?

Three themes were identified reflecting the participants’ major test preparation practices: test familiarization, English skill improvement, and stress management.

Test familiarization. Because all students aimed to improve test scores—their top motivation for taking test preparation courses—test familiarization as a practice is well addressed by test preparation courses, and serves to familiarize students with the test and practice the tasks for each test section. Both teachers and administrators agreed that sufficient knowledge about, and practice for the test are the essential components for instruction of test preparation courses.

The knowledge about the test presented to students during test preparation courses can be generally summarized as the answer to the following two questions: “What is the test?” and “What does the test look like?” The teachers and administrators at Centre L made conscientious efforts through instruction or supervision to communicate with the students that the TOEFL iBT “evaluates whether you can perform and study in an English-speaking environment.” The teachers in particular understood that many students had the misconception that some test-taking
strategies could “crack” the TOEFL iBT. Therefore, to help students “re-think” their preparation, the teachers emphasized the clear communication of what the TOEFL iBT actually is: a test for which they need to “seriously prepare” rather than view it simply as one that is “tricky crack.”

All the teachers at Centre L communicated this “appropriate” understanding of the TOEFL iBT in their first session of the course, in order for students to accept and follow their instruction.

Knowledge about the test features was also introduced to help students understand what the test looked like. The features mainly included: (1) the sections (e.g., listening, reading, etc.); (2) the structure (e.g., number of test tasks, order of tasks, designated time to complete test tasks); (3) the formats (e.g., multiple-choice questions, short-answer questions, essay); (4) the required skills (e.g., summarizing ideas, drawing inferences, demonstrating arguments); and (5) scoring criteria if applicable (e.g., scored samples of speaking and writing tasks). The introduction of these test features gave the students a concise overview of the TOEFL iBT, which many found difficult to obtain by themselves because they felt they had “no where to start” in the hundreds of pages of the Official Guide to the TOEFL iBT.

The practice of test tasks in each test section was another important component in test familiarization. The teachers at Centre L were observed to demonstrate the necessary steps for completing tasks, and led in-class drills for the students to familiarize themselves with these steps. This kind of demonstration-drill pattern was observed in all eight teachers’ instruction practices, and the students found that the targeted demonstration-drill pattern helped them understand “how to complete test tasks.” For example, when introducing the independent writing task, Henry (CL-T3), one of the TOEFL iBT teachers, first analyzed the task direction, then reviewed the scored writing samples to direct the students’ attention to the two important aspects in the scoring criteria: (1) which ideas are presented in the writing sample? and (2)
how are these ideas logically connected (e.g., connecting words/sentences)? He then demonstrated the two steps to completing the writing task: (1) forming ideas, and (2) presenting ideas in a clear logical way. Although Henry’s demonstration was targeted to preparing for the specific writing task, this approach was rated highly by his students for practicing English writing skills. As Liam (CL-S2), a student in Henry’s class, recalled,

The method that Henry taught us—the logical structure writing—I have never heard that before [from my teachers at school]. The writing method I heard most was to use certain template—presenting topic at the beginning, using the words like ‘first’ and ‘second’ to continue. Henry emphasized to use our own logical thinking [emphasized tone] to write, to present the ideas in ascending level of importance. This is actually [the way used] in the full-mark writing samples in the Official Guide. I think, this method changed my ideas on how I write [in English]—writing should not be made up of, should not rely on the ready-to-use [writing] templates, he teaches me that writing should be based on my own thinking.

The classroom observation data and the interview data with the teachers of the TOEFL iBT test preparation courses also showed that the test-specific strategies were included when practising test tasks. With the teaching experience and the first-hand test-taking experience, the teachers at Centre L tended to review and analyze the test to “have a thorough understanding.” By reviewing and analyzing relevant resources, including the Official Guide to the TOEFL iBT, the TOEFL iBT Practice Online (TPO), and released sample test items, these teachers summarized key strategies for the specific test tasks. However, they also made it very clear that the test-taking strategies they summarized should be considered supplementary to developing strong English skills, rather than as the primary “trick.” Leo (CL-T2), one of the teacher participants, developed a strategy for answering multiple-choice reading task questions, based on one pattern he noticed. He recalled:

The distracting options [of a multiple-choice question] usually use the exact words, phrases, sentence structures, or even the logic of the text; however, the key information in these distracting options have been changed. Although the key information seems unobvious, it actually examines whether you understand the
corresponding part in the text. So I told my students to be cautious about the options having the exact words, phrases in the texts, because they are probably distracting options.

When practicing this strategy in class, Leo encouraged the students to first talk about the reason for choosing an option (of the multiple-choice question), and then he identified the distracting information and the underlying problems of choosing the distractors (e.g., difficulty in distinguishing the logical relationship). Using this explanation, Leo advised his students to use this strategy properly. However, Leo said that not all the students took his advice, especially those students who had very limited time before their test dates and those who still expected to increase test scores by using “tricks.”

Of all the test preparation practices the students were asked in the questionnaire (Appendix H), there were the practices that taught the students test-specific strategies for each CET4 test section, although there was no direct observation or interview data to demonstrate how the additional test features were presented to them. I present these test-specific practices in Table 14. As shown in this table, the students taking CET4 test preparation courses rated the practices targeted at completing test tasks with the highest means among all the practices for each particular test section. For example, the students rated the highest means for the practices targeted for the CET4 reading tasks, including 17g “reading questions before looking for key words and sentence of questions” (M=4.70), 17i “scanning key words in text before reading and understanding individual sentences” (M=4.61), and 17h “searching for answers in the text according to the sequence of questions” (M=4.54). These practices were the test-specific strategies, and the highest means indicated that such strategies were a major component in the CET4 test preparation practices. These high means also meant that the students paid great attention to these strategies.
### Table 14

**Descriptives of the Test-Specific Practices by CET4 Test Section (n=534)**

<table>
<thead>
<tr>
<th>Practice</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Listening</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16h. Going over the options beforehand to focus my attention accordingly</td>
<td>522</td>
<td>4.60</td>
<td>1.51</td>
<td>-1.00</td>
<td>.35</td>
</tr>
<tr>
<td>16f. Finding ways to keep concentrated on listening tasks</td>
<td>520</td>
<td>4.44</td>
<td>1.50</td>
<td>-.85</td>
<td>.23</td>
</tr>
<tr>
<td>16e. Noting down key words and important information</td>
<td>517</td>
<td>4.15</td>
<td>1.61</td>
<td>-.64</td>
<td>-.26</td>
</tr>
<tr>
<td><strong>Reading</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17g. Reading questions before looking for key words and sentence of questions</td>
<td>522</td>
<td>4.70</td>
<td>1.45</td>
<td>-.99</td>
<td>.25</td>
</tr>
<tr>
<td>17i. Scanning key words in text before reading and understanding individual sentences carefully</td>
<td>524</td>
<td>4.61</td>
<td>1.43</td>
<td>-.89</td>
<td>.21</td>
</tr>
<tr>
<td>17h. Searching for the answers in the text according to the sequence of questions</td>
<td>524</td>
<td>4.54</td>
<td>1.47</td>
<td>-.86</td>
<td>.03</td>
</tr>
<tr>
<td><strong>Writing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18f. Keeping my writing neat and tidy</td>
<td>487</td>
<td>4.55</td>
<td>1.48</td>
<td>-.83</td>
<td>.06</td>
</tr>
<tr>
<td>18i. Checking grammar and spelling mistakes</td>
<td>492</td>
<td>4.50</td>
<td>1.49</td>
<td>-.90</td>
<td>.30</td>
</tr>
<tr>
<td>18j. Paying attention to unfamiliar words and grammar</td>
<td>486</td>
<td>4.46</td>
<td>1.55</td>
<td>-.85</td>
<td>.05</td>
</tr>
<tr>
<td><strong>Translation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19c. Checking spelling or grammar mistakes</td>
<td>491</td>
<td>4.47</td>
<td>1.53</td>
<td>-.87</td>
<td>.18</td>
</tr>
<tr>
<td>19d. Paying attention to using unfamiliar words</td>
<td>493</td>
<td>4.41</td>
<td>1.55</td>
<td>-.87</td>
<td>.20</td>
</tr>
<tr>
<td>19a. Reviewing retired translation test items</td>
<td>494</td>
<td>4.30</td>
<td>1.72</td>
<td>-.76</td>
<td>-.41</td>
</tr>
<tr>
<td>19b. Reviewing frequently-assessed key sentence structures and phrases</td>
<td>493</td>
<td>4.22</td>
<td>1.59</td>
<td>-.72</td>
<td>-.12</td>
</tr>
<tr>
<td><strong>Cloze</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20e. Using logical lines of passages</td>
<td>490</td>
<td>4.41</td>
<td>1.50</td>
<td>-.87</td>
<td>.27</td>
</tr>
<tr>
<td>20c. Grasping the main idea of the passage</td>
<td>487</td>
<td>4.39</td>
<td>1.54</td>
<td>-.91</td>
<td>.38</td>
</tr>
<tr>
<td>20b. Making inference from the context</td>
<td>491</td>
<td>4.39</td>
<td>1.52</td>
<td>-.80</td>
<td>.11</td>
</tr>
<tr>
<td>20a. Looking for frequently-assessed questions</td>
<td>492</td>
<td>4.14</td>
<td>1.66</td>
<td>-.65</td>
<td>-.35</td>
</tr>
</tbody>
</table>

**English skill improvement.** Practices to improve English skills were also included in test preparation courses and were strongly emphasized by the participants who believed that the improvement of English skills ultimately contributed to an increase in test scores. The practices targeted at improving English skills included: reviewing the English skills required in test tasks,
demonstrating strategies to practice English skills, and providing additional resources for English skill improvement.

All eight teachers at Centre L had explicit instructional focus on the specific English skills required in their respective teaching modules. Grammar was one such focus in both the reading and writing modules because the teachers believed that grammar was the foundational skill required to understand texts and communicate ideas. For example, compound and complex sentences were reviewed in the reading module to identify relationships between and among sentences in the reading tasks; complex and compound sentences were also reviewed in the writing module to highlight “making effective sentences” in writing. Pronunciation is another example, and was emphasized in the listening and speaking modules because listening tasks and speaking tasks required “accurate pronunciation.”

In addition, these teachers demonstrated detailed strategies of how to practice English skills. The teachers did not claim that the methods were “something new” about English language learning; rather, such methods were “something effective” with which to promote students’ English language learning. Lindsay (CL-S1), a student, said all four teachers of the course she took demonstrated a “step-by-step method” for improving respective English skills. She recalled the “four-step intensive listening” that her teacher, Kerry (CL-T1), demonstrated in the listening module: (1) read transcripts silently (without listening), (2) read transcripts aloud (without listening), (3) read along with the transcripts, and (4) listen and repeat. Kerry explained how each step practiced listening skills: reading silently helps the students to clear the “obstacles” in vocabulary and pronunciation; reading aloud trains the coordination of the vocal/acoustic organs (such as lips, teeth, and tongue) so students can read the transcript fluently and clearly; reading along focuses on speed, intonation, pronunciation, and fluency; and listening
and repeating allows the students to (self-)evaluate the extent of their skills in recognition (e.g., recognition of vocabulary and pronunciation) and in understanding (e.g., understanding key ideas).

Similarly, the questionnaire data from the CET4 test preparation practices also showed that the practices targeted to improve general English skills were included in the instruction of CET4 test preparation courses, even though this data (e.g., means) only revealed what these practices were, not how they were performed.

Table 15

*Descriptives of General Practices for English Skills in CET4 Test Preparation Courses*

<table>
<thead>
<tr>
<th>Practice Description</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>16c. Reading listening transcripts to enhance understanding</td>
<td>520</td>
<td>3.92</td>
<td>1.66</td>
<td>-.60</td>
<td>-.24</td>
</tr>
<tr>
<td>16a. Repeatedly listening to listening materials to familiarize pronunciation and intonation</td>
<td>527</td>
<td>3.91</td>
<td>1.78</td>
<td>-.45</td>
<td>-.75</td>
</tr>
<tr>
<td>16b. Trying to understand listening materials that I have practiced</td>
<td>522</td>
<td>3.70</td>
<td>1.79</td>
<td>-.44</td>
<td>-.67</td>
</tr>
<tr>
<td>17d. Focusing on understanding difficult and complex sentences in the passages</td>
<td>518</td>
<td>3.77</td>
<td>1.64</td>
<td>-.50</td>
<td>-.38</td>
</tr>
<tr>
<td>17a. Memorizing vocabulary including parts of speech, roots and pre-suffixes</td>
<td>521</td>
<td>3.67</td>
<td>1.82</td>
<td>-.28</td>
<td>-.88</td>
</tr>
<tr>
<td>18g. Using more advanced vocabulary</td>
<td>495</td>
<td>4.12</td>
<td>1.64</td>
<td>-.69</td>
<td>-.20</td>
</tr>
<tr>
<td>18h. Using more complex sentences</td>
<td>489</td>
<td>4.10</td>
<td>1.59</td>
<td>-.73</td>
<td>.07</td>
</tr>
<tr>
<td>20d. Understanding complex sentences</td>
<td>492</td>
<td>3.97</td>
<td>1.55</td>
<td>-.55</td>
<td>-.18</td>
</tr>
</tbody>
</table>

As shown in Table 15, among the test preparation practices for CET4 test sections of listening, reading, writing, and cloze, the students rated general practices targeted at improving the corresponding English skill with lower means than they gave the test-specific practices reported in Table 15. For example, the lower means of two practices that aimed for improving reading skill—17d “understanding complex sentences” (M=3.77) and 17a “memorizing
vocabulary including parts of speech, roots and pre-suffixes” (M=3.67)—indicated these practices were not as well attended to by students as they prepared for the CET4. This pattern (i.e., lower means on the general practices for particular English skills) was consistent across the test preparation practices for the test sections of listening, writing, and cloze. That the students placed emphasis on the test-specific practices more frequently than they did the more general practices reflected their primary motivation for taking the CET4 test preparation courses—to improve test scores—even though the general practices were a component in the CET4 test preparation courses and were available for the students.

It was understood by both the students and the teachers at Centre L that practices for improving their English skills required additional support beyond in-class instruction of test preparation courses they took; therefore, extra resources were provided through a series of courses by Centre L that students could chose to attend. Two supervisors mentioned that Centre L offered a series of English skill courses, including English Grammar, Core Vocabulary in Reading and Writing, Learning English through Watching Movies, and Oral English. These courses were designed to reach students in addition to those preparing for various English language tests. However, it turned out that many of the students in the test prep courses subsequently registered for one among the series because they realized “some aspects of English skills” still needed improving. English Grammar received the most attention because many students thought they needed to improve their command of grammar. The course description stated, “this course will tell you why grammar rules are formulated and how to apply these rules [italicized indicating the exact English words used in original course description in Chinese].” Hayden (CL-A3) said this course received the most positive evaluation, and he remembered the students’ comments on the evaluation forms that this course helped students understand the
English grammar rather than simply have them reciting English grammar rules. One particular evaluation comment that Hayden showed in the interview read, “I start to love grammar after taking this course. I take a look at emails that were written in English last week, and I am enthusiastically and excitedly revising the emails,” suggesting that the student’s efforts to reinforce the learning of grammar rules.

**Stress management.** Stress management was also found to be a primary practice for preparing for high-stakes English tests. This practice was named as ‘stress management’ because it included all the strategies to help the students boost confidence and relieve anxiety, and these strategies made contributions for the students to manage their frequently reported ‘stress’ that was related to both English learning and test taking.

The stress and the anxiety that students experienced were well understood and attended to by both teachers and administrators. The interview data and the classroom observation data revealed that test preparation is an “unpleasant” and “very tiring” process because of the intensity and consequences of test preparation, and repeated unsuccessful attempts could turn English skill improvement into a “frustrating” process. Therefore, managing stress has become an important component in the practices to help students achieve success.

All eight teachers at Centre L understood that TOEFL test preparation was a painstaking process that requires constant encouragement and continuous confidence; therefore, the teachers paid attention onto building the needed confidence in their instruction. To encourage their students, therefore, they shared their own experiences of preparing for the TOEFL pBT (paper-based test) and/or iBT. At the very beginning of his module, Kerry (CL-T1) shared his entire process of planning to studying abroad, in particular discussing the challenges (e.g., requirement for a very high TOEFL pBT score) and his frustrations (e.g., doubts about himself) in test
preparation. He believed his experience would resonate with situations his students might encounter, and his personal experience conveyed a confident message: “there is definitely a way out there,” as long as the students persisted. Confidence was also built through public lectures and in-class instruction, and many successful cases were shared demonstrating how high TOEFL iBT scores were achieved by targeted study plans and within reasonable timelines.

Voicing the thoughts of many students, a banner on Centre L’s website asks: “You are determined to learn English better but feel like giving up again?” The teachers and the administrators interviewed all shared the idea that building confidence was critical for contributing to English learning, and best achieved through hearing about experiences of success. Many students were impressed by how their teachers became successful English learners. Leo (CL-T2) and Henry (CL-T3), Centre L teachers who grew up in small towns, presented themselves as examples of successful English learning regardless of where one received early English education. Teachers Kerry (CL-T1) and Megan (CL-T4), who studied arts programs, presented themselves as successful cases of English learning regardless of program background. These teachers’ personal English learning experiences confidently portrayed to students the message: “If I can do it, you can as well.”

With years of teaching test preparation courses, the teachers at Centre L understood that the intensive schedule would influence students in many ways, e.g., fatigue, attention span, and anxiety. Therefore, they added “in-class breaks” to adjust the pace in class and to reduce the students’ fatigue and anxiety. These teachers typically used the extended discussions on a particular point/idea as such breaks. For example, Henry (CL-T3) strategically used an anecdote about himself and the local people of his hometown as a way to “give students’ minds a break.” He would first give instructions on presenting the main ideas in a writing task. He then would tell
the story about when he was invited by the Village Committee of his hometown to talk about his life in the big city and to give suggestions on local development, presenting three points as the main ideas of his talk. He then carried on to tell the students about his town’s local traditions about marriage and funeral ceremony that he disliked. The students showed great interest in his stories and burst into laughter three times during this three-and-half-minute break.

Some teachers’ breaks to mitigate stress and fatigue were “totally for fun”. These teachers, with years of teaching test preparation courses, understood the intensive schedule would influence students on many aspects, e.g., stress, fatigue, attention span, and mood. Some teachers said in the interviews that, “It is too tiring [for students] to sit through 2.5 hours every class which requires full concentration, and it is even harder to insist in this way through the whole course;” and “everyone’s attention might wander unconsciously every 20 minutes and it is impossible [for me] to take break every 20 minutes.” Therefore, these teachers inserted some episodes as in-class breaks to help reduce students’ fatigue and anxiety. For example, Will (CL-T8), was a fan of travelling and backpacking and used to be a tourist guide and an outdoor activity leader. He shared many of his traveling experiences with his students during class, including his excitement about beautiful sceneries, the stress and difficulties he encountered on self-guided tours, and the joy of making friends when traveling. He knew that the students needed a good atmosphere in class, and sensed the stress, anxiety, and fatigue from students’ reactions and responses. Therefore, he took time to discuss relaxing topics as a way to “make students happy.” A happy mood, these teachers believed, would help reduce anxiety and stress, and also promote learning.

One common idea shared among these teachers needed addressing, that these teachers emphasized they paid particular attention onto the frequency (how many times per session) and
the duration (how many minutes per break) of such breaks. They mentioned the episodes used as in-class breaks should be at the appropriate proportion of the entire class time; and they did not want to “linger too far away” from the focus of their instruction, rather they used these episodes to adjust the pace in class.

Section Summary

In this section, I presented my findings on the nature of test preparation practices in detail, with the descriptive structure of who, why, and what. Yin (2009) argued that the descriptive structure was appropriate to identify links among the data, and I combined this approach with the themes, categories and codes identified in Figure 2 for a concise understanding of how they were related to each other.

Figure 2. Themes, categories, and codes reflecting the nature of the test preparation practices

7 The data that contributed to generating themes, categories and codes in this figure were primarily from the qualitative data related to test preparation practices for the TOEFL iBT.
In Figure 2, each participant group was assigned a colour to indicate the categories and the coded that were relevant to the respective group, with green for administrators, blue for teachers, and magenta for students. The categories and the codes were presented in the colour-shaded boxes with the colour corresponding to the respective participant group. For the categories and the codes that were relevant to multiple groups of participants, a multiple-colour-shade was used.

As shown in Figure 2, the major relationship that emerged was that the who (those participants with a stake in how the test preparation practices were perceived, implemented and the outcomes achieved) and the why (reasons these participants engaged in test preparation practices and supported test preparation) influenced the what (what these participants did to contribute to test preparation practices). Specifically, the key groups of the participants (i.e., the students, the teachers, and the administrators) came to test preparation centres with a variety of background characteristics in teaching, learning, and test-taking experience; the participants’ motivations for, and perceptions of engaging in test preparation practices were grounded in these characteristics. Subsequently, all three groups of participants were involved in three major practices (i.e., test familiarization, stress management, and English skill improvement). The extent of their involvement (what), however, varied because of the differences in their characteristics (who), and their motivations and perceptions (why). The extent of their involvement further influenced whether these practices made focused contributions to test preparation for the specific test or extended contributions to English language learning.

**Perceived Effects of the Test Preparation Practices**

The perceived effects were explored through studying the relationship between test preparation practices and the students’ TOEFL iBT and CET4 test scores, both the perceived relationship reported through accounts by all the participants preparing for the TOEFL iBT, or
the observed relationship identified by the questionnaire and test score data from the students preparing for the CET4.

**Perceived Effects on the TOEFL iBT Test Scores**

Neither the teachers nor the administrators at Centre L thought that test preparation practices could produce immediate improvements in test scores if the students took the TOEFL iBT immediately after completing the test preparation courses (particularly the two-week intensive type of test preparation). They believed any potential increase in test scores should parallel improvement in English skills, and it was “unrealistic” to raise English skills to a higher level within a short period of time. Therefore, all eight teachers at Centre L guided their students to have reasonable expectations with respect to test score gains, recommending that they plan sufficient time (e.g., at least two months prior to taking the test) for practicing their new strategies for improving their English skills on a regular basis, if they wish to observe the potential effects of increasing their test scores. As Leo (CL-T2) commented,

> If you scored under 80 in TPOs (Test of Practice Online), don’t expect to increase your scores by learning certain test-taking strategies in this course, especially in such short period of time. Those strategies are based on a solid foundation of English proficiency.

Some students expressed their perception that the strategies they learned had a potentially positive effect on improving their English skills. For example, when commenting on the “four-step intensive listening” strategy, Lindsay (CL-S1) explained,

> I just started practicing this strategy, and I noticed that the step of repeating [fourth-step in this strategy] helps me improve the skill of recognition; and I believe it will have effects on improving English [listening] skills if I can continue practicing for longer time, and ultimately help with the TOEFL iBT test scores.

Some other students who were repeating the TOEFL iBT agreed that their test scores did not increase right after they completed test preparation courses. Young (CL-S3) took test preparation courses for the first two out of three times he took the TOEFL, and wrote the test
right after he completed the courses. None of his three total scores was over 90; therefore, he learned the lesson that test scores could not be increased unless he “improves English [skills].” The students who repeated taking the TOEFL iBT to “refresh” their best personal test score also agreed that the test preparation practices could not help increase test scores immediately. Rachel (CU-01), who was admitted into a Canadian university with the TOEFL iBT score of 104, experienced this improvement over time first hand. She took the TOEFL iBT three times. On her first attempt, she took the test immediately after completing a test preparation course; however, she only scored 86. Over the following six months, she took the TOEFL iBT twice more and achieved her best score.

**Perceived Effects on the CET4 Test Scores**

The relationship between the CET4 test preparation practices and students’ CET4 test scores was examined using factor analysis, correlation, and multiple regressions on the data collected from 534 students. Factor analysis was conducted to examine the underlying factor structure of CET4 test preparation practices. The correlation and multiple regressions were conducted to examine if there were significant relationships between these practices and test scores. The results are presented below, sequentially: (1) factor analyses, (2) correlations, and (3) multiple regressions.

**Factor analyses.** All 36 items of the CET4 test preparation practices for each test section of the questionnaire (Appendix H) were analyzed together. In this section, I will report the steps that I used to perform factor analyses, and then continue with the results of factor analyses.

**Steps of factor analyses.** I used principal component analysis and direct oblimin because they are most commonly used in language testing research (Ockey, 2014). The result showed that 5 factors were extracted. However, the review of the fourth factor indicated that the items loaded
onto this factor were doubled loaded onto the third factor, suggesting the fourth factor was not a unique factor. I re-ran the factor analysis with fixed 4 factors using principal component analysis and direct oblimin. However, the similar situation occurred with this extracted 4-factor structure. The 6 items loaded onto the fourth factor with the coefficients larger than .30, were double loaded onto the first three factors with larger coefficients; and these double loaded items indicated the fourth factor was not a unique factor. Therefore, I re-ran the factor analysis with fixed 3 factors using principal component analysis and direct oblimin. In this 3-factor structure, however, the third factor was not a unique factor because the 6 loaded items on this factor were double loaded onto the second factor.

Due to the issues in the previous steps of factor analysis using direct oblimin, the varimax rotation method was used. I re-ran the factor analysis with the fixed numbers of factors (4 factors) because the practices prepared the students for the four sub-skills that CET4 measures, including listening skill, reading skill, writing skill, and integrative skill (Xie, 2010). Two issues were found with this four-factor structure. First, there were two problematic items in the factor loading matrix. These two items were test preparation practices for listening skills, and they were doubled loaded onto the factor of test preparation practices for reading skills. I reviewed these two items and found they were related to general skills rather than particular to listening or reading skills. Second, a few items of test preparation practices for reading skill were split into two factors. Therefore, I removed these two items and re-ran the factor analysis with fixed 4 factors using principal component analysis and varimax. The items for listening skills and reading skills were loaded neatly onto their respective factors. However, most of items for writing skills were loaded with the items for integrative skills, with a few loaded as a separate factor. This suggested a 3-factor structure. I then ran the factor analysis with fixed 3 factors using
principal component analysis and varimax. In this factor structure, the items of test preparation practices for listening skills were loaded as one factor. The items of test preparation practices for writing skills and integrative skills were neatly loaded as one factor. The items of test preparation practices for reading skills were loaded as one factor, with one exception of one item loaded onto the factor of test preparation practices for listening skill. I then reviewed this item and found that, the practice of this item was not particularly for reading skills; rather it was transferable to other skills as well. Therefore, I removed this item and re-ran the factor analysis with fixed 3 factors using principal component analysis and varimax. The 3 extracted factors had clearly and neatly loaded items, and this 3-factor structure will be reported in the following section. I summarize the steps of these factor analyses in Table 16.

Table 16

*Steps of Factor Analyses*

<table>
<thead>
<tr>
<th>Step</th>
<th>Method</th>
<th>Actions</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Principal component analysis (PCA) + Oblimin</td>
<td>Factor analysis (FA)</td>
<td>Two items loaded onto the 4th factor were double loaded</td>
</tr>
<tr>
<td>2</td>
<td>PCA + Oblimin</td>
<td>FA with fixed 4 factors</td>
<td>Six items loaded onto the 4th factor were double loaded</td>
</tr>
<tr>
<td>3</td>
<td>PCA + Oblimin</td>
<td>FA with fixed 3 factors</td>
<td>Six items loaded on the 3rd factor were double loaded</td>
</tr>
<tr>
<td>4</td>
<td>PCA + Varimax</td>
<td>Re-ran FA with fixed 4 factors</td>
<td>Two items of test preparation practices for listening skills were double loaded</td>
</tr>
<tr>
<td>5</td>
<td>PCA + Varimax</td>
<td>Remove two items, re-ran FA with fixed 4 factors</td>
<td>Items of test preparation practices for writing skills loaded onto two factors</td>
</tr>
<tr>
<td>6</td>
<td>PCA + Varimax</td>
<td>Re-ran FA with fixed 3 factors</td>
<td>One item of test preparation practice for reading skills loaded with the factor of test preparation practices for listening skills</td>
</tr>
<tr>
<td>7</td>
<td>PCA + Varimax</td>
<td>Remove one item, re-ran FA with fixed 3 factors</td>
<td>3 factors extracted</td>
</tr>
</tbody>
</table>
Results of factor analysis. For the factor analysis with 3 neat factors, the Kaiser-Meyer-Olkin value was .960 and the Bartlett’s Test of Sphericity reached statistical significance.

Table 17

Factor Loadings of the CET4 Test Preparation Practices (n=534)

<table>
<thead>
<tr>
<th></th>
<th>Rotated Component Coefficients</th>
<th>Communalities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Integrative/writing practices</td>
<td>Reading practices</td>
</tr>
<tr>
<td>16a-listnpract1</td>
<td>.25</td>
<td>.25</td>
</tr>
<tr>
<td>16b-listnpract2</td>
<td>.22</td>
<td>.15</td>
</tr>
<tr>
<td>16c-listnpract3</td>
<td>.20</td>
<td>.24</td>
</tr>
<tr>
<td>16d-listnpract4</td>
<td>.31</td>
<td>.38</td>
</tr>
<tr>
<td>16e-listnpract5</td>
<td>.35</td>
<td>.41</td>
</tr>
<tr>
<td>16g-listnpract7</td>
<td>.21</td>
<td>.22</td>
</tr>
<tr>
<td>17a-readpract1</td>
<td>.14</td>
<td>.59</td>
</tr>
<tr>
<td>17b-readpract2</td>
<td>.29</td>
<td>.71</td>
</tr>
<tr>
<td>17c-readpract3</td>
<td>.34</td>
<td>.60</td>
</tr>
<tr>
<td>17e-readpract5</td>
<td>.34</td>
<td>.61</td>
</tr>
<tr>
<td>17f-readpract6</td>
<td>.22</td>
<td>.56</td>
</tr>
<tr>
<td>17g-readpract7</td>
<td>.36</td>
<td>.76</td>
</tr>
<tr>
<td>17h-readpract8</td>
<td>.27</td>
<td>.73</td>
</tr>
<tr>
<td>17i-readpract9</td>
<td>.38</td>
<td>.64</td>
</tr>
<tr>
<td>18a-wripact1</td>
<td>.57</td>
<td>.12</td>
</tr>
<tr>
<td>18b-wripact2</td>
<td>.68</td>
<td>.24</td>
</tr>
<tr>
<td>18c-wripact3</td>
<td>.65</td>
<td>.03</td>
</tr>
<tr>
<td>18d-wripact4</td>
<td>.63</td>
<td>.12</td>
</tr>
<tr>
<td>18e-wripact5</td>
<td>.54</td>
<td>.17</td>
</tr>
<tr>
<td>18f-wripact6</td>
<td>.69</td>
<td>.22</td>
</tr>
<tr>
<td>18g-wripact7</td>
<td>.63</td>
<td>.23</td>
</tr>
<tr>
<td>18h-wripact8</td>
<td>.63</td>
<td>.19</td>
</tr>
<tr>
<td>18i-wripact9</td>
<td>.69</td>
<td>.43</td>
</tr>
<tr>
<td>18j-wripact10</td>
<td>.65</td>
<td>.34</td>
</tr>
<tr>
<td>19a-translpract1</td>
<td>.63</td>
<td>.35</td>
</tr>
<tr>
<td>19b-translpract2</td>
<td>.60</td>
<td>.37</td>
</tr>
<tr>
<td>19c-translpract3</td>
<td>.70</td>
<td>.50</td>
</tr>
<tr>
<td>19d-translpract4</td>
<td>.72</td>
<td>.45</td>
</tr>
<tr>
<td>20a-clzpract1</td>
<td>.64</td>
<td>.34</td>
</tr>
<tr>
<td>20b-clzpract2</td>
<td>.71</td>
<td>.41</td>
</tr>
<tr>
<td>20c-clzpract3</td>
<td>.67</td>
<td>.34</td>
</tr>
<tr>
<td>20d-clzpract4</td>
<td>.53</td>
<td>.27</td>
</tr>
<tr>
<td>20e-clzpract5</td>
<td>.61</td>
<td>.50</td>
</tr>
</tbody>
</table>
The three factors represent the sub-skills that CET4 measures, as shown in Table 17. There were 19 items extracted as the first factor, including 10 items from 18a to 18j, 4 items from 19a to 19d, and 5 items from 20a to 20e. This factor accounted for 49.97% of the total variance. This factor matched with two intended measurement constructs of CET4 and also two sub-scores on the CET4 report card—the writing skill and the integrative skill. In CET4, translation and cloze are the two test sections that measure the same integrative skill (Xie, 2010). Therefore, the factor that was extracted from nine items—19a–19d of the test preparation practices for CET4 translation and 20a–20e of the test preparation practices for CET4 cloze—altogether represented the construct of test preparation practices for CET4 integrative skills. Because the items loaded onto this factor included the test preparation practices for writing skills and integrative skills, this factor is named *integrative/writing practices* in Table 17. The Cronbach’s alpha for this factor was 0.96.

The second factor was extracted from 8 items, including 17a, 17b, 17c, 17e, 17f, 17g, 17h, and 17i, and all the items were the test preparation practices for reading skills. This factor represented the reading skills that CET4 measures; therefore, this factor is named *reading practices*. This factor accounted for 5.44% of the total variance, and its Cronbach’s alpha was 0.91.

There were 6 items loaded onto the third factor, including 16a, 16b, 16c, 16d, 16e, and 16g. These items were the test preparation practices for the listening skills that CET4 measures; therefore, this factor is named as *listening practices*. This factor accounted for 4.56% of the total variance. The Cronbach’s alpha for this factor was 0.86.

The mean of all the items for each of the three factors was calculated and saved as the variables. These variables of factor mean scores were used in the follow-up analysis, including
correlation analyses and multiple regression analyses as the independent variables of test preparation practices for respective English skills.

**Correlations.** Correlation analyses were conducted between the dependent variables of the CET4 test scores and the independent variables of students’ background characteristics and CET-4 test preparation practices. These correlation analyses explored the relationship between a series of independent variables and the dependent variables of the CET4 test scores. There were eight independent variables: (1) three extracted factors of CET4 test preparation practices—listening practices, reading practices, integrative/writing practices (which used the factor mean scores of these factors as the variables), and (2) four students characteristics—gender, years of English learning, self-reported NMET\(^8\) score, and test repeater status. The five dependent variables were the CET4 test scores: the total score and the four sub-scores, including listening score, reading score, integrative score, and writing score. The guidelines for interpreting the magnitude of correlations were, .30 as small, .50 as medium, and .85 as large.

The results of the correlation between the independent variables and the dependent variables are presented in Table 18. As shown in this table, none of the three factors of the CET4 test preparation practices was significantly correlated with the CET4 scores, indicating that these test preparation practices had no significant relationship with CET4 test scores. On the other hand, some significant correlations were found between the student background characteristics of and the CET4 test scores. These characteristics were gender, years of English learning, self-reported NMET scores (university entrance score), and test repeater status.

---

\(^8\) The National Matriculation English Test (NMET) is the university entrance test of English that secondary school graduates who wish to gain entrance to Chinese universities and colleges take (for a comprehensive review, see Cheng & Qi, 2009).
Table 18

Correlations between Independent Variables and Dependent Variables (n=534)

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>Listening score</th>
<th>Reading score</th>
<th>Integrative score</th>
<th>Writing score</th>
<th>Total score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrative/Writing practices</td>
<td>.00</td>
<td>-0.03</td>
<td>-0.06</td>
<td>.00</td>
<td>-0.02</td>
</tr>
<tr>
<td>Reading practices</td>
<td>.02</td>
<td>.00</td>
<td>-0.03</td>
<td>.03</td>
<td>.06</td>
</tr>
<tr>
<td>Listening practices</td>
<td>-0.06</td>
<td>-0.04</td>
<td>-0.05</td>
<td>.01</td>
<td>.00</td>
</tr>
<tr>
<td>Gender</td>
<td>.10*</td>
<td>.14**</td>
<td>.11</td>
<td>.17**</td>
<td>.16**</td>
</tr>
<tr>
<td>Years of English learning</td>
<td>.20**</td>
<td>.09**</td>
<td>.07</td>
<td>.07</td>
<td>.14**</td>
</tr>
<tr>
<td>Self-reported NMET score</td>
<td>.51**</td>
<td>.57**</td>
<td>.33**</td>
<td>.55**</td>
<td>.62**</td>
</tr>
<tr>
<td>Status of test repeater</td>
<td>-.33**</td>
<td>-.39**</td>
<td>-.16**</td>
<td>-.39**</td>
<td>-.41**</td>
</tr>
</tbody>
</table>

Note: **p < .01, *p < .05

A general pattern was observed in the relationship between students’ background characteristics and the CET4 test scores. The CET4 total score and four sub-scores were significantly correlated with the background characteristic variables of self-reported NMET score, gender, and the status of test repeaters, indicating that male students scored significantly higher than female students, and test repeaters had significant lower scores than non-repeaters.

The test performance of the CET4 test repeaters was interesting to note compared to the TOEFL iBT test repeaters, who reported score increases after multiple attempts. It needed addressing some identified characteristics of the TOEFL iBT repeaters. First, according to the TOEFL iBT repeaters, many of them repeated the test more than twice until they achieved test scores “good” for academic program applications, and they spent months to complete these multiple test attempts. Second, the findings from the interviews and the classroom observations showed that the TOEFL iBT repeaters focused on improving English skills after failing on the test. The extended period of time for multiple test attempts and the re-directed focus, combined together, these characteristics associated with the TOEFL iBT repeaters might contribute to their increased test scores.
The highest correlation coefficients of these CET4 scores (sub-scores and total score) were with the self-reported NMET score—their English score upon admission, which ranged from .33 (with the sub-score of integrative skill) to .62 (with the total score). The variables, the self-reported NMET score, and the CET4 scores are all indicators of English language proficiency.

The listening score was significantly correlated with all students’ background characteristic variables, with coefficients for self-reported NMET score \( (r = .51) \), years of English learning \( (r = .20) \), and gender \( (r = .10) \) (male students performing better than female students). However, the correlation coefficient for the listening score was negatively correlated with the status of test repeater \( (r = -.33) \), indicating again that test repeaters scored significantly lower, this time on the skill of listening, compared to the students taking the CET4 for the first time. A similar pattern was observed across the correlations for CET4 reading score and CET4 total score with the same background characteristic variables.

The reading score was also significantly correlated with self-reported NMET score \( (r = .57) \), gender \( (r = .14) \) (that male students had higher scores than female students), and years of English learning \( (r = .09) \). The correlation coefficient of the reading score with the status of test repeater was significantly negative \( (r = -.39) \), indicating that test repeaters also scored lower on the skill of reading. The CET4 total score was again significantly correlated with self-reported NMET score \( (r = .62) \), gender \( (r = .16) \), and years of English learning \( (r = .14) \). Similarly, the correlation coefficient for the CET4 total score was slightly moderate negative \( (r = -.41) \), showing the lower total scores of test repeaters.

The CET4 integrative and CET4 writing scores shared the same correlation pattern. These two independent variables were significantly correlated with three background characteristic variables: gender, self-reported NMET score, and test repeater status. The correlation coefficient
of the CET4 integrative score with the self-reported NMET score was small \((r = .33)\), and that of the CET4 writing with the self-reported NMET score was medium \((r = .55)\). Gender was slightly correlated with the integrative score \((r = .11)\) and the writing score \((r = .17)\), with male students performing better than female students. However, the negative correlation coefficients of test repeater status with the integrative score \((r = -.16)\) and with the writing score \((r = -.39)\) indicated that, once again, test repeaters did not perform as well on these two skills as non-repeaters, which was the consistent pattern of the relationship across all the CET sub-scores.

**Multiple regressions.** Five individual multiple regression analyses were conducted using the CET4 total and four sub-scores as the dependent variables. The CET4 test scores of all 534 participants were included in the regression analyses. For the multiple regression analysis with the CET4 total score, the stepwise method was used and three sets of independent variables were entered into regression: (1) students’ background, including gender, years of English learning, self-reported NMET score, and test repeater status; (2) characteristics of test preparation courses, including type and timing; and (3) test preparation practices for the CET4 (3 factors of test preparation practices).

The independent variables of the students’ characteristics were entered as the first block to control for the contribution these backgrounds factors might make to students’ test scores. The independent variables of test preparation course characteristics and of test preparation practices were entered as the second and third blocks to examine what, if any, unique significant contribution these variables could make to CET4 test scores. For the other four individual multiple regression analyses, each of the CET4 sub-scores was used as the dependent variable, while the independent variables were the students’ background characteristics, test preparation
course characteristics, and test preparation practices. Table 19 presents the results of these regression analyses.

Table 19

*Summary of Five Multiple Regression Analyses (n=534)*

<table>
<thead>
<tr>
<th>Step</th>
<th>Independent variables</th>
<th>Final β of Dependent variables</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Total score</td>
<td>Reading score</td>
<td>Integrative score</td>
<td>Writing score</td>
<td>Listening score</td>
</tr>
<tr>
<td>1</td>
<td>Gender</td>
<td>.07</td>
<td>.08</td>
<td>.10</td>
<td>.04</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>Years of English learning</td>
<td>.10*</td>
<td>.09</td>
<td>.06</td>
<td>.02</td>
<td>.13**</td>
</tr>
<tr>
<td></td>
<td>Status of test repeaters</td>
<td>-.13*</td>
<td>-.15**</td>
<td>.05</td>
<td>-.13*</td>
<td>-.10**</td>
</tr>
<tr>
<td></td>
<td>Self-reported NMET score</td>
<td>.55**</td>
<td>.49**</td>
<td>.31**</td>
<td>.46**</td>
<td>.47**</td>
</tr>
<tr>
<td>2</td>
<td>Course types</td>
<td>.09</td>
<td>.10</td>
<td>.07</td>
<td>.06</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>Course timing</td>
<td>-.01</td>
<td>.00</td>
<td>.04</td>
<td>-.08</td>
<td>.01</td>
</tr>
<tr>
<td>3</td>
<td>Listening practices</td>
<td>.01</td>
<td>-.02</td>
<td>-.02</td>
<td>.06</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>Reading practices</td>
<td>.07</td>
<td>-.01</td>
<td>-.03</td>
<td>.07</td>
<td>.15</td>
</tr>
<tr>
<td></td>
<td>Integrative/writing practices</td>
<td>-.07</td>
<td>-.08</td>
<td>-.09</td>
<td>-.15</td>
<td>-.11</td>
</tr>
<tr>
<td></td>
<td>Adjusted R²</td>
<td>.45</td>
<td>.38</td>
<td>.09</td>
<td>.33</td>
<td>.32</td>
</tr>
</tbody>
</table>

*Note: *p < .05, **p < .01*

As shown in Table 19, the variables of the students’ background characteristics were significant predictors of CET4 scores, while none of the variables of test preparation courses characteristics and test preparation practices significantly predicted CET4 scores. The model of with these three blocks of variables explained 45% of the total variance of the CET4 total score, 38% of the reading score, 9% of the integrative score, 32% of the listening score, and 37% of the writing score.

Looking across the coefficients of the variables of students’ background characteristics, it was found that the students’ self-reported NMET scores were the significant predictor with the greatest positive β values, indicating that the previous English test score variable was the best predictor of students’ CET4 test scores among all the background characteristics. This variable was also the only background variable that could significantly predict both the CET4 total and
the four sub-scores. It should be noted that both variables, the independent variable of the NMET scores and the dependent variable of the CET4 total and sub-scores are the scores of English language tests, and thus both indicate the students’ English language proficiency; therefore, it is not surprising that the variable of NMET scores was the significant predictor of the variable of the CET4 scores.

The number of years of learning English was also revealed to be significant predictors of both CET4 total and listening scores, but with small coefficients of .10 and .15, respectively. The variable test-repeater status was the significant predictor of all CET4 scores except for the integrative score. However, the negative $\beta$ value indicated a negative significant correlation between this variable and the students’ CET4 test scores. This result suggested that test-repeaters scored significantly lower than first time test takers (non-repeaters) on the CET4 total score and the sub-scores of listening, reading, and writing. Therefore, the variable of test-repeater status could also be a significant predictor of some domains of CET4 test performance, but in a negative direction. This result was consistent with the previous correlation results. The status of being a test-repeater was significantly correlated with all domains of CET4 scores with negative coefficients; therefore, it was no surprising to note this variable had a negative $\beta$ value in this regression model.

None of the variables of test preparation course characteristics—e.g., types of courses or timing of taking courses—was a significant predictor of the students’ CET4 test performance. This result indicated that there was no significant correlation between type and timing of test preparation courses with students’ CET4 test scores. Put simply, no matter what type of course these students took, or what timeframe they chose, there were no statistically significant relationship between these choices and students’ test performance.
Value of the Test Preparation Practices

Three themes (Table 20) emerged from the primary data sources of the interviews with 12 students currently studying at a Canadian university and the supplementary data sources of the interviews with six students at Centre L (18 students altogether). Among the 12 students at a Canadian university, ten took TOEFL iBT test preparation courses at branches of the same national chain training centre, and two chose their local test preparation centres. The six students chose the TOEFL iBT courses at Centre L. The data from these students reflect their perceived value of the test preparation practices the students experienced: trustworthiness—professional teachers, and structured and reliable instruction; efficacy—investment in time and efforts, and necessary knowledge of the test and English skill improvement; and reorientation—learning communities, and sources of motivation. Each theme is elaborated upon in detail in the following.

Table 20

| Thematic Findings of Students’ Perceived Value of the Test Preparation Practices |
|---------------------------------|---------------------------------|
| **Trustworthiness**             | **Efficacy**                    |
| Code 1 “Professional” teachers  | Code 3 Investment in time and   |
| Code 2 Structured and reliable  | efforts                         |
| instruction                      | Code 4 Necessary knowledge of   |
|                                 | the test and English skill      |
|                                 | improvement                      |
| **Reorientation**               | Code 5 Learning communities     |
| Code 6 Motivation reorientation |                                 |

**Trustworthiness**

The students found the test preparation practices to be trustworthy, evaluated by the following two aspects: the “professional teachers” and the “structured and reliable instruction.”

“Professional teachers.” Many students used the word “professional” to describe their teachers because the teachers used their own first-hand educational and test-taking experience to
teach what the test is about and what it is used for. The teachers who most impressed the students shared one common characteristic—overseas educational background, giving them first-hand experience studying in the academic contexts the students hoped to attend themselves. These teachers shared their personal learning experiences to emphasize the importance of strong TOEFL iBT scores for entrance into academic programs and proficient English skills for the successful completion of these programs. Some teachers’ first-hand TOEFL iBT test-taking experience was valued because these teachers were able to provide many practical suggestions for managing various test-taking situations. Jack (CU-09) commented on a suggestion his teacher gave to deal with a possible situation in the TOEFL iBT speaking section:

The speaking section requires both listening and recording of your answers. The speaking [module] my teacher mentions, if the test taker beside you started the speaking [section] earlier, there was the risk that you might listen to the lecture while s/he had started talking. This will definitely interfere with your listening. So he suggested saving some time in the reading section by skipping the screens of test directions. Even though the saved time could not help start the speaking section much earlier, you might not be so left behind [to start speaking section]. This is a very important detail from my teacher’s real-time test taking.

Such first-hand personal experience, with respect to both education and test-taking, became these teachers’ greatest advantage in the classroom, and the experiential suggestions they made were attended to by their students. As Clark (CU-04), a PhD student, explained, “these teachers, who have their own experiences in real [English academic] situations, know what is needed; and compared to the teachers who strictly follow textbooks, I like and trust the teachers with experiential pedagogy.” However, it was noted that the availability of teachers with overseas educational background or TOEFL iBT test-taking experience was limited, even though such experience was definitely considered as asset when applying for the position. The test preparation centres with the teachers who had already demonstrated success at accomplishing the goals the students had for themselves was critical.
**Structured and reliable instruction.** All the students reported experiencing a similar course structure—the four modules of reading, listening, speaking, and writing that aligned with the TOEFL iBT test sections on the same. The students stressed that this structure provided “a starting position to assign and break down the entire test preparation plan into four major parts”, which they found especially helpful when facing limited time to prepare for the test preparation.

It was noted that many students (10 out of 18) chose branches of the same national chain training centre. The reason for making this choice was that this centre was perceived as being “professional” and “authoritative” in terms of its instructional content. The sources of the instructional content, according to the students’ accounts and their documents of in-class exercises, were mostly from the official website of the Educational Testing Service (ETS), the Official Guide to the TOEFL iBT, or the Test Practice Online (TPO) published by the ETS. To these students, the direct mapping of instruction onto the guides and practice tests was seen as the “professional” way to help them do well on the test. Justin (CU-07), a Master’s student, commented,

> Compared to the segmented pieces of information you put together to map out the picture of the TOEFL iBT, the instructional content and information presented to you by this educational institution is more reliable, and you tend to have more trust, … because they are professionals [in test preparation industry].

The instructional content these students described included: sequential illustrations of test features (e.g., test formats, scoring criteria, task content/topics, designated time); demonstrations of test tasks and test-taking strategies; management of affect (e.g., anxiety, confidence); strategies for developing English skills; and/or tips on the test-taking process.

While the content structure was similar across centres, he coverage and/or percentage of each aspects of the instructional content varied across teachers. For example, some teachers explained the importance of content knowledge for comprehending reading texts, and covered
some discussion of social values (e.g., gender equity, environment awareness, etc.) in their instruction. Some students thought such discussion and understanding helped them to answer relevant test items. However, other students showed negative attitudes towards such instructional content. Maggie (CU-11), a Master’s student, recalled when the teacher of a reading module extended a discussion from a reading passage about evolution to the presidential elections in the United States, and she had no idea how this discussion was related to test preparation; she thought the discussion was “totally waste of time.” This student’s perception that the in-class discussion was out of the instructional focus reflected the point that the teachers at Centre L emphasized with respect to the episodes as in-class breaks. The different extent that the discussions or episodes of this kind used from teacher to teacher contributed to students’ different perceptions of the value of the instruction, because the students might feel the diluted instructional focus or the adjusted instructional pace with the focused instruction.

**Efficacy**

The data revealed that the students judged their test preparation practices in terms of costs and benefits. As such, they valued the efficacy of these practices from two perspectives: (1) the efficient investment in time and effort, and (2) the necessary knowledge of the test and of English skills improvement.

**Investments in time and effort.** All the students indicated that they would still take test preparation courses if they were given one more chance to prepare for the TOEFL iBT because this was the most time efficient way to prepare for the test. They believed that, if they hadn’t taken preparation courses, they would have either taken much more time to familiarize themselves with the test or taken the test more times to achieve a satisfactory score. The time cost equation was an important consideration, which they balanced with the whole process of
academic applications; taking the TOEFL iBT was only one among many requirements. Clark (CU-04) clearly stated the importance he placed on efficacy, with regards to time investment:

I did a cost calculation. If I had not taken the TOEFL iBT test preparation course and if I had not got a satisfactory score, I would have probably had to wait till next year to apply [for academic program]; and the cost would be far more than the time and money I would spend now. In this case, I think it [taking test preparation courses] is worthwhile.

Some students commented on the effort investment, stating that they thought that the structured instruction was helpful for regulating their efforts. These students commented that lack of self-control was a big challenge when preparing for the test on their own because, without the help of test preparation courses, they might be disorganized and inefficient in terms of time management. Mary (CU-05), a PhD student, explained: “I did not want to look at it [the TOEFL iBT] at all if nobody guides me how to.” She found the intensive course very useful because she knew “what aspects [of test preparation practices] to direct efforts onto” by following the teachers’ structured instruction.

**Necessary knowledge.** Many students valued the necessary knowledge that their teachers presented about the TOEFL iBT and English skills improvement. The teachers and the administrators at Centre L repeatedly emphasized that high scores would not be within arm’s reach; they directed the students’ attention onto English skills because they understood that English skills were what the TOEFL iBT intended to measure. Many students admitted that knowledge of “what is tested by the TOEFL iBT” was very necessary for their test preparation, especially when they were immersed in cases of using test-taking strategies to achieve high scores.

Many students recalled their teachers explaining how test scores were related to English skills rather than test-taking strategies. Jenny (CU-08), an undergraduate student, remembered
her teacher giving detailed instructions on a scored writing sample in the Official Guide and explaining why this sample was given a high score by mapping it onto the scoring criteria. She said,

The [teacher’s] instruction made me understand that the clear logic throughout this writing sample ensures the coherence of the ideas presented, and one portion of the high score was awarded to its clear logic; and the high score is definitely associated with the skills presented in this writing.

Some students further added that they attended to the instruction on English skills after understanding what was tested by the TOEFL iBT, and their teachers demonstrated strategies in class on how to improve English skills, particularly through practicing test tasks. For example, Mary (CU-05) recalled her teacher demonstrating using a flow chart for a listening task. The teacher took note of the logical line of a lecture (one listening task) and the related details, and organized the notes using a flow chart. This strategy was quite different from the strategy that Mary had used previously—to note down everything she could catch in the listening task. Due to the limited time, Mary only tried this strategy once in class; however, she repeatedly practiced it on different listening tasks after class, and noticed that this strategy “worked well” for listening tasks and also “helped improve listening skills.”

Similarly, many students planned a period of time (e.g., 2–3 months or longer) after completing test preparation courses as a reinforcement period; improving their English skills was a common focus for reinforcement before they took the TOEFL iBT. Rachel (CU-01) increased her TOEFL iBT scores from 86 to 104 in six months, and she attributed the score gains to her continuous practice of her English skills and the correct understanding of the test that her teachers presented.

However, because of individual characteristics and contextual constraints, such necessary knowledge of the TOEFL iBT and English skill improvement did not reach every student. Some
students (4 out of 18) expressed that they “only needed a score”, so they focused on test familiarization and did not have the expectation of improving their English skills. Given this instrumental motivation, these students attended to the instruction related to the test (e.g., test formats). Some met many teachers who believed test preparation was test familiarization. When such individual characteristics (e.g., narrowed attitude toward learning, teaching beliefs) were confounded with contextual constraints (e.g., limited timeline, limited budget), it was unlikely that the students would include English skill improvement in their test preparation practices. As Justin (CU-07) commented, “I had a very tight schedule [when preparing for the TOEFL iBT], and I did not want to improve my English proficiency through taking this [test preparation] course; I just needed a score [emphasis in original tone], and this was most important to me.”

Reorientation

Reorientation included the learning communities where students gained access to from their experience of taking test preparation courses, and motivation reorientation where the students gained strength to persist.

Learning communities. One advantage of test preparation courses was the opportunity to meet peers who were also involved in the application process. Shared goals and experiences enabled discussion and the exchange of ideas among peers; therefore, communities often formed—led by either students or teachers—and served as platforms to facilitate discussions. Some communities were virtual, such as online chat groups or groups in online forums, and others were face-to-face, such as lunch-mate groups. For example, Mary recalled that she and five other girls in the same intensive course soon became a group after taking lunch break together everyday. She felt “better” to be in this group because she could talk through the
difficulties and stress she encountered during the test preparation process, and knew that she was not “the only one who had experienced those difficulties.”

However, some students admitted that the possibility of forming communities depended on interactions with peers in class, and not all the students in this study reported the emergence of, or participation in learning communities in their accounts. For example, Lindsay (CL-S1), a student at Centre L, mentioned that only after being assigned to a group for performing an oral task did she become familiar with the students sitting close to her, who then started an online chat group. In addition, some students mentioned that their teachers used social media platforms to form communities, such as blogs, Weibo (a Chinese version of Twitter), and public account subscriptions on WeChat (a mobile communication application similar to WhatsApp).

Participation in learning communities mainly included the sharing of information and resources among peers and teachers. The students commented that they valued the resources shared in these communities for two reasons: (1) they could access useful resources of which they had not been previously aware, and (2) the shared resources reviewed and recommended by peers and teachers offered a helpful starting point, given the overwhelming range of resources available on the market. Wilson (CU-02), for instance, who practiced listening skills with *60-second Science by Scientific American*, said he came to know this resource from a post by *Ji-jing* shared by his peers in an online chat group. He clearly remembered that this *Ji-jing* was a detailed reflection of a test taker’s experience of preparing for the TOEFL iBT, and the podcast of *60-second Science by Scientific American* was the resource this test taker used. Wilson appreciated this shared resource because he thought it helped him improve his listening skills.

The learning communities that teachers initiated (e.g., an online group) served as an extension of class, and many students said they read posts related to the test and English
language learning. Clark recalled his writing teacher posting the transcripts of a speech Obama delivered in Russia on her blog, and highlighting some sentences in blue to draw students’ attention to the sentence structure and vocabulary. Clark particularly mentioned that he had regular interaction with his teachers on social media platforms, including this writing teacher, and he admitted that the rapport he built with them through the interactions increased his trust in both them and in their instruction.

Motivation reorientation. Many students recalled their test preparation courses as being an experience of gaining inspiration, encouragement, and mental support from peers and teachers. For them, attending test preparation was more like a ritual to equip themselves with “spiritual powers”, in William’s (CU-12) words, and this kind of motivational reorientation became an anticipated and expected ‘right of passage’ to be experienced by those wishing to study abroad. Classes in these test preparation courses bring together groups of youth with similar goals who work hard together to achieve them. Commenting on the class as a platform for encouragement, Clark noted:

I came to the class before 8 a.m. and found that four or five students had been there and started reciting vocabulary. Then I immediately knew what I should do. [...] I also found that my English is not that good compared to many students in class. Those whose English were far more better worked really hard, how could I waste time? This atmosphere was the source of my motivation.

Students reported that many of their teachers shared their own successful experiences or those of their previous students. The students found these narratives to be very inspirational because they could see themselves in those who succeeded. Knowing the outcomes achieved through the efforts others made and how they succeeded encouraged the students to work harder. William recalled,

Those successful stories my teachers told mentally supported me to insist on sitting in a classroom with 300 students in hot summer days; [...] it sounded tough but I did not feel so
because it was where I got inspired. [...] The mental power from these successes also conveyed a message of persistence to me, which I think is so important not only for taking the TOEFL iBT but also for each step of the academic application process.

Chapter Summary

This chapter reported the results of the data collected from multiple stages of Chinese students’ test preparation experiences (i.e., during taking test preparation courses, right after completing test preparation courses, and after achieving successful test scores), from multiple participants (i.e., students, teachers, and administrators), and using multiple methods (i.e., documents, classroom observations, interviews and questionnaires). The results were organized according to the nature of the test preparation practices, the perceived effects of these practices, and the value of these practices.

The major findings with regards to the nature of the test preparation practices revealed that students chose to take test preparation courses because they associated a high level of importance of the test scores to the test uses (e.g., language requirement in academic application for English-speaking universities, job applications). The participants’ characteristics, particularly students’ previous test-taking experience and the teachers’ and administrators’ educational and personal backgrounds, were closely related to why the different participants engaged in test preparation practices as well as how students chose test preparation courses and teachers taught those courses. The students who had previous unsuccessful attempts on the tests focused their test preparation practices on English skill improvement and also planned more time for preparing the tests. The teachers and the administrators with overseas educational experience and successful English learning experiences communicated with the students the appropriate strategies to achieve high test scores. Three test preparation practices were observed: test familiarization, English skill improvement, and stress management; these participants’ characteristics, their
different motivations, and the various choices they made with regards to test preparation courses (e.g., types, and timing), all contributed to shaping these practices.

The test preparation practices were found not to have an immediate effect on increasing test scores. The students taking the TOEFL iBT test preparation either thought their test scores would not increase if they took the TOEFL iBT immediately after completing their test preparation courses, or reported a similar score to that of previous test attempts. The CET4 test scores of students recently completing the CET4 test preparation courses were examined in relation to their reported test preparation practices, and no significant relationships were found. The students valued their teachers’ professionalism, the efficacy, and the reorientation that the test preparation practices provided. The structured and reliable instruction that their teachers gave enabled the students to efficiently prepare for the tests, to understand the necessary knowledge about the tests, and to focus on English skill improvement. Additionally, the test preparation practices allowed the students to participate in learning communities and also to gain motivational strength to achieve success.
Chapter 5 Discussion

This chapter discusses the major research findings of this study, and addresses their implications. I begin by discussing the findings for each of the three research questions by connecting to the existing literature and the Chinese context. Next, I revisit the purpose of this study by discussing the major findings through the lens of Green’s (2007) washback model to contribute to understanding the phenomenon of Chinese students’ test preparation for high-stakes English language tests. Finally, I conclude by discussing the implications of this study, its limitations, and recommendations for future research.

Discussion of the Three Research Questions

This section is organized into three sub-sections, one for each research question. I discuss the results for each research question in relation to the existing literature to understand how these findings contribute to the test preparation literature. I further focus on discussing these findings in the context of test preparation for high-stakes English tests in China to understand the specificity of each of the three aspect of the test preparation practices: nature, effects, and value.

Research Question 1: Nature of the Test Preparation Practices

The first research question asked, “What is the nature of test preparation practices employed to help Chinese students achieve success in the two tests?” The literature has suggested that among the three major test preparation practices that were identified in this study, test familiarization was the most common practice. Teachers build their instruction around test features and test demands to help students become familiar with the target test (e.g., Alderson & Hamp-Lyons, 1996; Chau, 2008; Green, 2007; Wall & Dorák, 2006); students repeat drills to become familiar with the requirements and necessary strategies to complete the test items and tasks (e.g., Doe & Fox, 2012; Liu, 2014; Powers, 1985; Powers & Swinton, 1982; Xie, 2010).
Previous studies also reported that test preparation teachers included English skill improvement in their instruction (e.g., Gibson & Swan, 2008; Green, 2007; Wall & Dorák, 2006; Yu, 2012). My research similarly revealed the practice of including English skills improvement in the test preparation courses, and my findings provided detailed evidences about the factors (i.e., participants’ characteristics, perceptions, and the contextual influence) related to the practice of English skills improvement. These factors contribute to a deeper understanding of the conditions and actions that may contribute to the quality of test preparation experiences. Furthermore, although many studies in the reviewed literature report that maximizing motivation and minimizing anxiety are the most common reasons for engaging in test preparation practices (e.g., Cole, 1982; Smith, 1990), stress management was not often identified as a consideration. In previous studies, reference was made only to students’ expectations that taking test preparation courses would reduce test anxiety and boost their confidence (e.g., Green, 2007; Yu, 2012). My findings revealed that Centre L used stress management as one particular test preparation practice to help students achieve success.

The three major test preparation practices identified by the findings—test familiarization, English skills improvement, and stress management—are best understood by considering participants’ characteristics, their perceptions, and the context that influenced their choices. All of these aspects combined, including participants’ characteristics, perceptions, and contextual characteristics of test preparation centres, provide unique insights into test preparation practices in the Chinese context. I will discuss each of these aspects in the following section.

**Students’ Characteristics.** Previous studies in cultural contexts other than China reported that, among the students enrolled in coaching programs or test preparation courses, some had low English language proficiency and/or limited English learning experience, and formal English
language instruction was not typically part of either their academic preparation or extracurricular experiences (e.g., Cole, 1982; Ferman, 2004; Gan, 2009). My study of students in China found that those with regular access to English instruction and with sufficient experience of English learning still chose to take test preparation courses. These findings illustrated the unique characteristics of Chinese students in test preparation courses.

First, the research data revealed that students chose test preparation courses even though they had regular English courses at their academic programs on campus. The Chinese students in this study, in particular the students at Site 1 (Centre L) and Site 2 (eight universities and colleges), were in undergraduate or graduate programs when they chose to take test preparation courses for the TOEFL iBT or the CET4, and therefore had access to regular English courses on campus. The finding of this particular characteristic further probed the question of why: i.e., why do these students need extra instruction in English language, in addition to the regular English courses already included in their programs and tuitions?

Second, my results showed that the students chose test preparation courses even though they had learned English for many years. All of the students at Site 1 (Centre L) and Site 3 (a Canadian university) with the TOEFL iBT test preparation experience had studied English for more than 10 years; among 534 students at Site 2 (eight universities and colleges in China), 26.6% (n=142) reported more than 10 years of English learning. This finding added uniqueness to the characteristics of Chinese students choosing to take test preparation courses, and may speak to an issue of instructional quality in the school systems for teaching English, and also raised an additional question of why these students still choose extra hours of instruction despite the many years they have already spent on English learning.
The learning style of Chinese students may answer the above questions. Hu (2002) extensively reviewed English language teaching and the Chinese culture of learning, and argued that Chinese students commonly used the four Rs learning strategy: reception, repetition, review, and reproduction. Chinese students tend to receive and retain the knowledge imparted by teachers, repeatedly study to acquire knowledge, review what they received and repeat it to consolidate and deepen knowledge, and accurately reproduce the knowledge on demand by teachers or tests. The Chinese students in this study, who had been immersed in this culture of learning and who had practiced such learning strategies at every stage of their learning, preferred to use these same learning strategies to prepare for the test. In addition, the instruction for the test preparation courses matched the four R learning strategies: the teachers passed the knowledge of the test and of English skills improvement to the students; drills were repeatedly practised so the students could reproduce what each specific test tasks requires, or what the teachers demand to improve their English skills; during this process, the students were able to review if and what they had acquired. Chinese students, who are accustomed to a specific set of learning strategies, take test preparation courses that match such learning strategies, and it is reasonable to understand that taking preparation courses is a way for them to review what will be tested before they actually take the test. As such, the students take test preparation courses as a mental preparation, and also follow these routinized strategies to achieve success on high-stake tests.

Previously reported characteristics of the students in coaching programs or test preparation courses include students without access to regular instruction on the subject to be tested, those students with low language proficiency level, and those who took the test for the first time (e.g., Cole, 1982; Ferman, 2004; Gan, 2009; Liu, 2014; Yu, 2012). Similarly, my findings also showed that the students who took the CET4 or the TOEFL iBT for the first time chose to take test
preparation courses. It should be noted, however, that the substantial importance associated with the test scores contributed to the students’ decisions on types and timing of taking test preparation courses. Therefore, for the Chinese students of this study—who rated test importance very high because of the consequences related to test score uses (e.g., the successful TOEFL iBT test scores for academic application)—their decision making could not be influenced by their previous test-taking experience only. Their understanding of test score use and the subsequent academic opportunities related to these scores also greatly contributed to their decision making.

It should also be noted that students with previous unsuccessful test attempts focused considerable attention on English skills improvement when they repeated the tests. One possible interpretation is that these students came to understand that if their English skills did not improve, they were not likely to see a substantial increase in their test scores; and also they realized that test familiarization and test-taking strategies alone would not be enough to improve their scores. The students who did their first test attempts paid more attention on test familiarization and test-taking strategies than on English skills improvement, because they needed such guidance to reduce the influence on their performance due to unfamiliarity and to ensure their English proficiency better demonstrated.

**Teachers’ characteristics.** In keeping with previous research (e.g., Gibson & Swan, 2008; Spratt, 2005; Watanabe, 2000), the findings of the study showed that the test preparation practices the teachers used were greatly influenced by factors related to their characteristics and perceptions. These factors included the teachers’ educational background (e.g., overseas study, major programs), their professional background (e.g., professional training, teaching experience), their understanding of the test demands, and their level of English language learning. As Spratt (2005) argued, these factors influenced how teachers teach test preparation classes, and that
subsequently influenced how (e.g., positive or negative) and to what degree (e.g., no, intense, or “heavy”) washback occurred.

One characteristic displayed by some teachers in this study deserves particular attention: the lack of a degree in English language or professional training in English language education. This characteristic was related to current status of the test preparation industry in China. The significant market share and business potential of the test preparation industry has resulted in a rapid growth in the number of test preparation centres as well as a variety of types and sizes of test preparation courses. With the market booming, a huge number of teachers is needed to fill the teaching positions. A group of “professional” teachers has emerged on the test preparation market. They are “professional” in that they treat teaching test preparation courses as a profession by which to make a living (He, 2010). Indeed, they can “make a good living,” as the teachers at Centre L noted during this study.

Since the test preparation industry is privatized, there is no administrative supervision of test preparation centres by the educational authorities. A professional certificate in English language education, therefore, is not a compulsory requirement when recruiting teachers, and the specific criteria for recruiting teachers vary across test preparation centres. At Centre L, a high level of English skills (e.g., high scores on the TOEFL iBT) and successful teaching experience (e.g., from previous students’ evaluations) were prioritized during the hiring process to balance the non-professional background in English language education. The understanding of the TOEFL iBT and of the English language learning these teachers shared with their students were thus based on these teachers’ own successful experiences. However, with thousands of test preparation centres in China, it remains questionable whether each centre can properly balance teaching qualifications and the number of vacancies for teaching test preparation courses.
**Contextual Characteristics of the Test Preparation Centre.** Spratt (2005) argued that factors related to the school, such as atmosphere, cultural factors, and the pressure on test results, greatly influenced teaching. Similarly, the contextual characteristics of Centre L identified by the interview data and the documents showed positive support for the teachers’ teaching practices. The administrators shared an understanding with the teachers regarding “what the students needed” and “what the students wanted.” They therefore worked together to provide “appropriate” guidance to establish reasonable expectations of test preparation courses and test score gains (e.g., public lectures) and to provide extra resources to meet these expectations (e.g., courses targeted at improving English skills); the atmosphere at the Centre was that of a shared understanding and a culture of close collaborations that supported the teachers in following their belief of “teaching students” rather “pleasing students.”

Although administrative demands for test results were reported in some studies as the factor pushing teachers to “teach to the test” (e.g., Deng & Carless, 2010; Firestone, Monfils, & Schorr, 2004), neither the administrators nor the teachers at Centre L reported experiencing such pressure. This finding can be attributed to the test preparation course context in China. First, as discussed earlier, there is no administrative supervision of test preparation centres by educational authorities. Thus there is no top-down push to achieve good test results. Second, the Centre pays its teachers by the workload (e.g., hours of instruction, number of modules, size of classes) rather than based on student test results. Teachers reported they did not feel pressured to produce test results. Such a pressure-free environment contributed positively to the teachers at Centre L, and allowed them to follow their own instruction.
Research Question 2: Perceived Effects of the Test Preparation Practices

The second research question posed in this study was: “What are the perceived effects of these test preparation practices on Chinese students’ scores on the two tests?” The findings of the TOEFL test score changes reported by the students at Site 1 (Centre L) and Site 3 (a Canadian university) and the CET4 test scores of the students at Site 2 (eight universities and colleges) did not show immediate substantial test score increases or the significant relationship between test preparation practices and students’ test scores. Becker (1990) has urged further studies examining the effects of coaching/test preparation programs to determine “what constitutes coaching”, in order to reach a clear understanding of the coaching effects on test scores. Messick (1982) made detailed classification of the effects of test preparation (coaching) according to its relation to score validity. The effects of test preparation were found to be beneficial to score validity when test preparation: (1) helps students employ test-taking strategies and manage test anxiety to reduce the construct-irrelevant variance that influence students’ demonstration of their ability; and (2) helps improve students’ abilities (Messick, 1982). In contrast, Messick found that when test preparation works in the way that students’ increased test scores are not equivalent to the improvement in the measured ability, the effects will threaten the score validity because the test scores are inaccurately higher than students’ actual ability.

Following Messick’s (1982) classification, this section starts with the discussion on how the perceived effects of test preparation practices are related to the enhancement of test scores and English language skills. The discussion is in relation to the nature of test preparation practices to reach a legitimate interpretation of the test preparation effects. This section then continues with a discussion on the importance of students’ background characteristics, including
gender, years of English language learning, NMET test scores, and status of test repeaters, to highlight how these characteristics influenced test performance.

**Test scores.** Both the classroom observation data of the TOEFL iBT test preparation and the questionnaire data of the CET4 showed that test familiarization was one important test preparation practice, which included reviewing test structure, test formats, task requirement, sometimes scoring/rating criteria, and demonstrations of test tasks and drills. Test familiarization helps students have a clear understanding of the test tasks, which, in turn, helps them to manage testing situations that might be unrelated to their English language proficiency yet still influence their test performance. Thus one impact from test preparation practices is that the students are familiar with the structure of the tests, and this would likely reduce construct irrelevant variance as it would reduce unfamiliarity of the test as a barrier to achievement. For example, some TOEFL iBT teachers gave suggestions on how to avoid conflict with other test takers in the integrated speaking tasks (e.g., listening the scripts while others are starting to speaking to the computer to record their answers) to reduce the interference (e.g., ambient noise) that might influence how they demonstrate their English language proficiency. If the effect of such test preparation practices, as Messick (1982) argued, could reduce variance that was irrelevant to the construct (e.g., English language proficiency), the accuracy of test scores could be enhanced.

Both the correlation analyses and the multiple regression analyses revealed no significant relationship between test preparation practices and CET4 test scores. Because all of the CET4 students in this study completed test preparation courses; therefore, the non-significant results indicated the lack of differential relationship across test preparation practices. The choices that these students made with regards to the timing of taking the test preparation courses could provide an explanation for understanding the non-significant relationship. Specifically, many
students crammed in their test preparation very close to the test date (e.g., less than a month before). With such limited time, the effects of the test preparation practices were difficult to judge or achieve. The findings of the interview data from the students at Site 3 (a Canadian university) showed that while substantial increases in TOEFL iBT scores were achieved after test preparation courses, this rise in test score took six months to happen.

**English language skills.** My findings showed that instruction for the test preparation courses included a focus on English skills. The questionnaire data revealed there were specific practices in test preparation courses to improve English language skills (such as understanding complex sentences as a strategy for improving reading skills). The classroom observation data of the TOEFL iBT test preparation classes further demonstrated both how and why some strategies were beneficial for improving the students’ English skills (e.g., the four-step intensive listening strategy, understanding the TOEFL iBT). The students interpreted the increase in their test scores (e.g., Rachel’s TOEFL iBT score change from 86 to 104) as resulting from an improvement in English skills, attributed in turn to the instructional focus on the intended construct of the test. The test preparation practices with such a focus, as Messick (1982) and Powers (2012) argued, were thus beneficial to the validity of test score interpretation and use.

However, the question of interest is: to what extent were the teachers devoted to improving specific English skills? The students gave different accounts of their teachers’ instruction, from the positive experience of “this method [the teacher introduced] changed my views on how I write [in English]” to the disappointing accounts of “I don’t know what he is talking about.” The range of student responses was related to the range of improvement of English skills, which, in turn, was related to the different effects on test scores. Liu (2014) examined the relationship between the Chinese students’ test preparation practices for the TOEFL iBT and the TOEFL iBT
test scores, and found that attending coaching schools was a significant predictor of test scores, but with a small effect size. Although no further data were available in Liu’s study, the variation in teachers’ instructional focus across different test preparation centres or coaching schools may explain the small effect, or the differential effects on the students’ test scores and English skills.

Students’ background characteristics. The results of correlation analyses and multiple regression analyses on the questionnaire data revealed the importance of students’ background characteristics to students’ test performance. These characteristics include gender, years of English learning, NMET test score (the score of an English language test for entrance to university), and status of test repeaters. Gender is not a significant predictor of students’ CET4 test scores despite the significant relationship identified in the correlation analyses. This result is similar to the finding of a recent study that examining the relationship between Chinese students’ motivation, test anxiety and their test performance (Cheng, Klinger, Fox, Doe, Jin, & Wu, 2014). Among a number of students’ characteristics, Cheng et al. found that gender was not a significant predictor of CET4 test scores.

The variable of years of English learning is also identified as a significant predictor in this study. It is interesting to note that this variable is significant predictor of CET4 total score and listening-skill score, but with small coefficients (.10 and .13 respectively). One possible explanation of the small coefficients is that there is not much variance in the variable of years of English learning. The majority of the CET4 students reported they have either more than 10 years of English learning (starting from kindergarten) or around 4-6 years (starting from secondary school or Grade 7). This finding reflects the situation of English language education in China and the difference in the English curriculum and syllabus across various regions of China. Students in large cities have various resources and access to English classes within the school.
system and in the English language training market. Comparably speaking, students in some remote areas have limited resources of English learning, as Henry (CL-T3) mentioned that he did not receive training in listening and speaking when he started English learning at the small town he grew up; subsequently, the students in the remote area probably start to learn English late (e.g., from secondary school) or receive limited trainings in English skills. The result that years of English learning is the significant predictor of CET4 listening-skill score clearly reflect the influence with respect to years of English learning on students’ test performance.

NMET test scores and status of test repeaters are identified as significant predictors across all domain of CET4 test performance (i.e., total score and four sub-skill scores). NMET is an English language test that high-school students in China take in order to enter into universities and colleges (Mei, 2015). The positive relationship of the NMET test scores with the CET4 test scores, which is shown as the positive coefficients, indicate test scores on both tests reflect students’ English language proficiency. Students who perform well on the NMET have good English skills, and they can have good performance on the CET4 as well. The findings showed that test repeaters scored significantly lower on all domains of CET4. This is because the low English proficiency test repeaters had, which was supported by their low NMET scores. These two background characteristics both indicate students’ English proficiency; therefore, they became the most powerful predictors of students’ CET4 test performance.

**Research Question 3: Value of the Test Preparation Practices**

This study’s third research question asked: “How do Chinese students perceive the value of these practices with respect to helping them achieve success?” The findings clearly demonstrated that students valued the structured and reliable instruction that students had the confidence to accept, the efficiency of familiarizing the test and improving English skills, and also the
motivation to strive throughout the stressful test preparation experience. However, the students interviewed gave various accounts of their test preparation experience (e.g., instrumental motivation for taking the preparatory courses versus a desire to improve English language skills), and this experience directly influenced their perceived value.

Similar to many test-takers in previous studies (e.g., Green, 2006a, 2006b), the students in this study paid great attention to what the test was (e.g., test structure, test format) and how the test was administrated (e.g., test-taking process). This attention was related directly to their motivation for taking such courses, e.g., “I just need a score.” However, it is interesting to note the shift from this instrumental motivation to the understanding and accepting the need to improve English skills; the teachers played an important role in broadening students’ goals. The students believed their teachers “know what is needed” because these teachers had successful educational and test-taking experiences. Therefore, the students attended to the strategies their teachers demonstrated (e.g., using logic flow chart to practice listening skill), and their attention was rewarded by the self-perceived improvement in specific skill or by the substantial gains in the test scores.

Some teachers of the TOEFL iBT test preparation courses (e.g., all eight teachers at Centre L) had taken the TOEFL iBT themselves; this experience was a highly valued attribute a teacher could have according to students. The importance of putting testing researchers and/or practitioners through the tests has been demonstrated in recent studies (e.g., DeLuca et. al., 2013; Fox & Cheng, 2015). For example, the testing environment (e.g., the different pace of test-takers’ completing test sections, as Jack (CU-09), one student participant at Site 3, mentioned) would interfere with test-takers’ performance on the tasks and, therefore, on the test, thus introducing variance to test-takers’ scores. Teachers sharing their experiential recommendations
on the test-taking process was found to help students become aware of situations that might be barriers to demonstrate their actual English language proficiency.

Timeline was closely related to how the students perceived the value of the test preparation courses, i.e., the efficiency of test familiarization and the acceptance of the need for improving English skills. With a longer time frame, students were more likely to practice the strategies targeted at English skills (e.g., six months for Rachel (CU-01), a student participant at Site 3). However, this question awaits further study by language testing researchers and language teachers to explore: for how long do students need to improve English proficiency to see their improvement reflected in a higher test score?

Engagement in learning communities and becoming motivated and inspired by peers and teachers are perceived as being both important and useful during the test preparation process. Learning communities offer a venue for extending in-class instruction, which might be limited by time constraints; students do appreciate the information sharing and interactions with peers and teachers when participating in such learning communities. An important set of documents some students collected through these communities (e.g., online forum) is a collection of narratives by high TOEFL iBT achievers of their test preparation process. Many of the accounts demonstrated how they achieved high scores, starting from a medium-level of English proficiency (as indicated by a score of a national English test in China). The successful experience of peers is a source of strength and inspiration a shared perception of value among the students, including the students at Centre L and those taking the TOEFL iBT courses at other centres. As William (CU-12) said, “taking test preparation courses was more an experience of being inspired.” Students gained encouragement and inspiration from examples of their peers experiencing success, and drew on such narratives for motivation to persist throughout the test
preparation process. Although the benefits of collaborative learning (e.g., working in groups or with peers) have been recognized (e.g., Cheng & Fox, 2008), an understanding of learning communities initiated by students and/or teachers, especially those that offer support from these communities both within and beyond class boundaries, remains limited (e.g., Potts, 2005).

**Discussion by Research Purpose and Green’s Washback Model**

As stated in Chapter Two of the literature review, test preparation research in language testing field has been understood within the broader washback research as both deal with the core issue of the interrelationship of testing, teaching and learning. Test preparation is an educational phenomenon in which testing influences teaching, learning, and, subsequently, bringing consequences to test score interpretation and uses. Test preparation studies thus aim to examine the interrelationship between testing, teaching, and learning (Yu, 2012) as washback research studies do. However, not many test preparation studies have explicitly stated the specific washback framework(s) or conceptual model(s) that guided the empirical investigations (e.g., Ferman, 2004; Gan, 2009; Tsagari, 2012). Consequently, as discussed in the literature review, the understanding of test preparation practices is limited to particular participant groups (i.e., mostly on teachers, some on students, and few on administrators) and also limited to the kinds of evidences that focus on either teaching (e.g., Gibson & Swan, 2008; Tsagari, 2012), learning (e.g., Mickan & Motteram, 2009; Rahimi & Nazhand, 2010), or outcomes in relation to test performance (e.g., Gan, 2009; Robb & Ercanbrack, 1999). This current study is one of the first test preparation empirical studies that make this explicit connection with and address the guidance of washback studies on test preparation research.

Green’s washback model (2007) includes three dimensions (direction, variability, and intensity) to conceptualize multiple variables in teaching and learning interacting with the
influence of test (Weir & Milanovic, 2007), and thus provides a theoretical lens with which to understand the complexity of test preparation phenomenon. Using Green’s (2007) washback model, my study on test preparation phenomenon is one of the extensive empirical studies that explicitly guided by a washback theoretical framework. Through collecting evidences from multiple sources with respect to the nature, the perceived effects, and the perceived value, my study unpacked the test preparation phenomenon and also addressed how testing, teaching and learning influenced each other.

In the following section, I will illustrate this contribution below. I will first recap the three interrelated dimensions of Green’s washback model—direction, variability, and intensity, and then re-examine the findings within the three dimensions—to address the broad purpose of this study, which is to understand the phenomenon of Chinese students’ test preparation for high-stakes English language tests. I will then continue with discussing the contribution of my findings to Green’s washback model.

**Three Dimensions in Green’s Washback Model**

Washback *direction*, as the first dimension in this model, conceptualizes that all tests have the potential to bring positive or negative influences to teaching and learning, and that the direction of being positive or negative is related to whether teaching and learning target the broad focus (e.g., English language skills) or the narrow focus (e.g., item format, content) on test features. Washback *variability*, the second dimension, is conceptualized as being the differences in participants’ characteristics and perceptions. Variability is highlighted as knowledge and acceptance of test demands, and the resources to meet these demands. Washback *intensity*, the third dimension, is conceptualized as the extent of washback on participants, and extent of intensity is directly influenced by participants’ perceptions of test importance and test difficulty;
intensity is also related to participants’ variability (e.g., knowledge of test demands). In the following, I discuss my findings for each of the three dimensions.

**Direction**

By triangulating the interview data and the classroom observation data, I was able to determine that there was both positive and negative washback associated with test preparation practices. Positive washback occurred when students worked to improve their English skills and were able to reduce the influence of the factors in the test or the test taking process that could interfere with an accurate estimation of their English skills. It should be noted that positive washback was related to students’ characteristics, including their English language proficiency and status of test repeaters. For some students who took the test for first time, they may already have had appropriate English skills, and they simply needed test familiarization and stress management and to reduce the influence of factors that could interfere with their performance. This accordingly enhances student performance on the focal constructs, which enhance the validity of test score use. For some students who repeated the test, they realized the need to improve English skills from their previous test attempts; therefore, the students’ efforts to improve English skills were interpreted also as the positive washback. Evidence of negative washback was found from teachers’ interviews that, some students might misinterpret or misuse the strategies that teachers demonstrated (e.g., using clues in distractors to do multiple-choice questions in reading tasks without reading the passage), and the teachers believed that these students might end up using the strategies *per se* to find an answer for the test task rather than using their English skills and being assisted by the strategies.

My findings of the perceived effects of test preparation practices on students’ test scores also revealed the dual directions in which the test preparation practices led. Despite no
immediate effects on increasing test scores, some teachers communicated with students that substantial test score gains were achievable by following appropriate strategies and targeted study plans. They further advised that substantial test score gains could be achieved through continuous practice that focused on English skills. Substantial test score increases, according to both teachers and students, were attributed to efforts made to improve English skills, thus reflecting the positive direction that test preparation practices brought to teaching, learning, and testing. Also the participants mentioned that they have found a learning community through test preparation, which can help them in the long run. The negative direction was noticed when some students tried to answer test items without using English skills; for example, to distinguish distractors for multiple-choice questions in TOEFL iBT reading tasks.

My findings suggest that the direction of test preparation practices was related to the understanding of test demands (i.e., to demonstrate English language skills). It was observed among the teachers at Centre L that they included a critical first step in their instruction—an understanding of what the test was or what the test demands were. Test demands were introduced as the essential language skills (e.g., forming ideas and presenting ideas in logical way that Henry introduced in the writing module) and the knowledge of the test (e.g., test format, requirement of test tasks) that were needed to perform test tasks. When the participants (i.e., teachers, students, and administrators) understood sufficient English skills was the key to successful test scores, the test preparation practices they participated in positively contributed to teaching and learning, and also to testing because the focus of their practices was on the intended construct—English language skills. It should be noted that the ultimate direction as positive or negative, specifically, the understanding of test demands was not directly determined by any single group of participants. Rather, this outcome was the result of the interaction of participants’
characteristics (e.g., teaching experience, test-taking experience), perceptions (e.g., teaching beliefs), and context (e.g. particular test preparation centres). The differences in these aspects (i.e., characteristics, perceptions, and context) contributed to understanding the ultimate direction in which test preparation practices could lead.

**Variability**

Green (2007) emphasized that variability is demonstrated as the difference in participants’ knowledge/understanding of test demands, resources to meet test demands, and acceptance of test demands. My findings from interviews and observations clearly showed that the participants brought different extent of understanding of test demands (e.g., some students’ claim of knowing nothing except for the importance of the TOEFL iBT test score versus some students’ knowledge of the TOEFL iBT from previous test-taking experience), and that the participants had a variety of resources to support test preparation practices (e.g., extra courses for improving English skills, online community), and reached varied perceptions of test demands (e.g., improvement in English skills, strategies to increase test scores). More importantly, my findings clearly presented how such differences contributed to the direction of influence on teaching and learning. The positive direction that test preparation practices focused on English skills was observed among the students who had previous unsuccessful test attempts; the teachers who had successful experience of overseas academic studies and believed that improvement in English skills ultimately contributed to test score gains; and the administrators who held the belief that educational value should be prioritized over business profit. All these participants made concerted efforts to emphasize the appropriate understanding of test demands and to offer additional resources with which to support such understanding (e.g., a series of English skills courses, teachers’ blogs). The positive direction that test preparation practices helped with test
familiarization was observed among the students who had limited knowledge of the test (e.g., test features, test-taking procedure), and the teachers who shared test-taking experiences and emphasized the relationship between test-taking strategies and English skills.

It should be emphasized that the participants’ characteristics (e.g., teachers’ and students’ test-taking experience) and perceptions (e.g., teaching beliefs, motivation of participating in test preparation) interacted with the context in which they were situated—specifically speaking, the particular test preparation centres in which these participants were situated. By triangulating the collected documents with regards to the operation of the test preparation centre (e.g., public lectures, personnel structure, structure of offered courses) with the interview data of the administrators, my findings revealed how a commercial test preparation centre influence test preparation practices. The business belief that educational value would ultimately contribute to profit observed in Centre L provided the contextual support for the positive influence of test preparation practices to occur. However, this result needs to be interpreted with caution, the similar contextual support at Centre L may not be observed in every test preparation centre in China, considering the huge scale of test preparation industry in China (which represented by the thousands test preparation centres).

**Intensity**

My findings support the relationship between the intensity and the perceptions of test importance and test difficulty. Both the participants’ interviews and the questionnaire responses showed they associated great importance with the tests; and some teachers conveyed the message that successful test scores were “achievable”, but it was not an easy task because many students needed to take the test two or more times to achieve success. Given the perceptions of importance and “achievable” difficulty, the participants made great efforts to participate in test
preparation practices. These findings confirmed Green’s conceptualization that washback to participants is most intense when they think the test is important and challenging, but achievable.

My findings also revealed that the intensity of test preparation practices was associated with the variability of these practices. Some teachers used various resources to present the appropriate understanding of test demands (e.g., the official guide to the TOEFL iBT, first-hand test-taking experience); some students repeatedly expressed their trust in these teachers because of the “reliable instruction” and the “experiential pedagogy” the teachers offered. Subsequently, students’ trust was reflected in the acceptance and following of teachers’ instructions. All of these practices indicated that the influence of test preparation practices on teaching and learning was intensified because the understanding of the test demands was well communicated between teachers and students.

**Contribution to Green’s Washback Model**

The contribution of my study to Green’s washback model is twofold. First, my findings confirmed that Green’s washback model is a succinct theoretical conceptualization to understand test preparation phenomenon within the broader umbrella of washback studies. The re-examination of my findings from three dimensions of Green’s washback model above reflects the directions that test preparation practices led, the variability related to test preparation practices, and the intensity with which test preparation practices worked; and also my findings suggest the interrelationship among three dimensions, specifically, how the direction that test preparation practices led and the intensity of test preparation practices on teaching and learning were associated with the variability demonstrated among the participants (e.g., characteristics, perceptions), which demonstrate the complexity of this educational phenomenon. Further, my study extended the utilization of Green’s washback model to a context that is different from the
English for Academic Purpose (EAP) context in which Green (2007) created this model to examine test preparation practices for the IELTS in UK. With the achieved understanding of test preparation phenomenon in the Chinese context, the findings of my study suggest that Green’s model is able to provide the theoretical lens to examine test preparation phenomenon in the English as a foreign language (EFL) context and can be applied to similar language learning contexts.

Second, my findings further provided a tentative component that may supplement the intensity dimension in Green’s model. My findings showed that the message of “there is definitely a way out there” to achieve success on the test was well conveyed to the students through in the following several ways through test preparation: the teachers’ personal “painstaking” but “successful” prior test preparation and English learning experiences; the previous cohorts of students’ outcomes that achieved high scores within reasonable timelines by following an “appropriate” targeted study plan; the detailed instructions on how to improve a particular English skill; the support and the shared resources within the test preparation communities that emerged amongst student peers and teachers. The students in my study expressed great value for all of the above aspects because they believed that success on the test was within reach (although not “within arm’s reach”) with the help of the specific strategies to keep focused, and also the motivational strength to keep persistent. Therefore, even though the students did not see immediate increases in test scores, the influence on their learning could be intensified because the students knew they had the access to the relevant resources and people to support their efforts, by helping them understand what to do and how to continue to do, until they achieved success. These findings indicate that the perceptions of accessibility to resources and people also played a role in influencing the intensity of test preparation practices on
students’ learning. Therefore, the findings of my study suggest that perception of accessibility of relevant resources and participants could be a component to supplement the intensity dimension of Green’s washback model; and the perception of accessibility could mediate with perception of test importance and perception of test difficulty, and altogether influencing the intensity that test preparation practices on the participants.

Implications

I will discuss the implications of my study with respect to two aspects: implications for practice and implications for research.

Implications for practice

My findings have implications for prospective students who are considering test preparation courses before they take high-stakes English language tests such as the TOEFL iBT and the CET4. The students in this study confirmed the value of attending test preparation courses, even though no student reported the immediate substantial test score gains that test preparation centres claimed. Prospective students who expect to gain a sense what the test and test-taking processes look like can benefit from short-term test preparation courses because the instruction of such courses helps students navigate the test sections and test-taking procedures, and also helps reduce students’ anxieties caused by limited knowledge of test. It should be noted that the short-term courses are probably best suited for the students who already have confidence in their English language skills. It is important for prospective students to know they should not expect short-term effects of substantially increasing test scores from the intensive test preparation courses (e.g., two-weeks courses), but rather to look for the long-term outcome of achieving successful test scores from appropriate and targeted strategies to improve English skills. My findings on the strategic practices that may help students achieve success on tests
(e.g., strategies for improving listening skills, resources shared in learning communities) could benefit the prospective test takers, who may not have the budget for expensive test preparation courses.

The findings of my study also revealed a positive washback in the test preparation practices of the TOEFL iBT. This finding is contrary to Chau’s (2008) study, which reported detrimental effects of the test preparation practices for the writing test tasks on developing Chinese students’ skills of writing in English. My study showed the participants had positive perceptions of test preparation practices: the teachers and the administrators repeatedly emphasized the need for English skill improvement as the ultimate strategy for increasing test scores, and they kept this strategy as the central focus in their instruction and supervision. More importantly, my study identified the contextual factors that contributed to such positive experiences (e.g., the teachers’ experience of English language learning, the necessary understanding of the test demands, the value of teaching and learning). The findings of this study can also be shared with current and prospective teachers of test preparation courses and the administrators of test preparation centres in China to help them work towards increasing positive experiences on teaching and learning, or engaging in “ethical test preparation practices” that target for the intended construct that Hamp-Lyons (1998) called for.

**Implications for research**

The findings of this study have implications for test preparation research by providing empirical evidence with which to understand the phenomenon of test preparation in general and in the Chinese context in particular.

First, the findings of the three interrelated aspects of test preparation practices—the nature, the perceived effects, and the value—provided much-needed empirical evidence on the
following: what the test preparation practices looked like (e.g., test familiarization, English skills improvement, stress management), what shaped the test preparation practices (participants’ characteristics, perceptions, and context they were situated), how the test preparation practices work (e.g., test score increase), and how valuable these test preparation practices were with respect to contributing to success on tests.

Green’s washback model provides a theoretical lens to further interpret the findings of this study: the positive (or the negative) test preparation experiences were closely related to the characteristics and perceptions of the participants involved in such experiences, and the contextual characteristics (e.g., the operational belief of test preparation centres) also contributed to supporting positive test preparation practices. Furthermore, the influence of intensity level on teaching and learning varied by the conditions with respect to the participants’ characteristics and perceptions (e.g., teachers’ and students’ shared understanding that the improvement in English skills contributed to the increase in the TOEFL iBT scores).

Second, my findings also contributed to the literature with respect to test preparation practices in the Chinese context. With the evidences extensively collected from multiple participants and multiple sources, my study provided the legitimate perspectives, including characteristics, perceptions, context, and practices, to explore and understand how the Chinese students were prepared at commercial test preparation centres. With the exploratory understanding of test preparation phenomenon in the Chinese context from these perspectives, the future studies may further the empirical investigations to the huge claim made by hundreds of test preparation centres in China: that students could achieve high scores after completing test preparation courses. The perspectives used in this study thus could help the investigations, which use rigorous quantitative research designs and methodologies to measure test score changes
before and after taking test preparation courses, to consider relevant variables in the research design and/or research finding interpretations.

**Limitations**

This study is one of the extensive large-scale empirical studies into exploring the phenomenon of test preparation in China, and the findings with regards to the nature, effects, and value of test preparation practices contribute to understanding whether these practices lead to positive or negative direction, how the variability in participants’ characteristics, perceptions, and situated context is related to the direction of test preparation practices, and how the extent of intensity influenced teaching and learning varies.

Despite the contributions, this study had two major limitations. The first limitation was due to the practical constraints of collecting data from the same groups of participants over their complete test preparation experiences. Ideally, the same groups of the participants should be followed through their complete test preparation experience. However, the fact that test preparation centres were not related to test administration became the practical constraints. For the students taking the TOEFL iBT or the CET4, they chose different test preparation centres to take test preparation courses, and they left the centres after completing these courses; they then took the TOEFL iBT at which ever test site they could secure a position (e.g., sometimes the students travelled to different cities to write the TOEFL iBT), or they took the CET4 at their own universities or colleges. Therefore, for the students who were approached at test preparation centres, it was difficult to continuously follow them after they left; for the students who were approached at universities or colleges, it was impossible to trace their records back to the previous test preparation centres they had attended.
Instead, I adopted a unique approach for three research sites to reflect the multiple stages along the whole test preparation experience, and employed multiple-method to collect data rather than doing so a linear way (i.e., during test preparation courses, recently completing test preparation courses, and achieving successful test scores after completing test preparation courses). The findings of my study, supported by the multiple-stage and multiple-method approach, clearly presented the connections regarding what the Chinese students had experienced at test preparation centres, what outcomes they had achieved, and how they valued their experiences in relation to their successful (or unsuccessful) test scores.

The second limitation of this study was that the findings of the CET4 test preparation practices were only based on the questionnaire data and the test score data of the students who had taken the CET4 test preparation courses prior to taking the CET4. I purposefully used this quantitative method on this group of the students to focus on examining the relationship between the test preparation practice and their test scores. I selected this approach to understand what relationship, if any, was between these practices and CET4 test performance. Although the statistical analyses clearly revealed non-significant relationships, classroom observations of the CET4 test preparation courses and interviews with the students—if were collected—could provide relevant evidence to understand these practices, especially how the students judged non-significant relationships. Further, the lack of a control group who did not complete test preparation courses before taking the CET4 test is also a related limitation; and the inclusion of a control group would produce the explicit findings that test preparation practices result in higher test scores for students who complete test preparation courses than those who do not.
Recommendations for Future Research

Green’s (2007) washback model provides a valuable framework with which to examine the phenomenon of test preparation from the dimensions of direction, variability, and intensity. Compared to the understanding of the interrelated relationship between the direction and the variability of test preparation practices, the understanding of the intensity needs further exploration of the factors, in addition to examining test importance and test difficulty. Specifically, future studies should continue examining the intensity in relation to the variability among participants, and collect further evidence of the conditions that lead participants to make the most (or least) efforts for test preparation.

Future research should continue exploring the contributing factors that shape teachers’ test preparation practices, especially with respect to their professional background (e.g., professional training experience) and sociolinguistic characteristics (e.g., sharing the similar non-native language background with students). My findings showed that the teachers who did not receive professional training in English language education and/or who did not have an educational background in English language, still received strong evaluations by their students. Given that a considerable number of the teachers in the test preparation courses in China have similar backgrounds (i.e., do not have professional training, or educational background in English language), this group of the teachers might be of interest for future researchers to discover what factors make those teachers equally successful in helping students achieve success.

The findings with regard to the TOEFL iBT test preparation practices at Centre L showed positive influences on the test preparation practices of teachers’ teaching and students’ learning. However, these positive influences were closely related to contextual factors (e.g., balance between educational value and business profit, recruitment criteria that prioritize educational
value). Future research may include multiple test preparation centres and use a comparison research design to further examine the operation of test preparation centres in order to further understand how the factors within these centres influence teaching and learning in test preparation experiences.

**Conclusion**

Chinese students, with strong desires to achieve success on large-scale, high stakes English language tests, make tremendous financial contributions to the massive-scale test preparation industry in China because test preparation centres make the enticing claim that test preparation courses and practices can help students achieve high test scores. However, insufficient empirical evidences with respect to what students experienced at test preparation centres in the Chinese context was neither able to describe what test preparation practices were, nor to contribute to understand test preparation phenomenon in China. This study examined the phenomenon of Chinese students taking test preparation courses for two high-stakes English language tests—the TOEFL iBT and the CET4—to understand this phenomenon. Guided by Green’s (2007) washback model, the phenomenon was understood from three interrelated aspects: (1) what were the test preparation practices Chinese students experienced and what was their experience like (nature), (2) what was the relationship between the test preparation practices and the students’ test scores (perceived effects), and (3) how did the students judge the value of these practices after achieving success on the tests (value).

Together, the findings showed that to help Chinese students achieve success on the two tests, test familiarization, English skills improvement, and stress management were the best-perceived test preparation practices used by the students, the teachers, and the administrators. However, because of the different characteristics (e.g., students’ previous test-taking experience,
teachers’ educational and professional background) and perceptions (e.g., motivation for participating in test preparation courses, expectations) the participants (students, teachers, and administrators) brought into the test preparation experience, the three test preparation practices were used and communicated (by the teachers and the administrators) and were accepted and practised (by the students) to varying extents, which mediated the perceived effects of test preparation. Although the test preparation practices did not contribute to an immediate substantial increase in test score, the students valued the benefits they received from the appropriate focus, strategies, and resources for improving their English skills as well as the motivational strength they developed to persist through the test preparation experience, because all of these aspects ultimately contributed to their success both on the tests and in English learning.

Hamp-Lyons (1998) argued that the problem of test preparation practices cannot be solved, but it can be better understood through research. This is the exactly appropriate quotation I would like to use to conclude. This study contributes to understanding the problem of test preparation practices in China, and explores the tension between business profits and educational value. Although this study cannot solve the problem, the understanding of the interplaying factors (i.e., participants’ characteristics, perceptions, and situated context) may help with the balance between gaining profits and maintaining educational value.
References


Appendix A Letter of Approval from the General Research Ethics Board

June 06, 2012

Dr. Liying Cheng
Professor
Faculty of Education
Queen’s University
Duncan McArthur Hall
511 Union Street
Kingston, ON K7L 3N7

GREB Romeo #: 6966079
Title: "CEDUC-562 11 Test Preparation: Does It Enhance Test Performance and English Language Proficiency?"

Dear Dr. Cheng:

The General Research Ethics Board (GREB) has reviewed and approved your request for renewal of ethics clearance for the above named study. This renewal is valid for one year from July 5, 2012. Prior to the next renewal date you will be sent a reminder memo and the link to ROMEO to renew for another year.

You are reminded of your obligation to advise the GREB of any adverse event(s) that occur during this one year period. An adverse event includes, but is not limited to, a complaint, a change or unexpected event that alters the level of risk for the researcher or participants or situation that requires a substantial change in approach to a participant(s). You are also advised that all adverse events must be reported to the GREB within 48 hours. Report to GREB through either ROMEO Event Report or Adverse Event Report Form at http://www.queensu.ca/ori/researchethics/GrebelRED/forms.html.

You are also reminded that all changes that might affect human participants must be cleared by the GREB. For example you must report changes in study procedures or implementations of new aspects into the study procedures. Your request for protocol changes will be forwarded to the appropriate GREB reviewers and/or the GREB Chair. Please report changes to GREB through either ROMEO Event Reports or the Ethics Change Form at http://www.queensu.ca/ori/researchethics/GrebelRED/forms.html.

On behalf of the General Research Ethics Board, I wish you continued success in your research.

Yours sincerely,

Joan Stevenson, Ph.D.
Professor and Chair
General Research Ethics Board

c.c.: Xia Ma (Doctoral Student), Co-investigator
Dr. Lesly Wade-Woolley, Chair, Unit REB
Erin Wickham, c/o Graduate Studies & Bureau of Research
Appendix B. 1 Letter of Information/Consent Form—Administrators and Teachers at Test Preparation Centres

Test Preparation: Does It Enhance Test Performance And English Language Proficiency?

Investigators: Liying Cheng, Hong Wang, Lynette May, Shahrzad Saif, Jia Ma

Principal Investigator: Dr. Liying Cheng
Faculty of Education
Queen’s University
Kingston, Ontario, Canada
(613) 533-6000 ext. 77431
E-mail: liying.cheng@queensu.ca

Research Sponsor: SSHRC Partnership Development Grant, SSHRC (Social Sciences and Humanities Research Council of Canada)

This study has been granted clearance according to the recommended principles of Canadian ethics guidelines, and Queen's policies.

Purpose of the Study
You are invited to take part in this study on the test takers’ experience of preparing for one of the four largest English language tests in the world—Test of English as a Foreign Language Internet-based Test (TOEFL iBT), the International English Language Testing System (IELTS), the Pearson Test of English (PTE), and/or the College English Test (CET), in commercial test preparation centres.

With the increasing opportunities of studying abroad and working internationally, you may have witnessed a large number of students are enthusiastic to pursue degrees abroad and jobs in some joint ventures. As the scores of some English language tests are required as the proof of English language proficiency when these students submit applications to institutions and industries, you may have the experience of coordinating, designing and/or delivering the preparation courses in order to help them to achieve the aim of obtaining their satisfactory scores of the required English language proficiency tests.

Your test preparation experience will help us (the investigators conducting this study, as listed above) have a better understanding of the nature and the extent of the test preparation you have designed and delivered, e.g., what specific preparation practices you have included in your instruction, how you have delivered or are going to deliver the preparation practices to your students.

What will happen during the study?
Depending on your role in this centre, e.g., administrators or/and teachers, your participation will be in the following three formats:
Collection of test preparation materials (Administrators and teachers)
If available, we would like to collect your test preparation materials, e.g., lesson plans for the test preparations courses (hard copy or electronic files in Word or PowerPoint), in-class handouts for students, resources for the preparation courses, minutes of the meetings with the teachers of test preparation courses, which can describe how you coordinate, design and deliver these courses. The original materials will be photocopied and returned to you.

Classroom observations (Teachers)
We would like to observe one week of your classroom instruction with your choice and agreement. The observations will focus on the contents, activities, interactions with students, and aim to understand how your preparation instructions are conducted. Your classroom instruction will be audio-taped and notes will be taken of the main events happening during the instruction. If you would like, you can review the audiotapes and the notes taken during the classroom observation (see Appendix C of Classroom Observation Schedule).

One-hour one-on-one interview (Administrators and teachers)
You will participate in a one-hour one-on-one interview with one of us. This interview will be roughly scheduled two weeks after the test preparation courses start. You have the full choice of the time and location for this interview. During the interview, you will be mainly asked how you think of your experience of managing, designing and delivering the preparation courses. Depending on your experience of teaching test preparation courses for one of the four English language tests we are going to investigate, you may encounter the following questions or similar ones for your own experience (A for administrators, and T for teachers):
(A) What do you pay more attention to when thinking of the TOEFL, IELTS, PET, or CET preparation courses?
(A) How do you match your preparation courses with students’ aim of achieving their satisfactory TOEFL, IELTS, PET, or CET scores?
(A) How do you select teachers who can deliver these TOEFL, IELTS, PET, or CET preparation courses? What are the key criteria for your selections?
(T) How do you decide what preparation activities should be included in your instruction?
(T) How do you organize your instruction on the preparation courses to benefit your students?
(T) How do you deal with the interactions with students or feedback from students on the classes?

In addition, some questions about your demographic information, e.g., education and work experience, about the test preparation materials, and about what being observed in your classroom instructions, will also be asked during this interview. We will audio-tape the interview and also take notes on your answers to the interview questions. You can book an appointment with us if you would like to review the interview transcripts and the notes we take during the interview.

Are there any risks to doing the study?
You may feel uncomfortable with or stressful to some interview questions, and/or classroom observation. If you feel so, you do not need to answer the questions. You are not obliged to participate in classroom observation if you are uneasy, and you can withdraw from this study at
any time. Your participation in this study is completely confidential. We will not use your name or any information that would allow you to be identified. No one except us will know whether you participated unless you choose to tell them.

**Are there any benefits to doing the study?**
The research will not benefit you immediately and directly. With hearing your perceptions of being involved in test preparation courses, and along with the voices from your students, you can help us to better understand how the test takers are prepared within your preparation centres.

Upon the completion of all or part of the interview and/or the classroom observation, if you would like, we will lead a one-hour seminar to the students on the general issues of applying for institutions abroad, or a one-hour seminar to the teachers on the updated resources and information about these tests, e.g., the TOEFL iBT, the IELTS, which we could gather but would not be easily accessible for the teachers at your centres.

**Who will know what I said or did in the study?**
You are participating in this study anonymously. We will make every effort to protect your confidentiality and privacy.

- We will not use your name or any information that would allow you to be identified.
- If needed, we will use the pseudonym you would like to be referred or assign one to you.
- The information you provide and present in the test preparation materials and/or the interview the will be kept in electronic files on a computer and protected by a password.
- Only the investigators and the designated research assistants have the access to the files.
- Once the study is complete, the files, without identifying information, will be destroyed.

**What if I change my mind about being in the study?**
Your participation in this study is voluntary. You are free to withdraw at any time, even after signing the consent form. There will be no consequences to you with your withdrawal. In cases of withdrawal, any data you have provided will be destroyed unless you indicate otherwise.

**How do I find out what was learned in the study?**
We expect to have this case study completed by approximately June, 2012. The results of this study will be presented at academic conferences. If you would like a brief summary of the results, please let us know your contact addresses in the consent form so we can send it to you.

**Questions about the study**
Any questions about study participation may be directed to the Liying Cheng at liying.cheng@queensu.ca. Any ethical concerns about the study may be directed to the Chair of the General Research Ethics Board at chair.GREB@queensu.ca or 613-533-6081.
CONSENT FORM

Please sign one copy of this Consent Form and return to us. Retain the second copy for your records.

I have read the information presented in the Letter of Information about the study Test Preparation: Does It Enhance Test Performance And English Language Proficiency? (Year One), which is conducted by Liying Cheng of Queen’s University. I have had the opportunity to ask questions about my involvement in this study and to receive additional details I requested. I understand that if I agree to participate in this study, I may withdraw from the study at any time. I have been given a copy of this form. I agree to participate in the study.

Signature: __________________________________________

Name of Participant (Printed): ________________________________

1. Would you like to receive a summary of the study’s results?
   ___ Yes, I would like to receive a summary of the study’s results. Please send it to this email address __________________ or to this mailing address __________. ___ No.

2. Would you like to have your identity kept confidential?
   ___ Yes.
   ___ No, I prefer to be identified or have a pseudonym used. Please refer to me as __________.
   ___ No, our test preparation centre prefers to be identified or have a pseudonym used. Please refer to our centre as ____________________.

3. Would you like to review the notes taken during the classroom observation?
   ___ Yes. Please contact me at ________________.
   ___ No.
Appendix B. 2 Letter of Information/Consent Form—Students at Test Preparation Centres in China

Test Preparation: Does It Enhance Test Performance And English Language Proficiency?

Investigators: Liying Cheng, Hong Wang, Lynette May, Shahrzad Saif, Jia Ma

Principal Investigator: Dr. Liying Cheng
Faculty of Education
Queen’s University
Kingston, Ontario, Canada
(613) 533-6000 ext. 77431
E-mail: liying.cheng@queensu.ca

Research Sponsor: SSHRC Partnership Development Grant, SSHRC (Social Sciences and Humanities Research Council of Canada)

This study has been granted clearance according to the recommended principles of Canadian ethics guidelines, and Queen’s policies.

Purpose of the Study
You are invited to take part in this study on the test takers’ experience of preparing for one of the four largest English language tests in the world—Test of English as a Foreign Language Internet-based Test (TOEFL iBT), the International English Language Testing System (IELTS), the Pearson Test of English (PTE), and/or the College English Test (CET), in commercial test preparation centres.

With the increasing opportunities of studying abroad and working internationally, you may be planning to study abroad or find jobs in joint ventures. As the scores of some English language tests are required as the proof of English language proficiency when you apply, you may aim to obtain a satisfactory score on those tests to smooth your applications. In order to achieve this aim, you may have the experience of taking some preparation courses.

Your test preparation experience will help us (the investigators conducting this study, as listed above) have a better understanding of the test preparation you have received, e.g., what specific preparation practices you have attended to, how you have utilized or are going to utilize the preparation practices.

What will happen during the study?
First, we would like to collect your test preparation materials, including your class notes, your writing samples, your logs, that can reflect and record your engagement in the test preparation courses. The original test preparation materials will be photocopied and returned to you.
Then, you will participate in a 45-minute one-on-one interview with one of us. This interview will be roughly scheduled two weeks after you start the preparation courses. You have the full choice of the time and location for this interview. During the interview, you will be mainly asked how you think of your experience of taking the preparation courses. Depending on your experience of taking test preparation courses for one of the four English language tests we are going to investigate, you may encounter the following questions or similar ones for your own experiences:

- Can you tell me why are you going to take these TOEFL, IELTS, PET, or CET preparation courses?
- What do you pay more attentions to when you take these preparation courses? Why?
- How do you integrate teachers’ instruction on the TOEFL, IELTS, PET, or CET preparation courses with your own preparation?

Some questions about your demographic information, e.g., age and education background, and about your test preparation materials will also be asked during this interview. We will audio-tape the interview and also take notes on your answers to the interview questions. You can book an appointment with us if you would like to review the interview transcripts and the notes we take during the interview.

**Are there any risks to doing the study?**
The risks involved in participating in this study are minimal. You may feel uncomfortable with or stressful to some interview questions. If you feel so, you do not need to answer them, and you can also withdraw from this study at any time. Your participation in this study is completely confidential. We will not use your name or any information that would allow you to be identified. No one except us will know whether you participated unless you choose to tell them.

**Are there any benefits to doing the study?**
This study will not benefit you directly. With hearing your perceptions of taking the test preparation courses, and along with the voices from your teachers and administrators at this test preparation centre, you can help us to better understand how the test takers are prepared.

Upon the completion of all or part of the interview, if you would like, we will provide you with a 30- to 45-minute consultation about the questions that you may confront with when you are planning to apply for institutions abroad.

**Who will know what I said or did in the study?**
You are participating in this study anonymously. We will make every effort to protect your confidentiality and privacy.

- We will not use your name or any information that would allow you to be identified.
- If needed, we will use the pseudonym you would like to be referred or assign one to you.
- The information you provide and present in the test preparation materials and/or the interview the will be kept in electronic files on a computer and protected by a password.
- Only the investigators and the designated research assistants have the access to the files.
- Once the study is complete, the files, without identifying information, will be destroyed.

**What if I change my mind about being in the study?**
Your participation in this study is voluntary. You are free to withdraw at any time, even after signing the consent form. There will be no consequences to you with your withdrawal. In cases of withdrawal, any data you have provided will be destroyed unless you indicate otherwise.

**How do I find out what was learned in the study?**

We expect to have this case study completed by approximately June, 2012. The results of this study will be presented at academic conferences. If you would like a brief summary of the results, please let us know your contact addresses in the consent form so we can send it to you.

**Questions about the study**

Any questions about study participation may be directed to the Liying Cheng at liying.cheng@queensu.ca. Any ethical concerns about the study may be directed to the Chair of the General Research Ethics Board at chair.GREB@queensu.ca or 613-533-6081.

---

**CONSENT FORM**

*Please sign one copy of this Consent Form and return to us. Retain the second copy for your records.*

I have read the information presented in the Letter of Information about a study *Test Preparation: Does It Enhance Test Performance And English Language Proficiency? (Year One)*, which is conducted by Liying Cheng of Queen's University. I have had the opportunity to ask questions about my involvement in this study and to receive additional details I requested. I understand that if I agree to participate in this study, I may withdraw from the study at any time. I have been given a copy of this form. I agree to participate in the study.

Signature: ______________________________________

Name of Participant (Printed): ________________________________

1. Would you like to receive a summary of the study’s results?
   ___ Yes, I would like to receive a summary of the study’s results. Please send it to this email address ________________________________ or to this mailing address ________.
   ___ No.

2. Would you like to have your identity kept confidential?
   ___ Yes.
   ___ No, I prefer to be identified or have a pseudonym used. Please refer to me as ________.

3. Would you like to review the interview transcripts and the notes taken during the interview?
   ___ Yes. Please contact me at ________________.
   ___ No.
Appendix B. 3 Letter of Information/Consent Form—Students at Universities in China

Test Preparation: Does It Enhance Test Performance And English Language Proficiency?

Investigators: Liying Cheng, Hong Wang, Lynette May, Shahrzad Saif, Jia Ma

Principal Investigator: Dr. Liying Cheng
Faculty of Education
Queen’s University
Kingston, Ontario, Canada
(613) 533-6000 ext. 77431
E-mail: liying.cheng@queensu.ca

Research Sponsor: SSHRC Partnership Development Grant, SSHRC (Social Sciences and Humanities Research Council of Canada)

This study has been granted clearance according to the recommended principles of Canadian ethics guidelines, and Queen's policies.

Purpose of the Study
You are invited to take part in this study on examining the effects of the test preparation you are involved in on your test performance and English language proficiency. Your participation will help us to achieve a better understanding of how the test preparations you have received benefit your test performance and English language proficiency.

What will happen during the study?
Your participation is to take one sample test of the English language test that you plan to take, (i.e., Test of English as a Foreign Language Internet-based Test (TOEFL iBT), the International English Language Testing System (IELTS), the Pearson Test of English (PTE), or the College English Test (CET)) (see Appendix F for a sample test description and one equivalent English language proficiency test (see Appendix G for a sample test description). The two tests will be administered before your start the test preparation and when you complete the test preparation (a period of 10-12 weeks).

Are there any risks to doing the study?
The risks involved in participating in this study are minimal. You may feel uncomfortable with or stressful to taking the tests you are invited to take even though taking the tests is part of your test preparation. If you feel so, you do not need to take them. You can also withdraw from this study at any time. Your participation in this study is completely confidential. We will not use your name or any information that would allow you to be identified. No one except us will know whether you participated unless you choose to tell others.

Are there any benefits to doing the study?
This study will not benefit you immediately and directly. With this study, you can help us to better understand how your test performance and English language proficiency are influenced by the test preparations you are involved in.

**Who will know what I did in the study?**
You are participating in this study anonymously. We will make every effort to protect your confidentiality and privacy.

- We will not use your name or any information that would allow you to be identified.
- Only the investigators have the access to the files on your test performance.
- Once the study is complete, the files, without identifying information, will be destroyed.

**What if I change my mind about being in the study?**
Your participation in this study is voluntary. You are free to withdraw at any time, even after signing the consent form. There will be no consequences to you with your withdrawal. In cases of withdrawal, any data you have provided will be destroyed unless you indicate otherwise.

**How do I find out what was learned in the study?**
We expect to have this study completed by approximately June, 2013. The results of this study will be presented at academic conferences. If you would like a brief summary of the results, please let us know how you would like it sent to you.

**Questions about the study**
Any questions about study participation may be directed to the Liying Cheng at liying.cheng@queensu.ca. Any ethical concerns about the study may be directed to the Chair of the General Research Ethics Board at chair.GREB@queensu.ca or 613-533-6081.
CONSENT FORM

Please sign one copy of this Consent Form and return to us. Retain the second copy for your records.

I have read the information presented in the Letter of Information about a study Test Preparation: Does It Enhance Test Performance And English Language Proficiency? (Year Two), which is conducted by Liying Cheng of Queen’s University. I have had the opportunity to ask questions about my involvement in this study and to receive additional details I requested. I understand that if I agree to participate in this study, I may withdraw from the study at any time. I have been given a copy of this form. I agree to participate in the study.

Signature: ______________________________________

Name of Participant (Printed): ______________________________

Would you like to receive a summary of the study’s results?

___ Yes, I would like to receive a summary of the study’s results. Please send it to this email address _____________________________ or to this mailing address _____________________________.

___ No.
Appendix B. 4 Letter of Information/Consent Form—Students at Universities in Canada

Test Preparation: Does It Enhance Test Performance And English Language Proficiency?

Investigators: Liying Cheng, Hong Wang, Lynette May, Shahrzad Saif, Jia Ma

Principal Investigator: Dr. Liying Cheng
Faculty of Education
Queen’s University
Kingston, Ontario, Canada
(613) 533-6000 ext. 77431
E-mail: liying.cheng@queensu.ca

Research Sponsor: SSHRC Partnership Development Grant, SSHRC (Social Sciences and Humanities Research Council of Canada)

*This study has been granted clearance according to the recommended principles of Canadian ethics guidelines, and Queen's policies.*

**Purpose of the Study**
You are invited to take part in this study on exploring the relationship between your obtained scores on the English language test (TOEFL iBT, IELTS, PTE, or CET) and your current English language proficiency in your academic study. Your participation will help us to achieve a better understanding of how the test preparation you have received has the effects on your English language proficiency and your current academic study at this English-speaking university.

**What will happen during the study?**
You are invited to participate this study in the following two ways: 1) answering a questionnaire (see Appendix J). The questionnaire will ask you for your demographic data, your obtained admission test scores (TOEFL, IELTS, PTE or CET), your test preparation experiences (if any), and your academic achievement such as credits earned and GPAs; and 2) participating a follow up focus-group discussion about the linguistic demands of your current university studies, and the extent to which your obtained scores reflected your ability to cope with the linguistic demands of studying in this English-speaking university.

**Are there any risks to doing the study?**
The risks involved in participating in this study are minimal. You may feel uncomfortable with or stressful to some questions in the questionnaire and/or during the focus-group discussion. If you feel so, you do not need to answer them. You can also withdraw from this study at any time. Your participation in this study is completely confidential. We will not use your name or any information that would allow you to be identified. No one except us will know whether you participated unless you choose to tell them.

**Are there any benefits to doing the study?**
This study will not benefit you immediately and directly. With this study, you can help us to achieve a better understanding as to how the test preparation you have received has the effects on your English language proficiency and your academic study in English-speaking countries.

**Who will know what I said or did in the study?**
You are participating in this study anonymously. We will make every effort to protect your confidentiality and privacy.

- We will not use your name or any information that would allow you to be identified.
- If needed, we will use the pseudonym you would like to be referred or assign one to you.
- The information you provide in the questionnaire and the focus-group discussion will be kept in electronic files on a computer and protected by a password.
- Only the investigators have the access to the files.
- Once the study is complete, the files, without identifying information, will be destroyed.

**What if I change my mind about being in the study?**
Your participation in this study is voluntary. You are free to withdraw at any time, even after signing the consent form. There will be no consequences to you and your study with your withdrawal. In cases of withdrawal, any data you have provided will be destroyed unless you indicate otherwise.

**How do I find out what was learned in the study?**
We expect to have this study completed by approximately March, 2014. The results of this study will be presented at academic conferences. If you would like a brief summary of the results, please let us know how you would like it sent to you.

**Questions about the study**
Any questions about study participation may be directed to the Liying Cheng at liying.cheng@queensu.ca. Any ethical concerns about the study may be directed to the Chair of the General Research Ethics Board at chair.GREB@queensu.ca or 613-533-6081.
CONSENT FORM

Please sign one copy of this Consent Form and return to us. Retain the second copy for your records.

I have read the information presented in the Letter of Information about a study Test Preparation: Does It Enhance Test Performance And English Language Proficiency? (Year Three), which is conducted by Liying Cheng of Queen’s University. I have had the opportunity to ask questions about my involvement in this study and to receive additional details I requested. I understand that if I agree to participate in this study, I may withdraw from the study at any time. I have been given a copy of this form. I agree to participate in the study.

Signature: ______________________________________

Name of Participant (Printed): ____________________________

1. Would you like to receive a summary of the study’s results?
   ___ Yes, I would like to receive a summary of the study’s results. Please send it to this email address ____________________________ or to this mailing address __________.
   ___ No.

2. Would you like to have your identity kept confidential?
   ___ Yes.
   ___ No, I prefer to be identified or have a pseudonym used. Please refer to me as ________.
Appendix C. Document Review Form

<table>
<thead>
<tr>
<th>Document</th>
<th>Date</th>
<th>Type</th>
<th>Purpose(s)</th>
<th>Contents</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

200
## Appendix D. Green’s (2007b) Observation Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Activities &amp; Episodes</th>
<th>Participant organization</th>
<th>Content</th>
<th>Test acts</th>
<th>Content control</th>
<th>Student modality</th>
<th>Materials</th>
<th>Notes / Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher actions</td>
<td>Student actions</td>
<td>Class</td>
<td>Grp</td>
<td>Ind</td>
<td>Mgmt</td>
<td>Lang</td>
<td>Other</td>
<td>T-C/S</td>
</tr>
</tbody>
</table>

Note: The explanations of some terms and some abbreviated terms: T-C/S: teacher-led central activity to student; S-C/S: student-led central activity to student(s); Chor: choral work by the whole class; Mgmt: management; Proc: procedural directives; Disc: disciplinary statements; Form: the grammar, vocabulary and pronunciation; Func: function, focusing on the purpose of speech acts (requesting); Disc: discourse, focusing on the way sentences combine into sequences; Soc: sociolinguistics on how to make sentences appropriate for particular contexts; Imm: topics immediate to teachers; Acad: topics related to academic subjects; Stat: test-taking strategies; T/Text: teacher controlled text; T/S/Text: teacher and student controlled text.
Appendix E. Interview Guide for Student Participants at Centre L

Section One—Background

• Can you briefly introduce yourself? Where do you study? Any plans after graduation?
• Have you ever taken TOEFL iBT before? If so, when did you take it? How did you feel about your performance on the test?

Section Two—Experience of taking TOEFL iBT test preparation courses at this centre

• Why do you want to take TOEFL iBT preparation courses?
• Why do you choose the TOEFL iBT preparation courses at this school? Teachers? Materials provided to you? Intrigued motivation in English learning? Can you tell me something about this?
• Having been in the preparation course for a couple of weeks, have you had the moment, “Oh, yes, this is exactly why I want to be here”? Or the moment, “Oh, no, why I am here? What the hell am I doing?” Can you give me some examples?
• What course(s) or teacher(s) do you like or dislike most? What make them attractive or unattractive to you?
• How do you practice what teachers instruct you about TOEFL iBT preparation?
• What other materials/activities do you use to get you prepared for TOEFL iBT outside the preparation classrooms? Do the TOEFL iBT preparation courses or the teachers at this school make an influence on those materials and activities?
• How do you think of the in-house textbook? Other courses related with TOEFL iBT preparation? What aspects of these textbooks or other related courses you think are particularly helpful for your TOEFL iBT preparation?
• Comparing where you start two weeks ago, have you noticed some changes in the way you perceive the TOEFL iBT in your English learning, or in any aspects of your study and life? Can you give me some examples?
• Till now, what are the benefits you have received from taking these preparation courses? Can you give me some examples?
• How do you think of the TOEFL iBT preparation courses in preparing you for the future academic studies abroad?
Appendix F. Interview Guide for Teacher Participants at Centre L

Section One—Background

- Can you introduce yourself?
- How has you become a TOEFL iBT teacher? Any related experience (e.g., study abroad)? What is your motive of teaching these preparation courses?
- Have you taken TOEFL iBT before? If so, when did you take it?
- How is your experience of taking TOEFL iBT influence your teaching TOEFL iBT preparation courses?
- Why do you choose to work in this test preparation centre? How long have you been teaching in this centre?

Section Two—Experience of teaching TOEFL iBT test preparation courses at this centre

- What specific contents do you pay attention to when you plan your courses or your part of the in-house textbook? Can you give me some examples?
- What resources do you use to plan your classes? Why?
- What specific activities do you use to teach these contents?
- What are the factors you keep in mind when you choose some particular tasks or materials for your courses? Any examples?
- What do you want your students to learn from your preparation courses?
- How do you think the interactions with students on classes?
- How do you think of the TOEFL iBT preparation courses in the broader sense of preparing students for their future academic studies?
Appendix G. Interview Guide for Administrator Participants at Centre L

Section One—Background

• Can you briefly introduce yourself? Like education background, work experience?
• How long have you been in this centre?
• What is your position at this centre? What is your responsibility of this position?
• How have you become interested in centre and in this position?

Section Two—Experience of working at this centre

• What administrative support do you provide to the teachers of TOEFL iBT preparation courses?
• What support and help do you provide to the students of TOEFL iBT preparation courses?
• What other support do you provide to the TOEFL iBT preparation courses in general?
• Do you report what you have observed on the students to teachers or supervisors of the TOEFL iBT preparation courses? If so, what do you report? Why do you think this is important to report?
Appendix H. 1 Questionnaire of CET4 Test Preparation Experience (English)

The purpose of this questionnaire is to understand your experience of preparing for College English Test Band 4 (CET4) at test preparation centres. This questionnaire has three sections: Section One Background Information, Section Two Test Importance, Section Three CET4 preparation experience. The data of this questionnaire are for the research purpose only, and have no relationship with your CET4 test scores. Thanks for your cooperation!

Section One—Background Information

1. My student number is: __ __ __ __ __ __ __ __ __ __ __ __ __ __ __ __ __ __ __ __ __ __ __ __.
2. I am ________: a: female b: male _______.
3. I started to learn English in Grade ____ (circle one)
   a. Pre-school  b. 1  c. 2  d. 3  e. 4  f. 5  g. 6  h. 7
4. I am currently studying (e.g. Engineering, Business, etc.) ____________________________.
5. My score in National Matriculation English Test (NMET) was ________________.
6. I took CET4 before (check one): No ____    Yes____
   If yes, when was your last time of taking CET4?  ____  (Year) ____(Month); and your score on this test was ____.

Section Two—Test Importance

This section explores your perceptions about CET-4. Please circle a number between 1 and 6 that best describes your views. NOTE: If the response does not apply to you, circle 0 (N/A). For example, in response to question 1 below, you may not have any children, if you do not, you would circle 0.

<table>
<thead>
<tr>
<th>N/A</th>
<th>Not important at all</th>
<th>Extremely important</th>
</tr>
</thead>
<tbody>
<tr>
<td>7. How important is my CET4 test score to</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Me</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>• My teachers</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>• My friends</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>• My parents</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8. How important is my CET4 test score to</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Gain entrance to graduate school</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>• Obtain diploma</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>• Obtain university degree</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>• Obtain academic award/scholarship</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>• Obtain a job</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>• Study/work abroad</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Section Three: CET4 Test Preparation Experience at Test Preparation Centres

This section explores your CET4 preparation experience at test preparation centres. Please answer the questions in one of the two following methods: (1) circle the provided answer that suits your experience; and (2) circle a number between 1 (strongly disagree) and 6 (strongly agree) that best describes your experiences and views. If the response does not apply to you, circle 0 (N/A).

9. Which type of CET4 test preparation courses did you take?
   Ji Cha Qiang Hua (Building English skills)  Zhen Ti Jiang Jie (Retired Test Items)  Mo Ni Kao Shi (Mock Test)  Ying Shi Ji Qiao (Test-taking Strategies)  Ci Hui Jiang Jie (Vocabulary)
   Other (Please Specify): ________________________________

10. How many hours of instruction did you receive in your CET4 preparation course?
   Less than 10  11—20  21—30  31—40  41-50  >50

11. Excluding time on CET4 preparation courses, on average how many hours per day have you spent?
   On practicing CET4 listening | None | 0.5 | 1 | 2 | 3 |
   On practicing CET4 reading | None | 0.5 | 1 | 2 | 3 |
   On memorizing vocabulary | None | 0.5 | 1 | 2 | 3 |
12. Excluding the practice of CET4 tasks on test preparation courses, so far, how many CET4 ___ have you practiced?  

<table>
<thead>
<tr>
<th>Activity</th>
<th>None</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4-6</th>
<th>7-9</th>
<th>&gt;10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening</td>
<td>None</td>
<td>0.5</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4-6</td>
<td>7-9</td>
<td>&gt;10</td>
</tr>
<tr>
<td>Reading</td>
<td>None</td>
<td>0.5</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4-6</td>
<td>7-9</td>
<td>&gt;10</td>
</tr>
<tr>
<td>Cloze</td>
<td>None</td>
<td>0.5</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4-6</td>
<td>7-9</td>
<td>&gt;10</td>
</tr>
<tr>
<td>Translation</td>
<td>None</td>
<td>0.5</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4-6</td>
<td>7-9</td>
<td>&gt;10</td>
</tr>
<tr>
<td>Writing</td>
<td>None</td>
<td>0.5</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4-6</td>
<td>7-9</td>
<td>&gt;10</td>
</tr>
</tbody>
</table>

13. When did you take CET4 test preparation course?  

- December  
- November  
- October  
- September  
- Before September  
- Before February  

14. The instructions I received in CET4 preparation courses included…

<table>
<thead>
<tr>
<th>Instruction</th>
<th>N/A</th>
<th>Strongly disagree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Teaching past CET4 test papers</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(b) Predicting CET4 tasks or topics of tasks</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(c) Having CET4 mock tests</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(d) Teaching general test-taking strategies (e.g., time management, review answers)</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(e) Teaching test-taking strategies targeting the CET4 test formats and test items (e.g., making inference using test format features or clues in the test materials)</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(f) Teaching strategies to improve general English language proficiency</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(g) Extend the discussions on topics related with the themes of the test items or test materials (e.g., topics on education, science and technology, society which are covered by some passages in reading comprehension)</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(h) Other (Please specify and circle a number):</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

15. I took CET4 test preparation courses because…

<table>
<thead>
<tr>
<th>Reason</th>
<th>N/A</th>
<th>Strongly disagree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) I wanted to familiarize CET4 test formats</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(b) I wanted to learn CET4 test-taking strategies.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(c) I wanted to improve CET4 test scores.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(d) I wanted to obtain in-house test preparation resources.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(e) I wanted to improve English language proficiency</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(f) I wanted to learn strategies to improve English language proficiency.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(g) I wanted to reduce the anxiety of taking CET4.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(h) I wanted to boost the confidence of taking CET4.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(i) I wanted to meet the parents’ expectation that I can pass CET4.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(j) I wanted to meet new friends.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(k) Others (Please specify and circle a number):</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

16. From CET4 preparation courses, I have learned to practice ___ by…

<table>
<thead>
<tr>
<th>Practice</th>
<th>N/A</th>
<th>Strongly disagree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) repeatedly listening to CET4 listening materials to familiarize myself with its pronunciation and intonation.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(b) trying to understand fully at the CET4 listening materials I have practiced.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(c) reading CET4 listening transcripts to enhance understanding.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(d) writing down whatever I heard.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(e) noting down key words and important information.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(f) finding ways to keep concentrated on the listening tasks.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
17. From CET4 preparation courses, I have learned to practice reading by…

| (a) memorizing vocabulary including parts of speech, roots and prefixes. | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (b) memorizing CET4 vocabulary or high-frequency words. | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (c) timing myself to improve my reading speed. | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (d) focusing on understanding difficult and complex sentences in the passages. | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (e) choosing options through logic elimination | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (f) guessing based on my life experience and background knowledge. | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (g) reading questions before looking for key words and sentence of questions. | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (h) searching for the answers in the text according to the sequence of questions. | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (i) scanning key words in the text before reading and understanding individual sentences carefully. | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (j) Others (Please specify and circle a number): | 0 | 1 | 2 | 3 | 4 | 5 | 6 |

18. From CET4 preparation courses, I have learned to practice writing by…

| (a) reciting sentence patterns for writing. | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (b) memorizing linking words and phrases. | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (c) reciting model essays or beautiful sentences in them. | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (d) summarizing the structures of CET4 model essays. | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (e) composing essays using past retired CET4 essay topics. | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (f) composing essays using predicted CET4 essay topics. | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (g) keeping my writing neat and tidy. | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (h) using more complex sentences | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (i) avoiding grammar and spelling mistakes. | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (j) avoiding using unfamiliar words. | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (k) Others (Please specify and circle a number): | 0 | 1 | 2 | 3 | 4 | 5 | 6 |

19. From CET4 preparation courses, I have learned to practice translation by…

| (a) reviewing the translation test items of past CET4. | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (b) reviewing the key sentence structures and phrases that are frequently assessed. | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (c) avoiding making spelling or grammar mistakes. | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (d) avoiding using unfamiliar words. | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (e) Others (Please specify and circle a number): | 0 | 1 | 2 | 3 | 4 | 5 | 6 |

20. From CET4 preparation courses, I have learned to practice cloze by…

| (a) looking for those frequently assessed questions | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| (b) training the skills for making inference from the context intensively. | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
(c) training the skills for grasping the gist intensively.
(d) selecting answers out of the options by elimination strategies.
(e) Others (Please specify and circle a number):

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

21. I think the usefulness of the test preparation practices I learned from CET 4 test preparation courses for ______ is:

| Not | | | | | Very |
|-----|---|---|---|---|
|     |   |   |   |   |

(a) CET4 listening
(b) CET4 reading
(c) CET4 writing
(d) CET4 translation
(e) CET4 cloze

- End of the Questionnaire -
Appendix H. 2 Questionnaire of CET4 Test Preparation Experience (Chinese)

大学英语四级考试辅导经历的问卷调查

同学：您好！

本问卷调查的目的是了解您参加大学英语四级考试辅导的经历。问卷由三个部分组成：个人背景、四级考试重要性，和参加四级辅导的经历。本问卷数据仅用于研究项目，您的回答与您的考试成绩没有任何关系。感谢您的支持和配合！

第一部分：个人背景

本部分了解您的个人背景，请按您的情况填写信息或选择符合的选项。

1. 我的学生证号码是 ____________________。
2. 我是 __________： a. 女性 b. 男性
3. 我从 __________开始学习英语。
   a. 幼儿园或小学前 b. 小学一年级 c. 小学二年级 d. 小学三年级 e. 小学四年级 f. 小学五年级 g. 小学六年级 h. 初一
4. 我现在所学的专业是 ____________________。（如计算机，金融）
5. 我的高考英语成绩是 ______________。
6. 我曾经考过四级考试： a. 否 b. 是
   若回答“是”，请您最近一次参加四级考试的时间是： _______年 ______月；最近一次四级成绩为 ________。

第二部分：四级考试重要性

本部分了解您对四级考试重要性的看法。请在 0（与我的实际情况不相关）和 1～6（重要性程度）号码中，圈选最符合您个人情况和观点的号码。

7. 四级考试成绩对 ________的重要性：
   • 我自己 0 1 2 3 4 5 6
   • 我的老师 0 1 2 3 4 5 6
   • 我的朋友 0 1 2 3 4 5 6
   • 我的父母 0 1 2 3 4 5 6
8. 四级考试成绩对于我 ________的重要性：
   • 报考研究生 0 1 2 3 4 5 6
   • 取得毕业证书 0 1 2 3 4 5 6
   • 取得学士学位 0 1 2 3 4 5 6
   • 争取学术奖项/奖学金 0 1 2 3 4 5 6
   • 求职 0 1 2 3 4 5 6
   • 出国进修/海外工作 0 1 2 3 4 5 6

第三部分：参加四级辅导班经历

本部分了解您参加四级辅导班的情况。下列各题请按照以下两种方式回答：(1) 从所提供的选项中，圈选适合您情况的一个选项；(2) 请在 1（与我的情况完全不一致）～6（与我的情况完全一致）号码中，圈选最符合您个人情况和观点的号码。如与您的实际情况不相关，请圈选 0 。

9. 我参加的四级辅导班是偏重：
   a. 基础强化班 b. 真题讲解班 c. 模拟考试班 d. 应试技巧班 e. 词汇讲解班 f. 其他（请具体说明）

10. 我一共参加了 ______小时的四级辅导。
    少于 10 11～20 21～30 31～40 41～50

11. 除了参加四级辅导班以外，我每天大约花 ______小时背单词。
    0.5 小时 1 小时 2 小时 3 小时
12. 在我参加的四级辅导班上，迄今为止我…

- 练听力：没有
  - 练达题：没有
  - 练阅读：没有
  - 练写：没有
  - 练翻译：没有
  - 练完形填空：没有
  - 练全套模拟题：没有

13. 我是参加的四级辅导班。

- 今年 12 月
- 今年 11 月
- 今年 10 月
- 今年 9 月
- 上半年
- 更早以前

14. 我参加的四级辅导班的教学内容包括…

- (a) 讲解历年四级真题：
  - 不相关：0
  - 完全一致：6

- (b) 预测四级考题和考点：
  - 不相关：0
  - 完全一致：6

- (c) 进行四级模拟考试：
  - 不相关：0
  - 完全一致：6

- (d) 讲授总体考试技巧（比如合理分配时间、仔细答题检查等）：
  - 不相关：0
  - 完全一致：6

- (e) 讲授针对四级题型和题目的应试技巧（比如利用题目的特点或材料中的线索进行推测）：
  - 不相关：0
  - 完全一致：6

- (f) 讲授提高英语基础知识和技能的方法：
  - 不相关：0
  - 完全一致：6

- (g) 延伸与四级考试题目主题内容相关的知识和信息（比如阅读题目中涉及的教育、科技、社会等方面的话题）：
  - 不相关：0
  - 完全一致：6

- (h) 其他（请具体说明并选择符合您情况的号码）：
  - 不相关：0
  - 完全一致：6

15. 我参加四级辅导班是因为…

- (a) 想要熟悉四级考试题型：
  - 不相关：0
  - 完全一致：6

- (b) 想要学习四级考试解题技巧：
  - 不相关：0
  - 完全一致：6

- (c) 想要提高四级考试分数：
  - 不相关：0
  - 完全一致：6

- (d) 想要获得内部备考资料：
  - 不相关：0
  - 完全一致：6

- (e) 想要在四级考试中取得理想的成绩：
  - 不相关：0
  - 完全一致：6

- (f) 想要提高英语基础能力：
  - 不相关：0
  - 完全一致：6

- (g) 想要了解英语学习方法：
  - 不相关：0
  - 完全一致：6

- (h) 想要减少考试紧张感：
  - 不相关：0
  - 完全一致：6

- (i) 想要满足父母想让我通过四级考试的期望：
  - 不相关：0
  - 完全一致：6

- (j) 想要认识新朋友：
  - 不相关：0
  - 完全一致：6

- (k) 其他（请具体说明并选择符合您情况的号码）：
  - 不相关：0
  - 完全一致：6

16. 我从四级辅导班学到的提高四级听力的方法是：

- (a) 反复听，熟悉四级的语音语调：
  - 不相关：0
  - 完全一致：6

- (b) 尽量将做过的听力题完全听懂：
  - 不相关：0
  - 完全一致：6

- (c) 听不懂时，看听力文字资料，加强理解：
  - 不相关：0
  - 完全一致：6

- (d) 做题时，边听边记关键词：
  - 不相关：0
  - 完全一致：6
做题时,边听边记重要资讯;
做题时,想办法保持注意力集中;
做题时,边听边猜测接下来会说什么;
做题时,快速预览选项,以便听的时候有针对性;
做题时,用逻辑推理排除干扰选项;
遇到难题,运用生活和背景知识猜答案;
先看问题,再到文中找题中的重点字眼;
做题时,根据题目先后顺序在文中找答案;
抓住关键字句,再认真阅读和理解个别句子;

17. 我从四级辅导班学到的提高四级阅读的方法是:
(a) 背词性词根词缀; 0 1 2 3 4 5 6
(b) 背四级词汇或高频词; 0 1 2 3 4 5 6
(c) 定时做题,提高阅读速度; 0 1 2 3 4 5 6
(d) 强化理解文中的虚词、复杂句; 0 1 2 3 4 5 6
(e) 做题时,用逻辑推理排除干扰选项; 0 1 2 3 4 5 6
(f) 遇到难题,运用生活和背景知识猜答案; 0 1 2 3 4 5 6
(g) 先看问题,再到文中找题中的重点字眼; 0 1 2 3 4 5 6
(h) 做题时,根据题目先后顺序在文中找答案; 0 1 2 3 4 5 6
(i) 其他(请具体说明并选择符合您情况的号码); 0 1 2 3 4 5 6

18. 我从四级辅导班学到的提高四级写作的方法是:
(a) 背作文套句; 0 1 2 3 4 5 6
(b) 背连接词和短语; 0 1 2 3 4 5 6
(c) 背范文或其中的句子; 0 1 2 3 4 5 6
(d) 总结范文的文章结构; 0 1 2 3 4 5 6
(e) 用往年的写作题目训练构思文章; 0 1 2 3 4 5 6
(f) 确保书写工整/卷面整洁; 0 1 2 3 4 5 6
(g) 尽量多用高级词汇; 0 1 2 3 4 5 6
(h) 尽量多用复杂句型,如从句; 0 1 2 3 4 5 6
(i) 尽量避免语法和拼写错误; 0 1 2 3 4 5 6
(j) 尽量避免用生词和不熟悉的语法; 0 1 2 3 4 5 6
(k) 其他(请具体说明并选择符合您情况的号码); 0 1 2 3 4 5 6

19. 我从四级辅导班学到的提高四级翻译的方法是:
(a) 练习往年的四级翻译题; 0 1 2 3 4 5 6
(b) 练习常考句型和词组; 0 1 2 3 4 5 6
(c) 做题时,尽量避免语法和拼写错误; 0 1 2 3 4 5 6
(d) 做题时,尽量避免用不熟悉的单词; 0 1 2 3 4 5 6
(e) 其他(请具体说明并选择符合您情况的号码); 0 1 2 3 4 5 6

20. 我从四级辅导班学到的提高四级完形填空的方法是:
(a) 做题时,注意常考考点; 0 1 2 3 4 5 6
(b) 强化练习上下文推理的技能; 0 1 2 3 4 5 6
(c) 强化训练全文把握大意; 0 1 2 3 4 5 6
(d) 强化理解文中的难句、复杂句；
(e) 做题时，用逻辑推理排除干扰项；
(f) 其他（请具体说明并选择符合您情况的号码）；

21. 我认为我从四级辅导班学到的...

<table>
<thead>
<tr>
<th>提高听力的方法</th>
<th>不知道有没有用</th>
<th>没有用</th>
<th></th>
<th></th>
<th></th>
<th>非常有用</th>
</tr>
</thead>
<tbody>
<tr>
<td>提高阅读的方法</td>
<td>不知道有没有用</td>
<td>没有用</td>
<td></td>
<td></td>
<td></td>
<td>非常有用</td>
</tr>
<tr>
<td>提高写作的方法</td>
<td>不知道有没有用</td>
<td>没有用</td>
<td></td>
<td></td>
<td></td>
<td>非常有用</td>
</tr>
<tr>
<td>提高翻译的方法</td>
<td>不知道有没有用</td>
<td>没有用</td>
<td></td>
<td></td>
<td></td>
<td>非常有用</td>
</tr>
<tr>
<td>提高完形填空的方法</td>
<td>不知道有没有用</td>
<td>没有用</td>
<td></td>
<td></td>
<td></td>
<td>非常有用</td>
</tr>
</tbody>
</table>

～问卷结束。感谢您参加本次问卷调查！～
Appendix I. Interview Guide for Student Participants at a Canadian University

Place:
Date:
Time of interview:
Interviewer:
Interviewee(s):

- Distribute Letter of Information/Consent Form;
- Once signed, distribute a recorder.

I am a graduate student from the Faculty of Education at Queen’s University in Canada. I am studying the Chinese students’ test preparation for high-stakes English language tests. The purpose of this interview is to gather perspectives from you - about your experience of taking the TOEFL test prep courses and how you think the values of the courses.

Questions:

1. First, I would like to ask you a few questions about your basic information.

   Gender:
   Age: _ ≤ 20 _ 21-25 _ 26-30 _ >30
   What is your major? What program are you in, undergraduate or graduate?
   Which year are you in your program?
   How many years have you been in Canada?

2. When did you take the TOEFL iBT?

3. How did you prepare the test? Did you take preparation courses? What type(s) of course(s) did you take? How many instructional hours were there for the course(s)?

4. Why did you decide to take the particular TOEFL iBT test preparation course(s)? Why did you choose this school or centre?

5. What did you find about the quality of the particular TOEFL iBT preparation course(s)?

6. What are the three things of this preparation course that really help you (or help you most)?
   1) What are the three things? (Such as teacher’s pedagogy, contents, resources, etc.)
   2) Can you give me some examples?
   3) Why do you think they are very helpful?
7. Is there something that you are not happy about the prep course you took? What is it or what are they? Why you are not happy with them?
8. How many teachers were there in your TOEFL iBT preparation course? Which teacher(s) impressed you most? In which way did the teacher(s) impress you?
9. How do you think about the course structure? The resources? The materials provided by the teacher or the school/textbooks?
10. Do you also learn from peers? What did you benefit from peers?
11. If you go back in time, will you prepare the TOEFL iBT the same way or differently? Will you choose a different type of the prep course? Why?
12. What were your expectations of taking the TOEFL iBT preparation courses? Do you think your expectations were met? Why or why not?
13. Now that you have prepared for the TOEFL iBT with this particular school, would you recommend the course to others? Why?
14. (a) You have the regular English language courses at your school/university, and these courses will help you improve your English ability and in turn should help you with the TOEFL iBT as well. Why do you still pay extra money to take the test prep course?
(b) Compared to the regular English course you take at school, do you think your TOEFL iBT prep course a better and effective way? Why?
15. Give the TOEFL iBT test score you got after taking this preparation course, do you think this course is worth the time, money and efforts you spent and made?
16. Do you think test preparation is a very important part in your application process? Do you think it helps you jump the hurdle? Can you get into this university without this test prep course?
17. (a) Consider all that you have learned from this particular preparation course. Are there some aspects of this course that benefited you beyond writing the TOFEL iBT? (did it teach you how to be more confident? More systematic in learning? More strategic thinking about assessment)
(b) Do think you have internalized the skills/strategies you learn from the TOEFL iBT prep course? What are the skills?
(c) Do you still use such skills/strategies/motivational influences in your current study? Are these skills still helping you with your current academic study (still supporting you)?
18. If you had the opportunity to work for this school or centre, how would you advertise the TOEFL iBT prep course you took? Can you use three words to promote the course?

That concludes our interview. I would like to thank you for coming to participate in today’s interview. Your responses will be valuable to the research I am doing. Please be assured that your responses will be confidential. Best of luck in your future work!
**Appendix J. Example of Descriptive Field Notes of Classroom Observation**

| Date: July 14, 2013 | Course: TOEFL iBT Course A  
Module: Writing  
Session #: 3rd | Teacher: Henry  
ID: CL-T3 |
|---------------------|-----------------------------------------------|
| **Time:** 8:20 am | **Recordings:**  
Folder: A  
Number: 01 |

**Observation before session starting/during breaks:**

It was a very hot day, 36 Celsius degree according to weather forecast. A ceiling fan and four standing fans were running to cool down; however, these fans did help much, with about 50 students packed in this 6 (L)*10 (W) classroom.

Beside the classroom door, one staff of the supporting team was checking each student’s registration card before the students entered into the classroom. One podium was in the most front of the classroom; and 50 chairs with writing pads were arranged in 5 rows, with a walking aisle in the middle to separate the 10 chairs in each row into left and right groups. Each student walked into his/her seat, which was indicated on the registration card.

Some students were finishing the breakfast; some were putting their notebooks on the writing pads. One boy, sitting in front of me, with sweats falling down the neck, used his ipad cover to bring more breezes to cool down. In between checking the registration card, the staff came to the podium, to check everything was ready for the teacher, the white board, the markers, the projector screen, and the microphone.

Henry came in (8:26 am); he was thin and medium height, and wore a checkered short-sleeve shirt and khaki short, looking not much different from the students in the classroom. Walking along the aisle, he smiled to one boy stuffing the mouth with a bun; while setting up his laptop and PowerPoint slides, he talked to the whole class, “I often got up late for the morning class when I was at school, and it didn’t feel good with stuffed mouth. Slow down; we have a couple of minutes to start.”

<table>
<thead>
<tr>
<th>Episode/Activity</th>
<th>Time</th>
<th>Descriptions</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introducing topic of this session</td>
<td>8:30</td>
<td>H (Henry) briefly talked about he was going to introduce one writing task—the integrated writing task.</td>
<td><strong>Material:</strong> PowerPoint slide showing (in English) “Integrated Writing Task” and “OG, Page 200” (OG— the Official Guide to the TOEFL 3rd edition)</td>
</tr>
<tr>
<td>Introducing the first two steps in this writing task</td>
<td>8:35</td>
<td>To the whole class, H introduced, to complete this task, first to read a passage, then to listen to a lecture. He then used one example in the OG, presenting to the whole class, what to pay attention to in the reading passage and listening passage.</td>
<td><strong>Material:</strong> PowerPoint slide showing (in Chinese) “Should/what should you write about what you read?” and “Does what you read help what you listen?”</td>
</tr>
<tr>
<td>Elaborating on the step of reading</td>
<td>8:47</td>
<td>H asked the whole class to read the reading passage; H used his</td>
<td><strong>Materials:</strong> The reading passage</td>
</tr>
<tr>
<td>Time</td>
<td>Activity</td>
<td>Details</td>
<td>Materials</td>
</tr>
<tr>
<td>-------</td>
<td>--------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>8:52</td>
<td>H asked the whole class about the main ideas of the reading passage;</td>
<td>(OG, p. 278) A boy and a girl, sitting two rows in front of me, shared one OG to look at the reading passage. One student answered; The main ideas summary of the reading passage (OG, p. 332) The sample writing response (OG, p. 332) PowerPoint slide: “Should/What should you write about what you read?”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H highlighted the sentences with the main ideas and also the words that indicated the logical lines of the passage;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H referred to the sample scored writing and summarized that, the main ideas in the reading passage could be included in the writing.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:02</td>
<td>H talked to the whole class the reading passage and the listening passage shared around 30% vocabulary because the topics were similar;</td>
<td>Materials: PowerPoint slide: “Does what you read help what you listen?” The listening transcript (OG p. 331) No student answered; The main ideas summary of the listening passage (OG, p. 332)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H played the listening passage to the whole class, while allowing the students to look at the transcript while listening;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H asked the whole class about the main ideas of the listening passage;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H read through the transcript and highlighted the main ideas; H emphasized the contrasting opinions between the main ideas of the reading and the listening passages.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:23</td>
<td>H explained the end of the first paragraph of the reading passage</td>
<td>Materials: The reading passage</td>
<td></td>
</tr>
<tr>
<td>and listening</td>
<td>usually include the main idea; H suggested to include such main idea in the writing; H explained the pattern of “example—explanation” in both passages; H suggested to pay attention to the ideas following a new example;</td>
<td>(OG, p. 278)</td>
<td>The listening transcript (OG p. 331)</td>
</tr>
<tr>
<td>Illustrating the pattern of presenting main ideas</td>
<td>9:50</td>
<td>H explained the main ideas were usually presented in the first paragraph in writing, whether in English or in Chinese; H used the example that some leaders in China usually said at the beginning of a meeting: “I will talk about three points in this meeting.” H talked about his own experience of being invited to a meeting at his hometown;</td>
<td>Some students showed great interest in H’s experience and burst into laughs three times in this 5-minute episode.</td>
</tr>
</tbody>
</table>
Appendix K. Example of Reflective Field Notes of Classroom Observation

<table>
<thead>
<tr>
<th>Episode/Activity</th>
<th>Reflections/Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introducing the first two steps in this task</td>
<td>Henry talked about the reading passage and the listening passage for the integrated writing tasks in quite a period of class time; I was wondering why he emphasized reading and listening and what he wanted to tell the students.</td>
</tr>
<tr>
<td>Elaborating on the step of reading</td>
<td>As Henry mentioned, the main ideas in the reading passage could be included in the writing. If so, why he still drew the students’ attention onto the words that showed the logical lines of the reading passage? What did he want to tell the students?</td>
</tr>
<tr>
<td>Illustrating the pattern of presenting main ideas</td>
<td>Henry talked about his personal experience of a meeting at his hometown, and this was not quite related to TOEFL test preparation; and what was his intention of adding this episode in the session? Was it very common to have similar episode in his instruction? If so, how did he decide what was the appropriate time to add such episodes? How did he think the students’ reactions to such episodes?</td>
</tr>
</tbody>
</table>
### Appendix L. Example of Integrating Interview Data and Questionnaire Data

<table>
<thead>
<tr>
<th>CET4</th>
<th>TOEFL iBT</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Questionnaire)</td>
<td>(Codes and quotations)</td>
</tr>
<tr>
<td>Mean (Max=6)</td>
<td></td>
</tr>
<tr>
<td>15c. To improve test scores</td>
<td>4.87</td>
</tr>
<tr>
<td>15b. To learn test-taking strategies</td>
<td>4.62</td>
</tr>
<tr>
<td>15e. To improve English language proficiency</td>
<td>4.21</td>
</tr>
<tr>
<td>15f. To learn methods of English language learning</td>
<td>4.17</td>
</tr>
<tr>
<td>15a. To familiarize tests</td>
<td>4.13</td>
</tr>
<tr>
<td>15h. To increase test-taking confidence</td>
<td>4.11</td>
</tr>
<tr>
<td>15g. To reduce test-taking anxiety</td>
<td>3.56</td>
</tr>
<tr>
<td>15i. To meet the parents’ expectations</td>
<td>3.49</td>
</tr>
<tr>
<td>15j. To meet new friends</td>
<td>2.36</td>
</tr>
</tbody>
</table>

*Note: This is the ID assigned to the student participants.*
Appendix M. Initial Versions of Figure 2

- English Language Learning
  - English skills improvement
- Test Preparation
  - De-stressing management
  - Test familiarization
- Beliefs
  - Recruiting teachers with same teaching beliefs
  - Conveying teaching belief (educational value) to students
- Strategies
  - Instruction on methods to improve English skills
  - Providing resources to continue support students’ practices
  - Offering English training courses
  - Offering public lectures to promote teaching beliefs
- Increasing confidence
  - Sharing personal experiences of English learning
  - Sharing test preparation experiences
  - Sharing test-taking strategies
  - Enhancing students’ experiences and personal stories
- Reducing anxiety
  - Balancing between English skills and test-taking strategies
  - Ensuring the possibility of improving English skills
  - Providing students to prepare for tests
- Helping goal setting
  - Experience in teaching students to prepare for tests
- Practice test tasks
  - To learn English better
  - To meet parents' expectations
- Knowledge of test
  - Desire to improve test scores
  - Aiming to prepare
- Resources of test
  - Limited knowledge of test-taking process
  - Preference on preparing

Test Preparation Centres

- English language learning
  - Test preparation
    - English skills improvement
      - Reviewing skills
      - Demonstrating methods
      - Providing resources
      - Reducing stress
      - Increasing confidence
    - De-stressing management
      - To explain the necessity of English skills
      - To share experiences
    - Test familiarization
      - To learn English better
      - To improve test scores
- Why
  - To support educational value
  - To help students “Think carefully”
- Who
  - Successful overseas educational experience
  - Extensive teaching experience
  - Successful English learning experience
  - Previous test-taking experience
  - Eager for success
- Test Preparation Centres
  - Administrators
  - Teachers
  - Students

220