

**Understanding test preparation behaviors of TOEFL iBT repeater test takers
in China**

by

Zhi (Stone) Chen

A thesis submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

MASTER OF ARTS

Major: Teaching English as a Second Language/Applied Linguistics (Computer Assisted
Language Learning)

Program of Study Committee:
Carol Chapelle, Major Professor
Jim Ranalli
Paul Koch

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this. The Graduate College will ensure this thesis is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2020

Copyright © Zhi (Stone) Chen, 2020. All rights reserved

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	iv
ABSTRACT.....	v
CHAPTER 1. INTRODUCTION.....	1
CHAPTER 2. LITERATURE REVIEW	6
2.1 Validity Issues of Repeaters’ Test Performance.....	6
2.2 Research on TOEFL Test Preparation	9
2.3 Research Questions.....	15
CHAPTER 3. METHODS.....	16
3.1 Participants.....	16
3.2 Data Description.....	17
Data Source	17
Authenticity, Truthfulness, and Accuracy of Reports.....	20
3.3 Procedures	22
Data Collection.....	22
Strategy Classification	24
3.4 Data Analysis	26
CHAPTER 4. RESULTS.....	29
4.1 Descriptive Statistics of Test Preparation Patterns and Strategies.....	29
Test-taking Patterns	29
Test Preparation Strategies.....	34
4.2 Qualitative Analysis.....	36
Newly Identified Strategies.....	36
Test Preparation Curriculum	40
Repeater Perceptions about Test Preparation Strategies	45
4.3 Quantitative Analysis.....	46
Test Preparation Strategies and Total Score Changes	46
Test Preparation Strategies and Subscore Changes	48

CHAPTER 5. CONCLUSION.....	53
5.1. Major findings	53
5.2 Limitations	54
5.3 Implications.....	55
REFERENCES.....	58
APPENDIX: AN EMAIL FROM ETS TO TOEFL IBT TEST-TAKERS	62

ACKNOWLEDGMENTS

I would like to first thank my committee chair Dr. Carol Chapelle for all of your guidance and patience throughout this project and its many drafts. Thank you as well to my committee members Dr. Jim Ranalli and Dr. Paul Koch for your support and valuable input during my defense.

ABSTRACT

Individuals who take the TOEFL test more than once are called TOEFL repeaters. TOEFL repeaters' test preparation behaviors are under studied and there is a lack of understanding about the interpretation of repeaters' multiple test scores. This study investigated TOEFL iBT (Internet-based test) repeaters' test preparation strategies and explored the relationship between repeaters' test preparation strategies and their test performance. TOEFL iBT repeaters who received satisfactory scores in the repeated test often share their test preparation experience in TOEFL-related online discussion forums. In this study, 170 Chinese TOEFL iBT repeaters' test preparation experience posts were manually collected from a major TOEFL test online discussion forum. The repeaters' test-taking behaviors, test-preparation strategies, and the implementation of strategies were presented by statistical and qualitative analyses. Using repeaters' test score changes as the outcome variable, this study ran simple and multiple regression analyses to find the relationship between repeaters' test preparation strategies and their test score improvements in the repeated tests. The results suggest that memorizing TOEFL vocabulary, getting feedback on output samples, speed listening, practicing TOEFL Practice Online (TPO), and shadow speaking are strong predictors of repeaters' test score gains. The findings may contribute to the understanding of the test preparation behaviors of the TOEFL iBT repeaters and provide relevant evidence to the support of the validity argument of the TOEFL iBT test. Implications of how the findings may benefit test takers, test users, and test developers of the TOEFL iBT test are also discussed.

CHAPTER 1. INTRODUCTION

The Test of English as a Foreign Language (TOEFL) exam is currently the most widely used English language proficiency test in the world for admitting international students into universities in English-speaking countries (Get the TOEFL Test Advantage, 2020). Many TOEFL test takers take the test multiple times hoping to improve their test scores. These test takers are called TOEFL repeaters. TOEFL repeaters make up a large proportion of the TOEFL population (Barkaoui, 2017). Wilson (1987) reports that among 221,744 candidates who took the TOEFL test between July 1977 and June 1980, 28% took the test more than once. An Educational Testing Service (ETS) survey of 14,593 test-takers who took the TOEFL iBT test in 2010 showed that 40% of them were repeaters (Liu, 2014). Many repeaters retake the test soon after their previous test. An ETS research report shows that around 10% of all TOEFL iBT test takers repeated the test during eight months in 2007 and about half of those repeaters retook the test within 30 days (Zhang, 2008).

The number of TOEFL repeaters has been increasing in the past decade due to growth in both the demand for the test and the ability of the businesses and institutions to increase the available supply. The demand for taking the TOEFL iBT test has been rising rapidly since more young people in developing countries are planning on entering universities in the USA and Canada than in the past. This is especially true in countries like China. In 2008 over 10% of the worldwide TOEFL population of one million were in China (Lin, 2008). The number of TOEFL iBT test-takers in China increased by 19% in the year 2011 and 32% in 2012, respectively (Yu et al., 2017). By 2012, China reached number one in the total number of TOEFL test takers in the world (Dai, 2016). In 2014, the number of TOEFL test takers in China represented about 20% of the TOEFL iBT test population (Liu, 2014).

To meet the growing demand for TOEFL, the supply of the TOEFL iBT test administrations has increased along with the steady expansion of internet-based language testing centers. According to the International Education Industry Research Report (2019), the number of TOEFL iBT test centers in China increased from 38 in 2013 to 47 in 2019. With adequate capacity to expand, testing companies have been continuously loosening the restrictions on retaking the test. For example, ETS has recently reduced the waiting time requirement before repeating the TOEFL test from 12 days to only three days and ETS no longer imposes restrictions on the total number of times a test taker can repeat the test. Seemingly, testing companies have been encouraging test takers to retake the tests. For example, some test takers in China report that they received emails (See Appendix) from ETS encouraging them to retake the TOEFL iBT test soon after they had received their latest test scores.

Since English language proficiency tests are the gatekeepers for international students entering English-speaking universities, obtaining a satisfactory English language proficiency test score is a major hurdle that young students must clear. Because of this, many test takers in China retake the TOEFL test more than twice before achieving test scores that they deem “good” for academic program applications and they often spend many months in the process of completing these multiple test attempts (Ma, 2017). When many candidates decide to repeat the TOEFL iBT test and can retake the same test as often as they want with few restrictions, issues regarding the validity of the interpretation of repeaters’ test scores arise. The validity argument for interpretation and use of the TOEFL iBT as outlined by Chapelle (2008) does not consider the fact that scores may have resulted from repeated test taking under current conditions. Moreover, factors contributing to the score gains of TOEFL repeaters are not well understood.

Gaining knowledge about TOEFL iBT repeaters' test preparation behaviors will benefit all stakeholders in this high-stakes language assessment. TOEFL test developers are concerned with whether the score increases of TOEFL iBT repeaters represent improvement in their language ability. TOEFL test users need to know if the improved scores are good indicators of university applicants' English ability to determine if applicants meet the language requirements for university admissions. TOEFL test takers care about effective test preparation strategies that can help them improve test performance. TOEFL iBT repeaters' test preparation behaviors are currently under investigated and the questions addressing validity issues associated with test preparation are rarely studied in the Second Language (L2) assessment field.

Studies on factors that contribute to repeaters' test-score improvement are also scarce. One major reason for the lack of research in this area is the limited availability of valid data on test repeaters and their preparation strategies. Understanding the effects of the TOEFL iBT repeaters' test preparation requires getting access to both quantitative and qualitative information on repeaters' test preparation behaviors between the different test occasions. However, such information is not readily available and is very difficult to obtain. Barkaoui (2017) stated in his review of TOEFL repeaters' test performance that: "Unfortunately, there is little to no research or data available on the numbers and characteristics of repeaters of English language proficiency tests" (p. 420). Additionally, information on students' behaviors outside the classrooms is also very difficult to collect. To better understand repeater behaviors, researchers need to find alternative ways to collect sizable qualitative and quantitative data that contains information about repeaters' language proficiency profiles and test preparation practices that can be linked to their test performance on different test occasions.

In China, such data linked to TOEFL iBT scores is made publicly available by the many Chinese TOEFL iBT test takers who write reports about their test preparation experience and share their information on TOEFL test discussion forums. A TOEFL discussion forum contains a large number of discussion threads. Each discussion thread is a collection of posts focused on a discussion topic initiated by the first post (also called the original post) of a thread. A TOEFL iBT test taker's test preparation experience report is typically the original post of a discussion thread and is often tagged as "TOEFL Test Tips" or "TOEFL Preparation Experience" by the forum administrators to facilitate searching. Research shows that many TOEFL test takers in China read test preparation reports and use some of the strategies recommended in these reports in their test preparation (Ma, 2017; Yu et al., 2017). Among these TOEFL iBT test preparation reports, about one quarter were written by TOEFL iBT repeaters. For example, of the 1,100 TOEFL iBT test preparation experience reports posted on the Chasedream TOEFL discussion forum (<http://forum.chasedream.com>) by August 2019, over 300 were repeater reports. The repeater reports contain information about test dates, the test takers' total scores and subscores of multiple test events, as well as detailed descriptions of the test preparation materials and preparation strategies used during their test preparation. In this study, I collected TOEFL iBT repeaters' test preparation reports from the Chasedream TOEFL discussion forum, extracted information about the test preparation strategies used by repeaters, and conducted qualitative and quantitative analyses to learn repeaters' test preparation behavior.

This thesis detailing the motivation, methods, and results of the analysis is composed of five chapters with the present introduction as Chapter 1. Chapter 2 reviews the TOEFL test preparation related literature. Chapter 3 describes the methodology used in this study. Chapter 4

presents the quantitative and qualitative findings of this study. Finally, Chapter 5 is the discussion of the findings, the implications, and limitations of the study.

CHAPTER 2. LITERATURE REVIEW

The present chapter gives an overview of previous research related to the topic of TOEFL test preparation and TOEFL repeaters. The validity issues of TOEFL repeaters' test performance are addressed in the first section. Next, the research on TOEFL test preparation is described. Finally, the research questions developed to guide this study are presented.

2.1 Validity Issues of Repeaters' Test Performance

The validity of a language proficiency test refers to the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores (Messick, 1989, p.6). The fundamental validity question about interpretation is whether the test scores reflect the test takers' language proficiency. It is essential to examine the extent to which test scores capture the learning effects that may differ by English language programs and English learners themselves (Chapelle, Enright, & Jamieson, 2008). The multiple test scores of TOEFL repeaters challenge the validity of the TOEFL iBT test. For TOEFL iBT test repeaters who obtained higher test scores in the repeated test, the critical issue is whether the increased test scores reflect an improvement in their English language proficiency. The validation process should therefore provide evidence that higher scores obtained by repeaters after their first scores actually reflect higher language proficiency as opposed to other factors.

A warrant is a statement which expresses a "rule for inferring claims of a certain kind from data of a certain kind" (Kane, 2013, p. 12). The generalization inference of the validity argument is based on the warrant that observed scores are estimates of the expected scores a test

taker would obtain on parallel versions of tasks and test forms and across administration and rating conditions (Chapelle, 2008). Evidence to support this inference should demonstrate that repeaters' multiple test scores obtained from different test occasions are generalizable and reliable. To address this issue, ETS researcher Zhang (2008) studied the performance of 12,385 test takers who repeated the TOEFL iBT within 30 days. The researcher used a 30-day time window for the data selection criteria based on the assumption that test-takers' L2 ability was not likely to change in such a short period of time and that the changes in scores would be mostly due to factors related to the test reliability. The findings show that changes in repeaters' test scores were very small (average 3.74 points of total scores) across different test occasions within the studied period, and the researcher concluded that the results support the generalization inference for the TOEFL iBT test. No other studies regarding the generalization inference concerning the repeating of the TOEFL iBT test were found in current literature.

The explanation inference of TOEFL iBT concerns the meaning of the TOEFL iBT test scores. Its warrant rests on the assumption that test scores vary depending on test takers' English language proficiency which was demonstrated in their test performance (Chapelle, 2008). With respect to repeaters' scores, studies supporting this inference would need to provide evidence that repeaters' score increases on test occasions after the first are mainly due to changes in the improvement of test-takers' L2 ability instead of construct-irrelevant factors (e.g., test preparation related to their previous test taking experience, or test-taking skills). Some construct-irrelevant factors lead to testwiseness, which is referred to by researchers as "using of one's knowledge of test formats and other peripheral information to answer test items" (Cohen, 2012, p. 97). Testwiseness has been measured in terms of the number of times the test was taken previously and the length of the interval between tests (e.g., Green, 2005; Wilson, 1987).

Testwiseness encompasses factors such as previous exposure to the test format and reduced anxiety (Messick & Jungeblut, 1981), memory of previous test item responses (Kulik, Kulik, & Bangert, 1984), and enhanced test-taking strategies (Sackett, Burris, & Ryan, 1989). Some researchers use the term ‘practice effects’ to describe testwiseness and suggest that the increase in test scores may attribute to the prior exposure to identical test or alternative forms of a test under the standardized conditions rather than true changes in the ability being measured (Cliffordson, 2004). Testwiseness should be seen as a concern when the TOEFL iBT is taken by repeaters who have had previous exposure to the test format and thus might have gained more testwiseness than first-time test takers. Questions concerning the repeater’s performance and the explanation inference remain largely unanswered or discussed in key literature in this field and no pertinent empirical studies are currently exploring such issues.

The extrapolation inference refers to predictions of how the expected score is to be interpreted as an indication of performance that the individual would receive in the target domain. It is based on the warrant that the construct of academic language proficiency, as assessed by the TOEFL iBT test, accounts for the quality of English language performance in the target domain, such as English-medium institutions of higher education (Chapelle, 2008). Backing for this inference is found in evidence that test scores can be related to other criteria of language proficiency in academic contexts. For example, evidence would show TOEFL iBT scores to distinguish as predicted among groups of students who had been placed into different levels of language classes after the admission decisions made by universities. Evidence for such distinctions has been found (Chapelle, 2008). Alternatively, using placement test scores rather than placements, Kokhan and Lin (2014) investigated the coefficients between repeaters’ multiple scores (the lowest, the most recent, the highest and self-reported scores) and the ESL

placement scores on newly enrolled international students by the University of Illinois at Urbana-Champaign. They found none of these scores had any predictive power to the placement test scores, although repeaters' self-reported highest scores had a relatively stronger association with the ESL placement results. Studies in this area are very scarce and many questions remained unanswered.

2.2 Research on TOEFL Test Preparation

Test preparation refers to “any intervention procedure specifically undertaken to improve test scores, whether by improving the skills measured by the test or by improving the skills for taking the test, or both” (Messick, 1982, p. 70). In the context of the language testing, test preparation is generally explored regarding the influence of testing on teaching and learning (e.g., Alderson & Hamp-Lyons, 1996; Elder & O’ Loughlin, 2003). Research on language test preparation courses has mainly revolved around the washback effect (e.g., Messick, 1996) and often involves test-related practices and curriculum that have been developed for language test candidates, test-specific training courses, and language learning and test-taking strategies used by test candidates (Green, 2007).

From the high-stakes language testing test takers' perspective, the major goal of the test preparation is to obtain a high score on the upcoming test. Test takers are willing to acquire any type of skills in test preparation that can help them achieve this goal, regardless of whether such skills focus on language learning or test taking. This is particularly true among Chinese test takers. China is a highly exam-oriented society and preparation for high-stakes tests is rooted in Chinese culture, even to the extent that test preparation is a nationally-accepted educational

phenomenon (Yu & Jin, 2014). Studies have found that many prospective Chinese students take language testing preparation courses for the purpose of getting scores necessary to meet language requirements for university admissions (Alderson & Hamp-Lyons, 1996). Furthermore, it is a common belief among test takers and test trainers in China that both language ability and test-taking skills determine the final score. Therefore, coaching in test preparation courses generally incorporates both language learning and test-taking strategies to help the candidate better prepare for the upcoming test and accordingly improve their test scores (Ma, 2017). Preparation training for high-stakes English tests has become a fast-growing and profitable industry in China since the early 1990s (Matoush & Fu, 2012). New Oriental Education (NYSE: EDU), for example, has reportedly trained 70% of Chinese students prior to their entering universities in the USA and Canada before 2010, with the majority of such students having taken TOEFL preparation courses (Tang, 2010).

Traditionally, TOEFL test preparation was predominantly conducted in the form of “coaching,” with test takers enrolling in test preparation courses where trainers coach them on test-relevant learning materials and test-taking skills. Studies on TOEFL test preparation courses often find limited effects on total score gains in the TOEFL tests. For example, Bachman, Davidson, Ryan, and Choi (1995) investigated the effect of a First Certificate in English (FCE) preparation course on both FCE and TOEFL tests and found that the preparation course had very little impact on the TOEFL scores. Ward and Xu (1994) investigated a 6-week training course on the summarization skills used in the TOEFL test and found that students had a 0.5 standard deviation score gain on TOEFL scores. Liu (2014) conducted an online survey of 14,593 Chinese test takers and found a correlation between attending training school and improvement in TOEFL iBT scores; however, researchers concluded that the contribution of 1.86 points to the

total test score was a very small effect. Ling, Powers, and Adler (2014) used TOEFL iBT practice test scores pre-test and post-test to compare the score improvements between two groups of students who received English language instructions and TOEFL iBT preparation courses in China and the United States respectively. The researchers found moderate score gains among students in China but substantial score gains among students in the USA, and the researchers believed that the score gains in their study captured students' learning effects that supported the validity argument for TOEFL iBT scores. The study also showed that students in the USA have greater score gains in writing and speaking, while students in China gained more in reading and listening scores. Yu et al. (2017) studied Chinese TOEFL iBT test takers who attended test preparation schools and also found a weak relationship between the test preparation training and the test performance.

With the development of modern technology and the wide adaptation of the new internet-based test format over the last decade, a question one may ask is whether taking courses in coaching schools is still the typical way to prepare for the TOEFL iBT. Chapelle (2008) argued that "as computer-assisted language assessment has become a reality, test takers have needed to reorient their test preparation practices to help them prepare for newest items" (p. 127). Indeed, along with the affordance of personal computers and internet access, test candidates' test preparation behaviors have greatly evolved, especially among the test takers in China. Anecdotal evidence suggests that in the past decade, more and more Chinese TOEFL iBT test takers use new technologies to help them prepare for the TOEFL iBT. For example, on a popular Chinese software portal pc6.com, there are 43 TOEFL iBT preparation software programs and mobile apps available for the public to download for free. The top TOEFL preparation mobile app *XiaoZhan TOEFL* has accumulated 10 million registered users since 2011 (zhan.com, 2020).

Most of these computer software programs and mobile apps have functions for multiple skills including practicing listening, speaking, reading, and writing. Some applications include new features such as multi-speed playback functions to facilitate intensive listening exercises and spaced learning models loaded with TOEFL vocabulary lists. Official and third-party TOEFL iBT simulation software is also widely available for candidates to practice TOEFL iBT simulation exams.

Apart from using software programs and mobile apps, test candidates' approach to test preparation has also changed dramatically. Many test takers now form online learning communities instead of attending coaching schools. A survey of Chinese students' TOEFL iBT test preparation experience found that, apart from getting preparation materials and exchanging test-taking experiences on TOEFL test-taking discussion forums, many test candidates form online learning communities within the discussion forums where participants share resources and motivate, coach, support and help each other (Ma, 2017). Common goals of learning, mutual support, and shared values and experiences bond members of learning communities (Jonassen, Peck & Welsion, 1999). Also, learning communities offer both instrumental help and emotional support to their members (Bruckman, 2006). Research finds that TOEFL test candidates use interactions in the online learning community as a way to reduce stress and to get encouragement from others, and many learning group participants motivate each other when their peer members express frustration or face difficulties (Ma, 2017).

Previous research on TOEFL test preparation focused more on coaching and curriculum and there were few studies on the candidate's test preparation behaviors such as the test taker's thought processes and their choice of test preparation strategies. Furthermore, there have been few studies focusing on analyzing these test-taking behaviors. In recent years, studies have

focused on test preparation strategies for test takers who prepare for the test both inside and outside of training schools. Building on the frameworks of previous research on test preparation strategies, Liu (2014) categorized TOEFL iBT test takers' test preparation strategies into two main types according to their different focuses: (a) general English learning strategies that aim to improve test takers' English ability as well as test performance, and (b) content-based test-specific preparation targeting the TOEFL iBT test. Based on a total of 14,584 email survey responses from Chinese TOEFL test takers, the researcher identified nine general English learning strategies (e.g., reading English books and watching English movies) and 19 test-specific preparation strategies (e.g., memorize TOEFL vocabulary lists) used by test takers. Using participants' Chinese College English Band-4 (CET4) exam scores as the control for their prior English proficiency level, the researcher ran multiple regression analyses to investigate the relationships between 28 individual test preparation strategies and the TOEFL iBT test scores. The finding was that writing emails, letters, and diaries; practicing speaking at English Salons; and reading books and magazines were positive predictors of both the total scores and subscores in the general English learning strategies category. For the content-based test-specific strategies category, although some strategies were found to be effective, they were identified as being domain-specific, meaning they were only positively correlated to each subskill section. With the exception of speaking, only vocabulary memorization and TOEFL Practice Online (TPO) appeared to be effective in predicting both the total score and subscores. Although several of these two types of strategies were statistically significant in predicting improved test scores, the researcher concluded that the increase in score points was very small. Regarding the test-taker status, the samples in this study contained both first-time test takers and repeaters; however, the repeaters' specific test preparation strategies were not separated from those of the first-time test

takers. Furthermore, the relationships between repeaters' test preparation strategies and their test score improvements were not investigated.

In another recent study on the TOEFL test preparation strategy, Yu et al. (2017) investigated Chinese test candidates' test preparation strategies and their TOEFL iBT speaking subscore gains. The researchers collected 1,514 questionnaires from both teachers and test takers, with questions covering speaking task activities as well as materials, courses, teaching, and learning strategies used in the test preparation for the speaking tasks. The researchers listed 17 test preparation activities in the questionnaire and asked the participants to answer how often they used each activity during test preparation based on a 5-point scale (5 points for very often and 1 point for never). Using the frequency points of the 17 test preparation activities as values of the independent variables and the test performance as dependent variable in the simple regression analyses, the researchers found that only the TOEFL Practice Online (TPO) activity seemed to be a strong predictor for students' TOEFL iBT speaking test scores. The researchers concluded that, although both teachers and learners have a high agreement in terms of the usefulness of test preparation activities, the results showed only a weak relationship between test preparation and test performance. However, the researchers acknowledged the limitation of missing the multiple regression analysis due to a lack of score improvement information from participants in this study.

The above mentioned two studies investigated test preparation strategies used by TOEFL test takers in general but did not distinguish first-time test takers from repeaters. One recent survey study on the Chinese high-stakes English language test preparation phenomena included a small subset of TOEFL iBT repeaters (Ma, 2017). Through classroom observations and interviews with 18 test takers and eight teachers, the researcher found that repeaters focused

more on improving English skills than first-time test takers. They hypothesized that the previous failed test attempts made repeaters redirect their focus to language learning and such behavior might contribute to their increased test scores later on. However, the author did not provide empirical evidence on the relationship between repeaters' test preparation strategies and score gains, nor did the author explore how repeaters used preparation strategies during their test preparation.

2.3 Research Questions

The lack of studies on language test repeaters' test preparation is a notable research gap. Barkaoui (2017) in his review of repeaters' performance on second language testing calls for "more research on the attributes and performance of repeaters of L2 proficiency test to inform the validity arguments to these tests, test score users, and test validation and L2 acquisition research in general" (p. 420). This study is a response to this call for research.

This study aims to answer the following research questions about the test preparation behaviors of TOEFL iBT repeaters in China:

- 1, What kind of test-taking patterns do Chinese TOEFL iBT repeaters have?
2. What test preparation strategies did TOEFL iBT repeaters in China use and how were the strategies implemented during test preparation?
3. What is the relationship between the test preparation strategies and TOEFL iBT test score improvement?

CHAPTER 3. METHODS

The methods used in this study were designed to overcome the limitations of previous research by examining the descriptions of TOEFL preparation strategies and test scores volunteered by participants in an online discussion forum on TOEFL test taking. Quantitative and qualitative approaches were taken in the analysis of information and test scores reported by the TOEFL iBT repeaters to address the stated research questions regarding test preparation strategies and the relationship between repeaters' test preparation strategies and test performance in the repeated tests. This chapter describes the participants, data description, procedures and the data analysis.

3.1 Participants

The sample in this study includes 170 individual TOEFL iBT repeaters who took their repeated TOEFL iBT tests between July 1, 2006 and June 30, 2019 and posted their test preparation experience reports on the TOEFL iBT discussion forum of the Chasedream website (<http://forum.chasedream.com>). All the discussion messages posted on the Chasedream forum are publicly accessible and free to download by forum visitors. However, only registered forum members can upload materials and publish posts on the forum. The forum membership registration is free for anyone with a valid email address.

Chinese TOEFL test takers who received satisfactory scores often voluntarily write and post their test preparation experience on TOEFL topic related online forums. The TOEFL discussion forum on Chasedream had over 1,100 test preparation experience reports posted by TOEFL iBT test takers from the beginning of 2003 till the end of 2019. Among all the test

preparation experience reports on the Chasedream forum, 32% were written by TOEFL iBT repeaters. The percentage of reports written by repeaters was similar to the previous research that found 28% to 40% of all TOEFL test takers were repeaters (Wilson 1987 and Liu 2014).

However, candidates who posted their test preparation experience on the internet might have different characteristics from those who did not. For example, TOEFL repeaters with improved test scores might be more willing to report their test preparation experience than repeaters who did not improve their test performance. Therefore, the participants in this study are likely be more representative of those TOEFL repeaters in China who made satisfactory improvements in test performance as opposed to the general TOEFL repeater population. However, given that the goal of this study is to investigate test preparation strategies and their relationships with test score improvements, the samples selected in this study may be good representatives of the target of interests, i.e., the TOEFL iBT repeaters who successfully improved their test scores.

The demographic information shows that among the 170 TOEFL iBT repeaters, 82 took the test only twice, 88 took the test more than two times, and only nine repeaters attended coaching schools during their test preparation for the repeated tests. Most repeater reports in this study contain other background information such as testing dates, scores of multiple tests, and information about coaching school attendance. However, gender and educational background information was rarely disclosed in the repeater reports.

3.2 Data Description

Data Source

This project collected data from the TOEFL test discussion forum on the Chasedream website. Chasedream is a study-abroad portal that serves as a free information hub for Chinese students to find information about overseas university programs, English proficiency tests, and

graduate admission tests. The portal runs its business by charging advertisement fees from clients such as universities, study-abroad service agencies, travel agencies, law firms, and language training schools (Advertising Service, 2004). The TOEFL test discussion forum in Chasedream is famous for its large number of TOEFL test-related discussion threads. From the portal's launching in 2003 until the end of 2019, its TOEFL discussion forum had accumulated over 55,000 TOEFL test-related discussion threads containing 1.08 million in-thread posts. These TOEFL test preparation discussion threads were grouped in the subforum tagged "TOEFL Preparation Tips." There were over 1,100 TOEFL preparation discussion threads on the forum. Among them, 350 were written by TOEFL repeaters. The TOEFL repeater reports on Chasedream usually attract a lot of attention from TOEFL test candidates. For example, among the top 30 most viewed test preparation reports on the forum, 21 (70%) were written by TOEFL iBT repeaters. The comments in the follow-up posts to these repeater reports showed that both first-time test takers and repeaters alike read and responded to the repeater reports. For example, a repeater report titled "The Journey from Scored 55 to 100 -Where there's a will, there is a way," posted in 2012, had attracted over 150,000 views and 2,400 responses from both first-time test takers and repeaters.

The repeater reports on Chasedream have an average length of 2,400 Chinese words. All reports were retrieved from <https://forum.chasedream.com> which were identifiable by their forum member ID. I have translated all example repeater report quotations from Chinese into English. Most of the TOEFL iBT repeater reports on Chasedream are written in an essay style with a similar writing structure. A typical repeater report begins with the dates, total scores, and all four subscores (reading, listening, speaking, and writing) of the repeater's most recent test event, as well as information on all previous test attempts. Many repeaters explicitly explained

the purpose of writing such a report in the introduction section. Example (1) shows the introduction section of a repeater report titled “From 72 to 120. Listening from 18 to 29. Old Student Made a Leap in 17 days.”

- (1) *Date 4/20: Reading 18, Listening 18 Speaking 15 Writing 21 Total 72;
Date 5/18: Reading 27, Listening 29 Speaking 20 Writing 24 Total 100.*

Finally, after a two-day delay, the scores are released at 7:00 am today. I did not have a good sleep for three days. The moment I saw my score report, I almost burst into tears. 100, not a high score, but it was a big leap for me. From the preparation of my GMAT test to my TOEFL test, all my test preparation materials, all the test preparation strategies of former test-takers, came from Chasedream, my dearest, most selfless Chasedreamers had given me so much help. Today, let me repay all of you with something to help you win (Sunsissy, 2013).

After the introduction section, most repeaters begin to describe the strategies they used during their preparation for the most recent test. All test preparation strategies are introduced in the sequence of subskills: reading, listening, speaking, and writing. This follows the same sequence as the sections in the TOEFL iBT test. Usually, the authors elaborate more on the section where they performed well or made the most score gains. The following paragraph (2) was the listening subskill section written by a repeater who scored 29 on the listening section in the repeated test:

- (2) *Summary of listening preparation steps: 1), Complete one volume of TPO, check your answers. 2), Identify 2-3 paragraphs that you made the most mistakes or contained parts that you did not understand, listen sentence by sentence at 120% of normal speed, or the fastest speed that you can handle. If you felt lost after several runs, then reduce the speed*

and listen repeatedly. Anyway, the key is repetition. This might be very boring, but sometimes learning is a physical exercise, you have to keep going. If you can, you are deserved to get higher scores than others. 3), Use the same 2-3 paragraphs, read the text, find out the question patterns, analyze the structure of the transcript, and learn the general structure of the listening texts. 4), Repeat the above steps, and do more TPO volume. Among the 20 TPO volumes I completed (volume 10-30), 10 of them were practices with the above steps (sherry, 2013).

Authenticity, Truthfulness, and Accuracy of Reports

The repeater reports in this study may have relatively a high degree of authenticity for several reasons. First, there are no incentives for post owners to make false claims about the test preparation strategies they used. According to the User Agreement of Chasedream, members cannot conduct any business through the forum and posts were not allowed to contain any promotions or advertisements. So, there was no financial gain for report authors. Second, the report authors have no intentions to attract viewers by boosting the test scores in their reports. All forum members are anonymous, and the respond posts in the threads show that many report authors do not revisit the thread after posting their TOEFL test preparation reports. The following example (3) showed that the author of a repeater report did not revisited his or her test preparation thread until three years later:

- (3) *PS: Three years have passed, today, during the coffee break at work in the afternoon, I suddenly thought about the post I wrote three years ago. I haven't visited Chasedream for 2 years. To my surprise, I saw over 1,000 responses in the thread. I was so excited and, at the same time, extremely happed that I was able to offer some encouragement to*

my “sisters and brothers in arms”, who are still fighting at the front line of the going-abroad battles. Your thread owner (me) is working at KPMG at Washington DC now. I have grown from a naïve student to someone who achieves his dream through years of struggles; only myself know how hard it had been and what it took to get here. As time goes by, we all have higher goals to reach. Unfortunately, I cannot offer further help to you, my brothers and sisters, on your preparation of the TOEFL or GRE/GMAT tests. After soon many years, I had forgotten how. But, all TOEFL and GRE/GMAT test-takers are family. If any of you Chasedream members come to study in DC, contact me by my WeChat ID XXX. Studying abroad is very challenging. I hope I can continue to provide more help to my dear brothers and sisters (Ikangning, 2014).

What supports the truthfulness of the information in these reports is the motivation for writing the test preparation reports. Most authors clearly stated that the purpose of composing the report was to repay the help they received from other forum members by helping others. For example, “to repay all of you with something to help you win” (Ex.1) and “offer some encouragement” (Ex.3). The popular “wanting by helping” mentality (Converse, Risen & Carter, 2012) among young people in China motivates many Chinese TOEFL test takers to provide truthful information in their test preparation reports. Many test takers in China believe that sharing test-taking tips is a good deed that will bring luck or good karma to their own journey of pursuing personal goals (Xie, 2010; Ding, 2017). Among the 170 repeaters in my sample, 30% explicitly expressed that they hoped their reports would “bring back good luck” if the strategies they provided in the reports could help others. Some repeaters even wrote such an intention in their post titles. For example, a repeater report titled “First Battle 95, Second Battle 108. Share My Experience to Help You and Bring Me Luck” was tagged as one of the most liked repeater posts in the forum. Although we have no way to confirm the authenticity of the information in the reports, it is

reasonable to assume such motivations support the authenticity of the information in the reports. The strong motivation to help others also drives the report authors to provide detailed information at their best, as we see in the above example (2) with elaborated descriptions of the preparation strategies for listening.

The scores reported in the posts were very likely to be accurate because upon receiving a satisfactory score in the repeated test, repeaters were often very excited and eager to share their experience immediately. The time tags of when the reports were posted show that most of the repeater reports were composed and posted on the same day of receiving the test scores, as shown in the example (1) above. Since the time of composing the reports and the time of taking the test were very close (within a couple of weeks), the recall of the strategies used during test preparation was still fresh with a likelihood of minimal false memories.

3.3 Procedures

Data Collection

I used Google Advanced Search Operators (SEO) to search for TOEFL iBT repeater posts in the target domain URL. The advanced search commands contained the following operators: *-site*, *-posttitle*, *-intext*, *-AND*, *-OR*, *-inurl*. For example, a within-title keyword search command: *"-posttitle: KEYWORDS site:chasedream.com"* limited the search in the Chasedream forum post titles that contained specified keywords; A within-post-text keyword search command: *"-intext: KEYWORDS site:chasedream.com"* restricted the search in the same URL but expanded the search into the post content containing the target keywords.

All the major content of repeater posts is written in simplified Chinese, although some TOEFL test-related terms and test questions are written in English. Some post titles also contain English words. However, the keywords referring to testing and the number of tests taken are all in Chinese. The target Chinese keywords used in searching for the repeater reports were 托福, 一战, 二战, 三战, 终战, 逆袭, 从...到...(TOEFL, *First Test*, *Second Test*, *Third Test*, *Final Test*, *Underdog Counterattack*, *Score from...to...*). The Chinese words “一战, 二战, 三战, 终战” which literally translate as “First Battle, Second Battle, Third Battle, and Final Battle,” are consistently used by all repeaters to refer the number of times they took the test. Each repeater post identified by the above-mentioned searching method was downloaded and saved as a txt file. Each txt file was a repeater report written by an individual TOEFL iBT repeater. In total, I downloaded 227 TOEFL iBT repeaters reports. To exam the relationship between test preparation strategies and test performance improvement, a repeater report must contain information of the test preparation strategies used between the test intervals and the score changes. Therefore, repeater reports missing information reflecting score changes were excluded. For example, I excluded those repeater reports that only reported the recent total test score but no previous test scores. Then I checked forum member IDs, testing dates, and test scores to exclude any duplicates or multiple reports written by the same author. The final sample contained 170 observations of TOEFL iBT test preparation reports written by 170 different test repeaters who took their final TOEFL iBT tests between January 2006 and August 2019. Among the 170 repeater reports, 82 came from repeaters who took the test two times and 88 came from repeaters who took the test for more than two times.

Strategy Classification

Following the strategy classification method in the Liu (2014) study, I classified test preparation strategies into the following two types, based on their different cognitive nature and skill improvement focuses (i.e., whether the strategy was used mostly to improve language skills or testing skills): (a) English learning strategies, and (b) test-specific strategies. For individual strategies under the two strategy types, I drew on the same definition principle used by Liu (2014), which was to define individual strategies based on specific test preparation activities or materials. For example, ‘read a variety of English books to improve overall English ability’ was defined as an *Extensive reading* individual strategy under the English learning strategy type, and ‘practice TOEFL TPO reading materials’ was defined as *Practice TPO reading strategy* under the test-specific strategy type.

However, my classification process for the individual strategies was different from those used in the Liu (2014) study. Liu (2014) developed a test preparation strategy list based on reviews of previous literature and popular materials used by coaching schools and then conducted a survey of test takers to find out if those individual strategies were used or not used by the yes or no response to the close-ended strategy list. In my study, when I identified a new individual strategy in a repeater report, I recorded it in a strategy coding book with a name, the definition, and an example of the strategy. Therefore, after reading all the repeater reports, all the test preparation strategies mentioned in the reports were recorded in the coding book. Table 1 shows an excerpt from the coding book.

Table 1

Excerpt of the individual test preparation strategy coding book

Strategy	Type	Definition	Example
Extensive Reading	English Learning	Read variety of books or magazines	<i>"To improve my reading skill, I sometimes read the Economist magazines..." (yanghaijian, 2014)</i>
Intensive Reading	English Learning	Read limited number of articles word by word	<i>"After reading a TPO article, do the intensive reading. Read word by word, write down all unfamiliar words ..." (wenzi5811, 2015)</i>
Speed Listening	English Learning	Playback listening materials at accelerated speed	<i>"Most importantly, I practiced all listening materials by listening at 1.3 times of the normal speed..." (1405, 2014)</i>
Note-taking for Listening	Test-taking	Practice note-taking skills for listening tasks	<i>"You need to practice how to use symbols and arrows to help you take notes while listening..." (cufe2004, 2010)</i>

The individual test preparation strategies are treated as categorical variables to represent the test preparation strategies as used or not used. Therefore, I recorded a value of one if a strategy was used by a repeater and a value of zero if a strategy was not used. The different kinds of individual test preparation strategies were described as (a) used, (b) recommended, (c) disapproved, or (d) not mentioned in the repeater reports. Since the repeaters did not indicate any difference in the frequency of use or perceived usefulness between strategies "used" and "recommended," I assigned the value of 1 on both "used" and "recommended" strategies. I assigned the value of 0 to the "disapproved" and "not mentioned" strategies to indicate that they were not used by repeaters. On an Excel worksheet in which each row represents a separate repeater record, I created a new column for each of the newly identified strategies as the independent variable and recorded the value of one in the column on the row of the individual repeater. If a strategy identified in a repeater report already existed in the coding book and in a worksheet column, I recorded the value of one under the existing column on the row of the new repeater. In the end, all columns representing the different strategies used by all the repeaters

were recorded on the worksheet. In total, I have identified and recorded 13 English language learning strategies and 21 TOEFL test-specific strategies.

To ensure the reliability of coding and the classifications of test preparation strategies in the repeater reports, 10 % of the data (17 reports) were also coded by a second coder. The second coder is a native Chinese speaker who holds a bachelor's degree in English and a Ph.D. degree in Business. Before coding the 17 reports, the second coder was trained to code using one report (not included in the 17 reports) and instructed to use the test preparation strategy coding book as the guideline. The inter-coder agreement was 95% (616 agreed cases out of 646 "used" and "not used" individual strategies). Cohen's Kappa (0.87) showed both coders were in near-perfect agreement about the individual strategies "used" or "not used" by repeaters.

3.4 Data Analysis

To address my Research Question 1 on TOEFL iBT repeaters' test-taking patterns, I conducted statistical analysis to investigate the repeaters' target test scores and score changes over test occasions, the number of tests taken, the waiting period between test occasions, and the time spent in preparation for the repeated test.

To address my Research Question 2, regarding what test preparation strategies were used by repeaters and how the repeaters implemented these strategies, I calculated descriptive statistics on language learning and test-specific strategies. Then, I conducted qualitative analyses to evaluate repeaters' test preparation strategies and investigated how some of the individual strategies were used by repeaters.

To address my Research Question 3 on the relationship between test preparation strategies and test score improvement, I conducted five simple and multiple regression analyses.

The difference between a repeater's first TOEFL iBT test score and the final TOEFL iBT test score (i.e., the score change) was the outcome variable in the regression analyses. Among all the 170 participants in this study, 166 reported their total score changes between their first and last tests, 131 repeaters reported the reading subscore changes, 136 reported the listening subscore changes, 132 reported the speaking subscore changes, and 133 reported the writing subscore changes. All missing data were treated with pairwise deletion.

The 34 individual test preparation strategies used to improve the specific sectional skillset were predictors in the analyses for sectional score changes. As mentioned in the previous section, each individual strategy is a categorical variable with the value of "1" if it was used by a repeater, and value of "0" if it was not used during the preparation for each of the subskill section.

When investigating the relationship between strategy use and the total score changes, I constructed seven sets of aggregated test preparation variables which involved using the same type of practicing activity or curriculum to improve skills for more than one skillset section. For example, *Intensive reading* and *Intensive listening* strategies both use an intensive learning activity which requires carefully studying a limited number of English texts, word by word, for many times. Therefore, an *Intensive learning* strategy was constructed as an independent variable to represent either *Intensive reading* or *Intensive listening* activity. The values of such aggregated strategy variables were calculated based on the number of sections a repeater used such strategies. For example, *Practice TPO materials* is an aggregated strategy variable. If a repeater practiced TPO materials in preparation for all four sectional skills, then for the analysis of the total score changes, the variable *Practice TPO materials* was included as an independent variable with a value of "4" in the simple and multiple regression models for the total TOEFL scores. The

range of this variable, thus, was from 0 to 4, indicating a repeater used this individual strategy in between none of the subskill section and all four of the subskill sections. Other aggregated strategy variables were: *Memorize TOEFL vocabulary* (reading and/or listening vocabulary lists), *Get feedback for output* (for speaking and/or writing samples), *Shadow speaking* (to improve listening and/or speaking), *Practice Ji-jing items* (used in any number or all of the four subskill sections), and *Taking the GMAT or GRE exams*. The *Number of times taken the test* and *Attending coaching schools* were also included as a control variable in the regression analyses of the total score changes. Correlation analysis between aggregated strategies in relation to the total score changes was also conducted.

CHAPTER 4. RESULTS

In this chapter, descriptive statistics are reported to examine repeaters' test-taking patterns, as well as to provide an overview of the test preparation strategies used by repeaters. Afterwards, results of the qualitative analyses of newly identified test preparation strategies are presented, including how repeaters used the test preparation curriculum and how repeaters perceived the test preparation strategies. Finally, the results on the relationship between test preparation strategies and score improvements are reported.

4.1 Descriptive Statistics of Test Preparation Patterns and Strategies

Test-taking Patterns

Overall, the TOEFL iBT repeaters in this study improved their test scores across all four skill sets being tested. Table 2 shows the descriptive statistics on repeaters' test performance in their first and last TOEFL iBT test occasions, including changes in the total scores and sectional subscores.

Table 2

Descriptive Statistics of Repeater Test Performance

Testing occasions and scores	Number of observations	Mean	SD.	Min	Max
First test total score	163	89.05	12.10	38	108
Last test total score	163	105.55	5.33	83	117
Total score change	166	16.40	9.97	-4	60
First test reading subscore	134	24.08	4.47	8	30
Last test reading subscore	158	28.08	0.15	22	30
Reading score change	131	3.87	4.29	-6	19
First test listening subscore	140	21.24	0.42	3	30

Table 2 (continued)

Last test listening subscore	157	27.17	0.18	20	30
Listening score change	136	5.93	4.60	-5	25
First test speaking subscore	132	20.58	0.25	8	30
Last test speaking subscore	158	23.24	0.18	17	30
Speaking score change	132	2.52	2.71	-4	14
First test writing subscore	134	23.30	0.29	13	30
Last test writing subscore	158	27.06	0.16	21	30
Writing score change	133	3.76	3.02	-4	13

One hundred sixty-three of the 170 repeaters reported the test scores of both their first and last tests. The repeaters in this study had an average score of 89 on their first test; they outperformed the population of TOEFL test takers in China, whose average score reported in 2017 was 79 (ETS, 2017). This suggests that either repeaters might have a relatively higher average English level than test takers in general, or repeaters who reported their test preparation experience had better-than-average English levels. The majority of these repeaters improved their test scores by an average 16.4 points. Only one repeater reported a negative total score change. The mean sectional score gains indicated that repeaters improved most on listening (5.93 points), followed by reading (3.87 points), writing (3.76 points), and speaking (2.52 points). The results are consistent with previous studies (e.g., Zhang, 2008; Ling, Powers & Adler, 2014) that find test takers obtained more score gains in the reading and listening sections than in the speaking and writing sections, although the writing and reading score gains in this study were very close.

The distribution of the first TOEFL iBT test scores (See Figure 1) shows that repeaters' first-time scores were clustered between 80 and 99. Fewer test takers retook the test when they achieved a score of 100, and no test taker retook the test when they reached the score of 110

during their first attempts. This score pattern suggests that many repeaters used 100 as a major score milestone and others chose 110 as their goal.

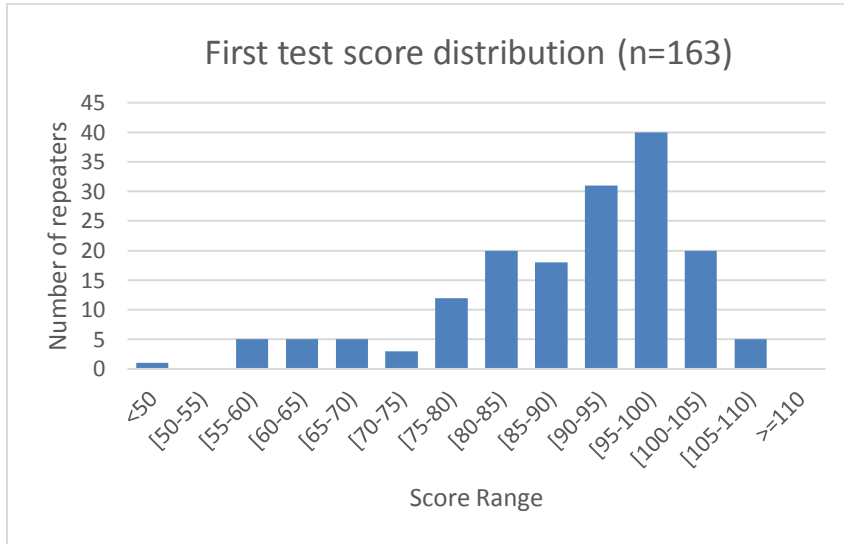


Figure 1 First test score frequency distribution

The total test score distribution of the last test (See Figure 2) suggest that repeaters were satisfied with their test scores if they reached the 100 or 110 score marks and were not likely to take the test again. Many repeaters indicated in their reports that they repeated the test because their previous scores were below 100.

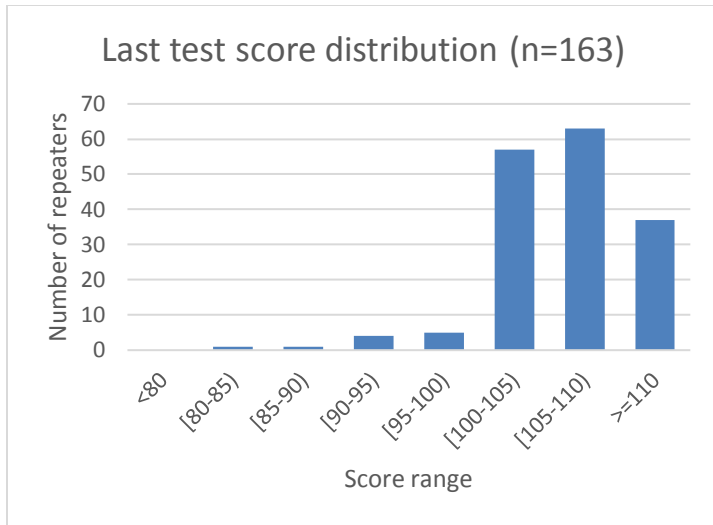


Figure 2 Last test score frequency distribution.

All the 170 repeaters in this study reported their number of times taking the TOEFL iBT test. Figure 3 shows the distribution of the number times the repeaters had taken the test. On average, repeaters took the test 3.11 times, ranging from two to 15 times. About half of the repeaters in this study took the TOEFL iBT tests only two times.

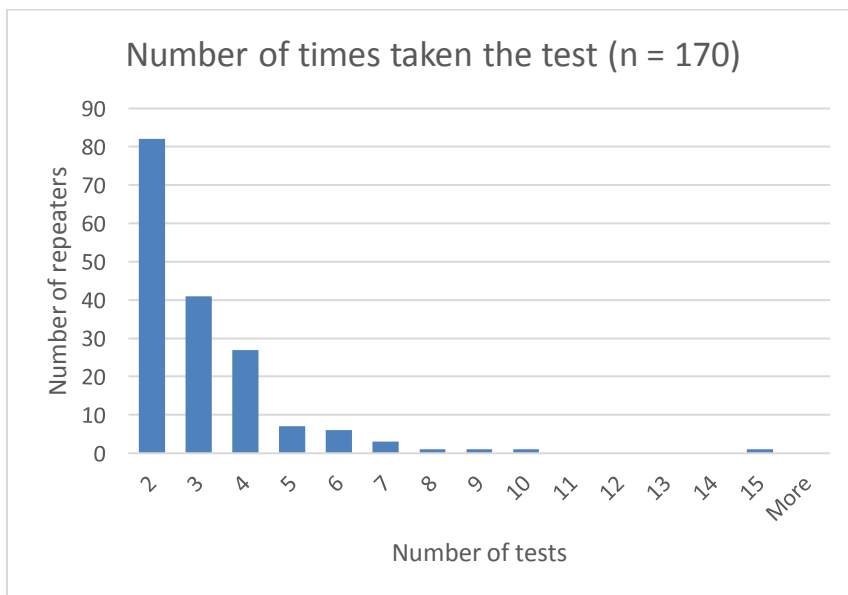


Figure 3 Number of times taken the test by repeaters.

Most repeaters reported the dates of their first and last tests, but many of those who took the test more than twice did not indicate the testing dates between the first and the last tests. Therefore, the waiting period between test events in this study was calculated by using the gaps between tests of the 66 two-time repeaters (excluding those who indicated that they repeated the tests due to the expiration of their first test scores). The information in the 66 two-time repeater report shows that the average waiting period between tests was four months, with 58% of the repeaters retaking the test within three months after their first test (See Figure 4). There may be multiple factors influencing the length of the intervals between tests, as suggested by Gu et al. (2015); however, investigating such factors is beyond the scope of this study.

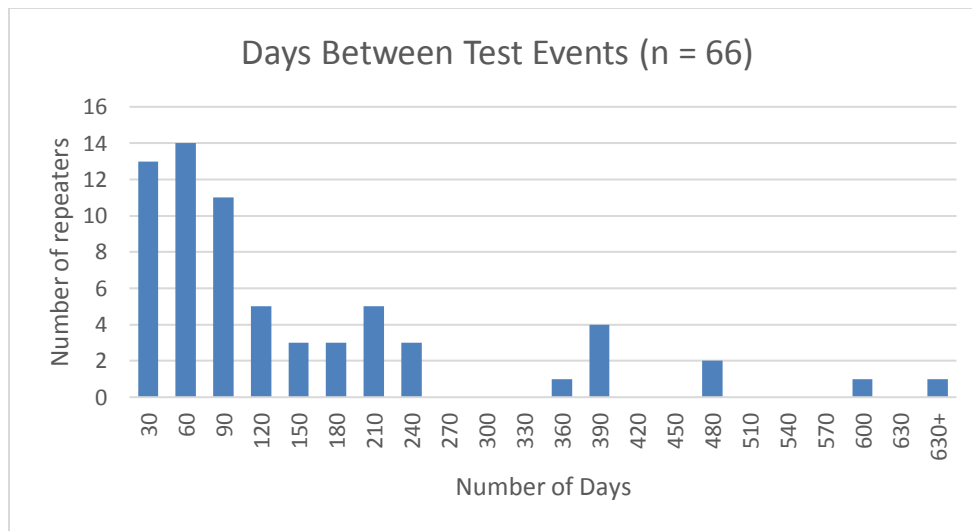


Figure 4 Time between two test events

Sixty-four out of the 77 two-time repeaters reported the number of days dedicated to test preparation. Most of these two-time repeaters spent between one and three weeks on test preparation, with an average of 12.6 days (See Figure 5). The average preparation time by

repeaters in this sample appears to be much shorter compared to the candidates (mostly first-time test-takers) in the Liu (2014) study.

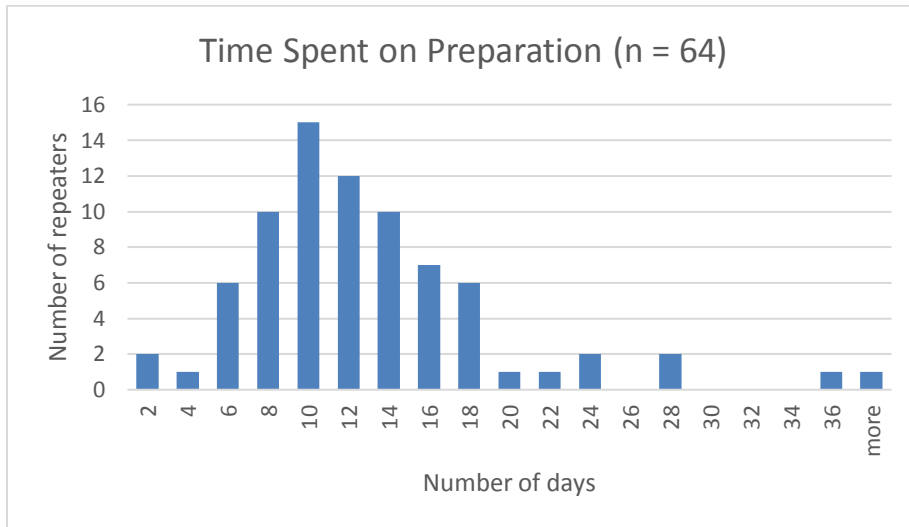


Figure 5 Time spent on preparation practices by repeaters

Test Preparation Strategies

The two types of test preparation strategies used by TOEFL iBT repeaters are shown in Table 3 and Table 4. In total, 13 language learning type individual strategies and 21 TOEFL test-specific type individual strategies used by repeaters during their test preparation were identified. The number/percentage of repeaters using these strategies during their test preparation are also reported. For example, of the 170 participants, 59 reported using an *Intensive listening* strategy, which means that this strategy was reported by 34.7 percent of participants. Table 3 lists all individual strategies in the language learning type and the number/percentage of repeaters who used such strategies. Table 4 lists all the test-specific types of individual strategies and the number/percentage of repeaters who used such strategies.

Table 3*Language learning strategies (n=170)*

Individual Strategy Description	N	%
Intensive listening (listen to limited number of articles word by word many times)	59	34.70%
Take GRE or GMAT exams (as a way to improve vocabulary and reading skills)	58	34.10%
Extensive listening (listen to English programs such as CNN News, TED Talks)	56	32.90%
Intensive reading (read limited number of articles word by word several times)	45	26.50%
Dictation activities to improve listening skills	43	25.30%
Speed listening (play English audio at accelerated speed. e.g., 1.3x of normal speed)	23	13.50%
Get feedback for writing (e.g., seeking comments from peers in a learning group)	23	13.50%
Record and replay own speaking for self-assess and correction	22	12.90%
Shadow speaking to practice listening (repeat after English audios)	20	11.80%
Shadow speaking to practice speaking (repeat after English audios)	12	7.10%
Practice pronunciation and intonation, imitate American accent for speaking tasks	11	6.50%
Extensive reading (Casual reading variety of English books, magazines e.g., Times)	8	4.70%
Get feedback for speaking (e.g., upload speaking sample online to seek comments)	4	2.40%

Table 4*TOEFL Test-specific strategies (n=170)*

Individual Strategy Description	N	%
Practice TPO reading materials	109	64.10%
Practice TPO listening materials	106	62.40%
Memorize TOEFL reading vocabulary list	78	45.90%
Practice speaking by using Ji-jing speaking items	59	34.70%
Write and memorize own examples for writing tasks	46	27.10%
Practice notes-taking skill for listening tasks (e.g., create short-hand, use keywords)	44	25.90%
Practice TPO speaking materials	40	23.50%
Practice writing topics by using Ji-jing writing items	37	21.80%
Use writing template written by others (e.g., templates found online)	33	19.40%
Practice TPO writing materials	32	18.80%
Write and memorize own writing template for writing tasks	31	18.20%
Memorize phrases and sentences used for writing tasks (e.g., "by the same token")	29	17.10%
Write and memorize own examples for speaking tasks	25	14.70%
Memorize TOEFL listening vocabulary list for listening	24	14.10%
Use speaking templates written by others for speaking tasks	22	12.90%
Write and memorize own spoken template for speaking tasks	21	12.40%
Use speaking examples written by others (e.g., examples found in books).	16	9.40%
Practice typing to improve typing speed for writing tasks in the test	16	9.40%

Table 4 (continued)

Practice Ji-jing items for listening	10	5.90%
Use writing examples written by others (e.g., found online)	10	5.90%
Practice reading by using Ji-jing reading items	7	4.10%

4.2 Qualitative Analysis

As mentioned in Section 3.2, the average length of the repeater reports in this study is about 2,400 words not written in response to any predetermined questions. Such detailed repeater reports revealed some information about TOEFL iBT test preparation that was not found in current literature, including some new test preparation strategies, how repeaters used the TOEFL test preparation curriculum, and the repeater's perceptions about the usefulness of various test preparation strategies.

Newly Identified Strategies

Several individual test preparation strategies widely used by repeaters, for example, *Speed listening*, *Dictation*, *Shadow speaking*, and *Get feedback for speaking and writing*, were not reported in current literature (e.g., Liu, 2014; Yu et al., 2017). Three newly identified strategies: *Speed listening*, *Dictation*, and *Shadow speaking*, were used for improving listening skills by repeaters. Some repeaters also used *Shadow speaking* to improve their speaking skills. The *Dictation* practice is conducted by playing English audios sentence by sentence, with the listeners stopping the audio after hearing each sentence and writing down what they heard. *Shadow speaking* is conducted by playing the English audio continuously while the listeners try to follow and imitate the sound they just heard. Many repeaters indicated that they practiced the *Dictation* and *Shadow speaking*, but these two strategies were simply mentioned in the repeater

reports without detailed descriptions of the practicing procedures, possibly because they were conventional learning approaches already widely used by many English learners in China. The *Speed listening* strategy, on the other hand, was strongly recommended and introduced with detailed instructions by many repeaters, especially those who substantially improved their listening scores. For example, a repeater who improved her listening score from 19 to 29 wrote the following (4):

(4) *Use speed listening. Save all your audio files on your mobile phone and listen to them many times. Start with 1.2 times the normal speed, then increase to 1.4 times the normal speed. If you can hear every word clearly at the 1.4-time speed, then in the real test, you will feel all the listening articles are played like slow motions in a movie, and you will be very close to getting the full score (ClaireZJY, 2019).*

Another repeater who improve their listening scores by 11 points gave the following suggestions (5):

(5) *My listening scores: 19-22-29-30. I suggest listening to a lot of Sixty Seconds Science Podcasts at 1.2 or 1.5 times the normal speed. If possible, you can follow up with some shadow speaking exercises, which I found to be very effective; Also, when you listen to the TPO materials, try not to use the normal speed. You should use either 1.2 or 1.5 times the normal speed, depending on your current listening ability. Believe me. If you use the normal speed during test preparation, you will cry in the real test (ccmianmian, 2018).*

The *Speed listening* strategy might be an effective way to improve the listening skills in a way that is more suitable for many Chinese TOEFL test-takers. Intuitively, playing English audio at a faster speed may help improve listening skills only if the listener understands the text but struggles with decoding the spoken form. This is exactly the case for many TOEFL test takers in China. Research shows that most Chinese students had difficulties understanding normal speed spoken English even after over 15 years of learning English and being able to read English (Li, 2002). Training to listen to high-speed English audio during the test preparation might help Chinese TOEFL test takers in overcoming such difficulties.

Another newly identified test preparation strategy widely used by repeaters in this study was *Get feedback for writing*. Interestingly, the feedback that the repeaters got during test preparation was mainly from peers in their online learning group instead of English teachers or TOEFL trainers. Online learning groups usually are formed among forum users who have a similar learning goal. These online learning groups usually run for a couple of weeks long with a maximum of 20 people in each group. Typically, one test taker initiates a learning group by posting a specific learning plan. For example, a “writing group for two weeks” would be an example of a TOEFL learning group with a focus on improving writing skills over a 14-day period. Forum users who have similar incoming test dates and the need to improve the proposed skillset voluntarily join the group by replying to the initiator’s post. By joining the online learning groups, test candidates turn their individual test preparation into a cooperative study process with peer support. For example, members of the learning group routinely post messages with a label “punch my card today” on their group message board and use the peer pressure to fight procrastination. The major group activities are: sharing learning materials, answering questions of practice test items, and, most importantly, editing each other’s spoken or writing

samples. A repeater commented (6) how the online learning group helped him/her improve the writing subscore from 22 to 27:

(6) Writing: I really appreciated the Chasedream forum because all the essays I wrote were carefully read and revised by members in my online writing group. Their feedback made me see very clearly where I need to improve my writing (lotusily, 2014)

Another repeater described the feedback process (7) when participating in the online group learning:

(7) Writing: two good (raw score), a total score of 28. Although I was not surprised, I still felt a little excited because I only got a fair and a good, overall score of 24 in my previous test. I would like to express my gratitude to my online writing group!! I only had 10 days to prepare for the test. I wrote 10 essays with my group, and I improved a lot. The biggest improvement came from understanding how to write. Sometimes it was not that I was unable to write good sentences, but rather from not familiar with the writing steps. I know what to write, but the writing looks awkward and full of mistakes. With the “writing + revising” activities in the group, at least I found out where I most likely to make mistakes and understood which writing structure worked better. Everyone in the group started writing the same topic at the same time, then we corrected each other’s writings. By comparing my own essays with those written by others also helped me learn how other people write the same topic, and I even used some parts written by other group members in my TOEFL test, for example, the beginning part of an essay (frankiyian, 2012).

The descriptions of how the *Speeding listening*, *Dictation*, *Shadow speaking* and *Get feedback for speaking and writing* strategies were used by repeaters during their test preparation clearly showed a focus on improving the English listening and writing skills rather than test-taking skills. Therefore, I categorized the above individual strategies as the language learning strategy type.

Test Preparation Curriculum

The repeater reports contain a lot of thorough and expounded descriptions of how repeaters used test preparation curriculum during test preparation and disclosed the connections between the curriculum and the TOEFL iBT test content. Two individual strategies used by repeaters: *Practice TPO materials* and *Practice with Ji-jing items* are particularly relevant to this aspect.

TOEFL Practice Online (TPO) is an ETS developed TOEFL internet-based simulation practice test that contains test items previously used in the true TOEFL iBT tests (e.g., TPO Volume 31 contains all the test items used in the January 2011 true test). Candidates can practice TPO on a computer with an internet connection and receive scores and feedback from ETS TOEFL raters within 24 hours on all the four subskills measured on the test. The authentic test items in TPO allow the user to see what kind of real questions are used in the test. Of course, these questions are removed from the ETS TOEFL item bank and would not appear in true tests again. Each volume of TPO has a different set of questions, and currently, there are 34 official volumes of TPO sold at \$45.95 per volume on ETS websites. The conventional sense of using TPO is to use it as a test simulation which gives test takers the experience of taking the real

TOEFL iBT (“Frequently Asked Questions about TOEFL® Practice Online”, 2019). Previous research on TOEFL preparation, which assumed TPO was mainly used in such a manner, found that the frequency of practicing with the TPO to be a strong predictor for improving test performance (Liu, 2014). The repeater reports in this project showed that most repeaters heavily relied on TPO in their preparation; however, the TPO was not used as a test simulation by TOEFL iBT repeaters in China. Almost all repeaters in my study used TPO as the primary test preparation curriculum and, for some, TPO was the only curriculum during their test preparation.

Using TPO as a curriculum may be particularly relevant to TOEFL test-takers in China. Due to the lack of copywrite protection enforcement systems, most TOEFL preparation materials are freely shared online among all test takers in China - this includes a free version of all the 34 volumes of TPO materials developed by ETS. The combined volumes of TPO materials provide test takers and repeaters with a huge number of true test questions to learn from and practice on, making TPO materials the set of test preparation curriculum that most resembles the TOEFL testing content. Many repeaters in China believed that they could find TOEFL test item patterns if they practiced a lot of TPO questions, and that this would lead to big improvements to their test scores. The following example (8) is the listening preparation strategy section in the report written by a repeater who achieved a whopping 11-point gain on the listening section in the most recent repeated test:

- (8) *A score of 29 on listening came as a surprise, there must be a bit of luck in it, but I want to tell you, it is possible to improve your listening scores in just 17 days. You all see how bad I did on my listening section in my First Battle (well, every section was bad), a score of 18 means that I did not understand most of the of them except for just a few sentences. But during my Second Battle, everything I heard was so clear, and I was able to identify the correct answers within seconds. I got distracted a little over the excitement when I*

started to answer the test questions, but I still got them right. So, there are skills in how to listen. Where do the skills come from? Do your TPOs! First, TPO will train you to be focused. You must listen to the whole set to make it effective. Second, TPO will allow you to get a grip on the test items. I used 4 hours to listen to each set of TPO for several times. Answer the questions after the first run regardless of how much you understood, then listen for the second time sentence by sentence and make sure you understood everything, I mean every word of it. This is basic. No matter how much testing tricks you have, they would be useless if you can't understand the listening. After this run, you listen to the whole set once again. Since you have already heard it and answered the questions, now you will be very sensitive to the logic behind the test items, then you will start to build your listening sensitivity. Actually, I read some high achievers test preparation strategies before my First Battle, they all mentioned "keywords," such as "but, except, for example...", but I did not get it at that time. Until in preparation for the repeated test when I found some good analysis on the listening materials, suddenly, I understood the meaning of "keywords." (I will attach it at the end of the post. including 26 sets of listening practices). If you read them carefully, there will be great rewards! Third, TPO will make us understand the pattern of how the developers in ETS write the test items. If you analyze a few sets of TPO practices you have already done, you will find the pattern. The correct answer can be found by eliminating the three distractors, which all bear the pattern that they are either "related" or "relevant" to the author's opinion. If you know this, you can get the right answers even you only understood part of the article! (Sunsissy, 2013).

Some repeaters used TPO as their only curriculum during test preparation and advised other test takers not to use other test preparation materials. For example, this TOEFL iBT repeater who improved a total test score from 78 to 101 gave the following advice in the report (9):

(9) *Reading has always been a strong sectional skill for Chinese students and easy to get high scores. To achieve this, TPO is the top priority. Promise me, do not read any other TOEFL materials!! If you cannot completely digest this official material (TPO), why bother reading other materials!! Read each of the 34 TPO volumes carefully, and make sure to analyze all the mistakes you made when practicing them!! Please memorized TOEFL vocabulary!! You must look up all the new words that appeared in TPO and memorize them. Some of the new words repeatedly appear in TPO. You will remember them if you look them up enough times (Stacey700, 2015).*

Ji-jing (机经 in Chinese, loosely translated as “computer experience” in English) is a slang word in the test preparation market that appeared soon after the launching of the TOEFL iBT test in China. It refers to a test taker’s recall of test questions used in true TOEFL iBT tests. ETS used to have a small TOEFL test-item bank and often reused the whole set of previous test questions in the upcoming TOEFL iBT exam. Test takers soon found out this loophole and started to share recalls of previous test items on TOEFL iBT discussion forums. The collective recalls of test questions from thousands of TOEFL test takers emerged as a bank of true TOEFL items – the TOEFL Ji-jing. Some test candidates who studied Ji-jing encountered the same test items on the TOEFL exams, thus artificially boosting their test scores. Such unethical behavior by some test takers brought unfairness to other TOEFL test takers, causing serious problems for the whole language test assessment system, heavily jeopardizing the validity of the TOEFL test.

In the past ten years, ETS has been tackling this issue by increasing the number of test questions in the TOEFL iBT test-item bank and starting to mix test items from many different versions of previous tests when creating the new test versions. This strategy used by EST seems

to be very effective. The online TOEFL Ji-jing forum became inactive around the year 2010 when test takers stopped posting recalls on the forum, since their recalled test questions would not be able to predict most of the items in the new tests. Today, the term Ji-jing is still in use but has a very different meaning. It now exists as a selection of items from the old Ji-jing bank as well as items from other resources that made “wild guesses” of questions that might appear on the new tests. Many TOEFL test takers stopped paying attention to Ji-jing. Even those who use Ji-jing in test preparation treat them as alternative learning materials rather than predictors of future test items.

Ji-jing is still widely mentioned in repeaters reports and many repeaters reported using Ji-jing questions to narrow down topics for practicing writing and speaking tasks. However, most repeaters warned test takers not to rely on using Ji-jing as predictors for test items. Many repeaters saw Ji-jing as a distraction and advised other test takers to ignore Ji-jing completely. The comments in this report (10) reflected the perception of TOEFL Ji-jing among many repeaters:

(10) To sum up, the magical days of the Ji-jing have long gone. Since ETS changed the test development strategy, I have noticed that Ji-jing had not been able to make many right predictions on testing items. Sometimes just predicted one speaking task question or one writing topic, not much on listening or reading at all. Not very helpful anyways. Some coaching schools or companies bragged about how magical their Ji-jing materials were, but none of them was credible! However, your attitude toward Ji-jing may heavily influence your test preparation and especially on your time management and confidence in the test. I failed my Second Battle because of this. I bought a lot of Ji-jing materials

from various sources, memorized all the test questions. But when I did not see any of them in the reading section during the test, I was so upset and was not able to perform at all (jsc405044049, 2013).

Repeater Perceptions about Test Preparation Strategies

Taking GMAT or GRE exams was usually listed as a test-taker characteristic or background information by previous research; however, in this study, it was categorized as an individual strategy because many repeaters strongly promoted it as a test preparation strategy in their reports. Over one-third of the repeaters took the graduate adaptive tests during their test intervals. Many repeaters stated in their reports that they purposefully took the GRE or GMAT tests during TOEFL preparation to help improve their TOEFL performance. Many repeaters felt that taking the GRE or GMAT exams improved their reading skills and increased their vocabulary base, which made them feel the TOEFL test was much easier to handle.

Several of the preparation activities, such as memorizing prewritten templates and memorizing examples to help complete the speaking and writing tasks in the TOEFL iBT test, were individual strategies well-documented by studies on TOEFL preparation (Liu, 2014. Yu et al., 2017). However, many repeaters strongly argued for the distinction between using a template or examples written by other people and using the ones written by the test taker themselves. In this study, I also made a distinction when categorizing individual strategies based on who wrote the templates and which examples were to be used in the speaking and writing tasks during test preparation. The following paragraph (11) is a repeater discussing her perception about the difference between the two strategies:

(11) *My experience told me: **Do not use a writing template written by others!!** Especially those templated distributed by coaching schools. Too wordy. Easily jotting down 200 words without any real stuff. For example: “this is the first difference between concept and fact bla, bla, bla.” Don’t use them. **You must use a template, but you have to write your own template** (alongoose86, 2012).*

4.3 Quantitative Analysis

Test Preparation Strategies and Total Score Changes

The relationship between the seven aggregated test preparation strategies and total score gains are shown in Table 5. Simple regression analyses indicate that when these aggregated test preparation strategies were examined separately, *Memorize TOEFL vocabulary (reading + listening)*, *Intensive learning (reading + listening)*, *Get feedback for output (speaking + writing)*, and *Shadow speaking (listening + writing)* strategies appeared to be strong predictors of the total score gains. In addition to the above strategies, *Coaching school attendance* is also positively related to test score changes. The *Number of times taken the test*, though not a test preparation strategy, also appeared to be a strong predictor of the total score gains. The multiple regression model showed that when including all the strategies in the same regression, *Intensive learning* is the strongest predictor of the total score gains significant at $p < .01$ level. *Memorize TOEFL Vocabulary*, *Shadow speaking*, *Practice TPO materials (reading + listening + speaking + writing)* are also strong predictors, significant at $p < .5$ and *Get feedback for output* is a predictor significant at $p < .1$.

Table 5.

Results from regression analyses on total score change (n=166)

Predictor Variables	Simple Regression			Multiple Regression		
	B	p	r ²	B	p	r ²
Practice TPO materials	1.306*	.066	.021	1.235**	.021	
Memorize TOEFL vocabulary	4.780***	.000	.095	2.405**	.037	
Intensive learning (learn word by word)	4.304***	.000	.095	2.571***	.005	
Get feedback for output	6.475***	.001	.065	2.963*	.071	
Practice Ji-jing materials	.107	.902	.000	.252	.728	
Shadow speaking	6.422***	.000	.073	2.578	.103	
Coaching school attendance	7.671*	.053	.034	3.167	.237	
Take GMAT or GRE exams	3.981**	.015	.036	1.94	.183	
Number of times taken the test	2.589***	.000	.191	1.975*	.082	
						.361

Note: **B** = unstandardized regression coefficient; * p<.1, ** p<.05, *** p<.01.

Among the significant predictors for total score gains, *Get feedback for output* has the largest magnitude, associated with an estimated total score increase of 5.92 ($2.96 \times 2 = 5.92$). The *Intensive learning* strategy and the *Memorize TOEFL vocabulary* strategy are associated with 5.14 ($2 \times 2.57 = 5.14$) points and 4.8 ($2 \times 2.4 = 4.8$) points of increases in the total scores, respectively. *Practice TPO materials* in all the four sections is associated with 4.94 ($1.235 \times 4 = 4.94$) points of the total score increase. The *Number of tests taken* is also associated with an increase in the total score by 1.975 points. *Attending coaching schools* has no effects once other strategies are included in the regression, which is consistent with previous studies that find that attending coaching schools has little effect in score gains (e.g., Liu, 2014; Bachman, Davidson, Ryan & Choi, 1995).

The correlation analysis results show that some strategies are significantly correlated (see Table 6). For example, *Memorize TOEFL vocabulary* and *Taking GMAT or GRE exams* are

significantly correlated. This was not surprising since taking the GMAT or GRE exams requires test takers to have a good command of an extensive vocabulary base, which would help these test takers perform better on the TOEFL test. Many repeaters explicitly stated in their reports that one of the reasons they took these graduate adaptive exams during their TOEFL preparation was to help them learn vocabulary and improve reading skills.

Table 6.

Correlation among aggregated strategies (n=166)

Variables	1	2	3	4	5	6	7	8	9
1. Taking GMAT/GRE exams	1								
2. Practice TPO materials	-.043	1							
3. Memorize TOEFL vocab	.218***	.100	1						
4. Intensive learning	-.021	.074	.175**	1					
5. Practice Ji-jing items	-.094	.216***	-.063	.001	1				
6. Get feedback for output	.081	.026	.091	.148	-.004	1			
7. Number of tests taken	.191	-.109	.215***	.107	-.082	.174**	1		
8. Shadow speaking	-.038	.123	.085	.261***	-.011	.240***	.133*	1	
9. Coaching school attendance	.085	-.029	-.044	.104	.06	.031	.199**	.193	1

Note: *p < .1. **p < .05. ***p < .01.

Test Preparation Strategies and Subscore Changes

Simple regression analyses and multiple regression analyses were conducted to investigate how each individual strategy is related to sectional subscore changes. The results are shown in Tables 7 through 10.

Table 7

Results for regression analysis on reading subscore changes (n=131)

Predictor Variables	Simple Regression			Multiple Regression		
	B	p	r ²	B	p	r ²
Practice TPO reading materials	0.603	-0.452	0.004	-0.123	-0.878	
Memorize TOEFL reading vocabulary	1.718**	-0.021	0.04	1.410*	-0.057	
Intensive Reading (read word by word)	2.247***	-0.007	0.055	2.202***	-0.01	
Practice Ji-jing reading materials	1.681	-0.442	0.005	2.201	-0.305	
Extensive Reading (variety of materials)	-2.659	-0.139	0.017	-2.506	-0.15	
						.108

Note: **B** = unstandardized regression coefficient; * p<.1, ** p<.05, *** p<.01.

Regression analyses on reading subscore changes (Table 7) show that *Memorize TOEFL reading vocabulary* (**B** = 1.718) and *Intensive reading* (**B** = 2.247) are strong predictors of the reading subscore gains. When including all strategies for improving reading in the multiple regression model, these two individual strategies remain significantly related to the reading score gains, especially the *Intensive Reading* strategy, which is associated with an increase of 2.2 points in the reading subscore.

Table 8

Results for regression analysis on listening subscore changes (n=136)

Predictor Variables	Simple Regression			Multiple Regression		
	B	p	r ²	B	p	r ²
Practice TPO listening materials	1.429*	.088	.022	.784	.348	
Memorize TOEFL listening vocabulary	2.253*	.053	.028	1.624	.155	
Intensive listening (listen word by word)	1.597*	.054	.027	1.346	.110	
Extensive listening (variety of materials)	-1.385	.0511	.003	-2.165	.302	
Speed listening (accelerated playback)	2.838***	.009	.050	2.738**	.012	
Note taking during test	-.514	.591	.002	-.225	.811	
Shadow speaking (to improve listening)	2.172*	.084	.022	1.422	.252	
Dictation Practices	.897	.322	.007	.861	.337	
Practice Ji-jing listening items	-2.491	.164	.014	-1.699	.330	
						.141

Note: **B** = unstandardized regression coefficient; * p<.1, ** p<.05, *** p<.01.

Among the individual strategies for listening, *Practice TPO listening materials* ($\mathbf{B} = 1.429$), *Memorize TOEFL listening vocabulary* ($\mathbf{B} = 2.253$), *Intensive listening* ($\mathbf{B} = 1.597$), *Speed listening* ($\mathbf{B} = 2.838$) and *Shadow speaking* ($\mathbf{B} = 2.172$) strategies all show significant relationships with the listening subscore increase, with *Speed listening* to be the strongest predictor of listening subscore gains. When including all strategies together into the multiple regression model, only *Speed listening* remains to be a strong predictor of the listening subscore changes (See Table 8).

Table 9

Results for regression analysis on speaking subscore changes (n=132)

Predictor Variables	Simple Regression			Multiple Regression		
	B	p	r ²	B	p	r ²
Practice TPO speaking materials	-.364	.542	.003	.146	.816	
Use speaking examples written by others	-.345	.654	.002	-.677	.380	
Use speaking templates written by others	-.210	.717	.001	-.299	.612	
Write own speaking examples	-.181	.769	.001	.037	.951	
Write own speaking templates	-.764	.146	.016	-.694	.181	
Get feedback for speaking samples	2.786*	.069	.025	3.172**	.042	
Shadow speaking (to practice speaking)	2.504**	.011	.049	2.748***	.006	
Practice Ji-jing speaking items	-.894	.108	.02	-1.001*	.086	
						.122

Note: **B** = unstandardized regression coefficient; * p<.1, ** p<.05, *** p<.01.

Regression analysis results show that *Get feedback for speaking samples* ($\mathbf{B} = 2.786$) and *Shadow speaking* ($\mathbf{B} = 2.504$) strategies are positively related to the speaking subscore gains. *Practice Ji-jing speaking items* is negatively related to subscore gains.

Table 10

Results for regression analysis on writing subscore changes (n=133)

Predictor Variables	Simple Regression			Multiple Regression		
	B	p	r ²	B	p	r ²
Practice TPO writing materials	.614	.355	.007	.452	.518	
Use writing examples written by others	1.610	.145	.016	1.260	.251	
Use writing templates written by others	1.507	.101	.020	.960	.147	
Write own writing examples	.275	.684	.001	-.003	.996	
Write own writing templates	.513	.387	.006	.613	.292	
Get feedback for writing samples	2.325***	.003	.066	2.432***	.002	
Practice Ji-jing writing items	.039	.951	.000	-.408	.530	
						.112

Note: **B** = unstandardized regression coefficient; * p<.1, ** p<.05, *** p<.01.

Finally, the regression analyses on writing subscore gains (Table 10) show that only *Get feedback for writing samples* (**B** = 2.325) demonstrated a significant relationship with the writing subscore gains.

In conclusion, the descriptive statistic analyses revealed that TOEFL iBT repeaters who reported their test preparation on the online discussion boards in China had relatively high average first test scores and achieved substantial score gains in the repeated test. Most repeaters set 100 or 110 as their target scores for the repeated test. Repeaters, on average, used less time to prepare for the repeater test than the first-time test takers, as found in previous research (Liu, 2014). Thirteen English language learning type strategies and 21 TOEFL test-specific strategies were identified, including several new test preparation strategies not recorded by previous studies on TOEFL iBT test preparation (Liu, 2014; Yu et al., 2017). The qualitative analysis showed how repeaters implement some of the test preparation strategies and particularly, how the repeaters used test preparation curriculum and perceived the usefulness of certain test preparation strategies. The simple and multiple regression analyses showed that *Intensive*

learning, Memorize TOEFL vocabulary, Studying TPO materials, Get feedback for speaking samples, Get feedbacks for writing samples, Speed listening and Shadow speaking strategies are strong predictors of the total score gains. It is important to note that no causal inference is implied in the findings. All the reported relationships are correlational.

CHAPTER 5. CONCLUSION

The current study investigated TOEFL iBT repeaters' test preparation strategies and the relationship between test preparation strategies and the improvement in performance on the repeated test. The results of the analyses carried out in this study contribute to the understanding of TOEFL iBT repeaters' behaviors and provide empirical evidence relevant to the validity argument of the TOEFL iBT test. Major findings are summarized in this chapter followed by implications of the findings and limitations of this study.

5.1. Major findings

For the first research question about TOEFL iBT repeaters' test-taking patterns in China, the findings show that most repeaters used 100 or 110 points as their target testing scores and spent between one to three weeks to prepare for the repeated tests. On average, repeaters achieved a 16.4 point score gain in the repeated tests and made the most subscore improvement (5.94 points) in the listening subskill section. Most (95%) of the repeaters did not attend coaching schools during their preparation for the repeated tests.

In response to the second research question, regarding what and how the preparation strategies TOEFL iBT repeaters were used by repeaters, this study identified 13 language learning strategies and 21 test-specific strategies. *Speed listening*, *Dictation*, *Shadow speaking*, and *Get feedback for speaking and writing samples* stand out as newly identified individual strategies that have not been reported by current literature on TOEFL test preparation. Interestingly, all these newly identified test preparation strategies require using new technology. For example, *Speed listening* requires using special mobile apps or personal computer software that can play audio files at various speeds. The detailed descriptions of how these new strategies

were implemented by repeaters revealed some new test preparation behaviors regarding the TOEFL iBT repeaters. Particularly, many repeaters joined online learning groups where group members share test preparation materials, answer test questions, and correct each other's speaking and writing samples.

Regarding my third research question on the relationship between test preparation strategies and test performance, the findings show that *Intensive learning*, *Memorize TOEFL vocabulary*, *Studying TPO materials*, *Get feedback for speaking samples*, *Get feedbacks for writing samples*, *Speed listening* and *Shadow speaking* strategies are significantly associated with the total score and sectional subscore improvements.

5.2 Limitations

A key limitation of this study is that the reports in the sample were selected to include only the successful repeaters from a discussion forum consisting of participants who had opted to join completely on their own. Repeaters who write about their test preparation experience might have different characteristics than repeaters in general. Therefore, the test preparation behaviors and outcomes of the repeaters in this study do not reflect those of the general TOEFL repeaters population. Furthermore, it is possible that some strategies the repeaters used during test preparation were not reported by the repeaters.

Due to space constraints, this study did not examine the cross-sectional influences of the individual strategies. For example, if the strategies targeting the listening section were correlated to performance on the speaking section. Future research could address longitudinal questions regarding how repeaters' language abilities development relates to the time spent, materials used, and activities conducted during test preparation, as well as how test preparation strategies change

across test occasions, and what are the relationships between cross-sectional domain focused strategies.

5.3 Implications

The findings of the actual strategies that repeaters reported using substantiate the inferences that support the validity argument of the TOEFL iBT test and provide information that may benefit all the stakeholders in the TOEFL iBT language assessment.

The study shows that individual test preparation strategies that are most significantly associated with repeaters' test performance improvements are the English language learning type strategies, except for *Memorize TOEFL vocabulary* and *Practice TPO materials*. However, although *Memorize TOEFL vocabulary* is categorized as a test-focused strategy type in the research field (e.g., Liu, 2014; Yu et al., 2017), it is difficult to separate factors contributing to language learning or test-taking skills when test takers memorize vocabulary. Research finds that the item format for vocabulary assessment in the TOEFL iBT encourages understanding and learning vocabulary in context, thus produces positive washback effects (Qian, 2008). Therefore, it is arguable that memorizing TOEFL vocabulary helps improve test takers' language ability. Also, as revealed by repeater reports, repeaters in China use TPO materials as curriculum rather than test simulation. So, it is more likely that studying a lot of TPO materials helps repeaters acquire more knowledge about academic English, rather than helping them get familiar with the test format. The above findings on repeaters' test preparation strategies suggest that repeaters' score gains may largely be a function of the growth in their English language proficiency. Since the key assumption underlying the explanation inference of the validity argument in relation to the repeaters' test scores is that the score changes over test occasions reflect the changes of

repeaters' L2 ability over time. The above results provide evidence that backs this assumption thus supporting the validity argument of the TOEFL iBT test.

For TOEFL test takers, the findings of the test preparation strategies that have a significant relationship with test performance suggest that when practicing TOEFL reading and listening skills, it might be more helpful to use an intensive learning approach. Test takers may benefit more by reading a limited number of English texts word by word for several times rather than skimming through a wide variety of materials. Listening to English at an accelerated speed may help improve listening skills, especially for intermediate or advanced students. As shown in the examples in section 4.2, after practicing speed listening, repeaters felt that English audio played at the normal speed sounded like “slow motions in a movie” (4) and became easier to understand. To improve speaking and writing skills, getting feedback for speaking and writing samples might be helpful. It is important to point out that getting feedback does not mean receiving instruction from English teachers or trainers. In fact, almost all the feedback the repeaters received in this study came from peer test takers in their online learning groups. Lastly, memorizing TOEFL vocabulary can also be a useful strategy.

In terms of repeaters' perceptions about test preparation strategies, the findings show that taking a graduate admission exam (e.g., GMAT or GRE) does seem to help improve TOEFL scores, which was consistent with the common belief among repeaters. However, *Taking the GMAT and GRE exam* was significantly correlated with *Memorize TOEFL vocabulary*; therefore, it is more likely that it was the increased vocabulary that helped improve repeaters' test performance. Contrary to most repeaters' perceptions, whether you use your own templates (or examples) or templates (or examples) written by others in the speaking and writing tasks did not seem to make a difference in test performance.

For the test users, this study provides information for the interpretation of repeaters' multiple TOEFL iBT test scores. The relationship between test preparation strategies and repeaters' test score improvements suggests that repeaters' score gains are strongly associated with test preparation strategies incorporating language learning focuses. Therefore, it is highly likely that the repeater's score gains are strongly associated with the repeater's English language learning efforts. This suggests that universities may use TOEFL iBT repeaters' highest test scores in making admission decisions without worrying about their poor performance on the previous tests or the number of times the repeaters had taken the test.

For the test developers, the findings that repeaters' score gains have a significant relationship with English language learning strategies but not TOEFL test-specific strategies suggest that repeaters' improved test scores may be largely attributed to the construct of academic proficiency, which is what the test is intended to measure. The result that *Practice Ji-jing items* strategy was negatively associated with test performance shows that developers' longstanding efforts to eliminate the negative impact of test-takers' recall of test items were highly successful. The qualitative analysis on how repeaters in China gained access to all the TPO materials and how TPO materials were actually used by repeaters during test preparation can help test developers re-evaluate the development, publishing, and distribution of TOEFL test preparation materials.

REFERENCES

- Advertising Service, (2004). Chasedream.com. Retrieved from:
<https://www.chasedream.com/show.aspx?id=6&cid=52>
- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13(3), 280-297.
- Bachman, L. F., Davidson, F., Ryan, K., & Choi, I. C. (1995). *An investigation of comparability of two tests of English as a foreign language*. Cambridge, England: Cambridge University Press.
- Barkaoui, K. (2017). Examining repeaters' performance on second language proficiency tests: a review and a call for research. *Language Assessment Quarterly*, 14(4), 420-431.
- Chapelle, C. A. (2008). The TOEFL validity argument. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 319–352). New York, NY: Routledge.
- Chapelle, C. A. (2008). Utilizing technology in language assessment. In N. H. Hornberger & E. Shohamy (Eds.), *Encyclopedia of language and education, volume 7 language testing and assessment* (pp. 123–134). New York, NY: Springer.
- Cliffordson, C. (2004). Effects of practice and intellectual growth on performance on the Swedish scholastic aptitude test (SweSAT). *European Journal of Psychological Assessment*, 20(3), 192–204.
- Cohen, A. D. (2012). Test-taking strategies. *The Cambridge guide to second language assessment*, 96-104.
- Company Intro (2020). Zhao.com. Retrieved from:
<https://www.zhan.com/aboutus/aboutus>
- Converse, B. A., Risen, J. L., & Carter, T. J. (2012). Investing in karma: When wanting promotes helping. *Psychological Science*, 23(8), 923-930.
- Dai, 2016. How Many TOEFL Candidates in China in 2016? International Education Web. Retrieved from:
<https://training.ieduchina.com/toefl/201608/13213.html>
- Ding (2017). From “Goodwill” to “Badwill”, the Superstitious Beliefs among Chinese Internet Users. *Today's Topic at Tengxun*. Vol. 3851. Retrieved from:
<https://view.news.qq.com/original/intouchtoday/n3851.html>

- Elder, C., & O' Loughlin, K. (2003). Investigating the relationship between intensive English language study and band score gain on IELTS. In R. Tulloh (Ed.), *IELTS Research Reports Volume 4*(6) (pp. 207–254).
- Frequently Asked Questions about TOEFL® Practice Online (2019), retrieved from: https://www.ets.org/s/toefl/pdf/toefl_tpo_faq.pdf
- Garcia, J., & Gustavson, A. R. (1997). The science of self-report. *APS Observer*, 10(1).
- Get the TOEFL Test Advantage (2020). Retrieved from: https://www.ets.org/toefl/teachers_advisors/advantage/
- Green, A. (2005). EAP study recommendations and score gains on the IELTS academic writing test. *Assessing Writing*, 10(1), 44–60.
- Green, A. (2007). Washback to learning outcomes: A comparative study of IELTS preparation and university pre-sessional language courses. *Assessment in Education*, 14(1), 75–97.
- International Education Industry Research Report. (2019, March). Duoqing Capital Research. Retrieved from: http://pdf.dfcfw.com/pdf/H3_AP202003101376070182_1.pdf
- Kulik, J. A., Kulik, C. L. C., & Bangert, R. L. (1984). Effects of practice on aptitude and achievement test scores. *American Educational Research Journal*, 21, 435–447.
- Kokhan, K., & Lin, C. K. (2014). Test of English as a Foreign Language (TOEFL): Interpretation of multiple score reports for ESL placement. *Papers in Language Testing and Assessment*, 3(1), 1-23.
- Li, Q.Y. (2002). Problems and Reflection on CET band 4 and Band 6. *Foreign Language Education*, 23(5),33-38.
- Lin, 2008. Sohu.com Interview with ETS Senior Representative in China. Retried from: <http://goabroad.sohu.com/20081020/n260142310.shtml>
- Ling, G., Powers, D. E., & Adler, R. M. (2014). Do TOEFL iBT® Scores Reflect Improvement in English-Language Proficiency? Extending the TOEFL iBT Validity Argument. *ETS Research Report Series*, 2014(1), 1-16.
- Liu, O. L. (2014). Investigating the relationship between test preparation and TOEFL iBT® performance. *ETS Research Report Series*, 2014(2), 1–13.
- Ma, J. (2017). *Understanding Test Preparation Phenomenon Through Chinese Students' Journey Towards Success on High-Stakes English Language Tests* (Doctoral dissertation, Queen's University, Kingston, Canada). Retrieved from: https://qspace.library.queensu.ca/bitstream/handle/1974/22703/Ma_Jia_201709_PhD.pdf

- Matoush, M. M., & Fu, D. (2012). Tests of English language as significant thresholds for college-bound Chinese and the washback of test-preparation. *Changing English, 19*(1), 111-121.
- Messick, S. (1981). The controversy over coaching: Issues of effectiveness and equity. In B. F. Green (Ed.), *Issues in testing: Coaching, disclosure, and ethnic bias* (pp. 21–53). San Francisco, CA: Jossey-Bass.
- Messick, S. (1982). Issues of effectiveness and equity in the coaching controversy: Implications for educational and testing practice. *Educational Psychologist, 17*, 67–91.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*(2), 5–11
- Messick, S. (1996). Validity and washback in language testing. *Language testing, 13*(3), 241-256.
- Messick, S., & Jungeblut, A. (1981). Time and method in coaching for the SAT. *Psychological Bulletin, 89*(2), 191.
- Qian, D. D. (2008). From single words to passages: Contextual effects on predictive power of vocabulary measures for assessing reading performance. *Language Assessment Quarterly, 5*(1), 1-19.
- Sackett, P. R., Burris, L. R., & Ryan, A. M. (1989). Coaching and practice effects in personnel selection.
- Swain, M., Huang, L. S., Barkaoui, K., Brooks, L., & Lapkin, S. (2009). The speaking section of the TOEFL iBT™ (SSTIBT): test-takers' reported strategic behaviors. *ETS Research Report Series, 2009*(2), i-118.
- Tang, W. (2010). Minhong Yu's talk on "happy learning and healthy grow-up" in Wuxi. *The New Oriental*.
- Test and Score Data. (2018). Education Testing Service. Retrieved from https://www.ets.org/s/toefl/pdf/94227_unlweb.pdf
- Ward, A. M., & Xu, L. (1994). *The relationship between summarization skills and TOEFL scores*. Paper presented at the annual meeting of the Teachers of English to Speakers of Other Languages, Baltimore, MD. Retrieved from ERIC database.
- Wilson, K. M. (1987). *Patterns of test taking and score change for examinees who repeat the test of English as a foreign language*. Princeton, NJ: Educational Testing Service.

- Winke, P., & Lim, H. (2017). The Effects of Test Preparation on Second-Language Listening Test Performance. *Language Assessment Quarterly*, 14(4), 380-397.
- Wu, A., & Pang, A. (2013, February 27). Test of English Easily Bested by Cheats in China. *The Epoch Times*. Retrieved from:
https://www.theepochtimes.com/test-of-english-easily-bested-by-cheats-in-china_1476987.html
- Xie (2010). "Accumulate Goodwill Credit" Becomes Popular among College Students in China. *The China Youth*. January 22, 2010. Retrieved from:
http://www.cyol.net/zqb/content/2010-01/22/content_3053111.htm
- Xie, Q., & Andrews, S. (2013). Do test design and uses influence test preparation? Testing a model of washback with Structural Equation Modeling. *Language Testing*, 30(1), 49–70.
- Yu, G., He, L., Rea-Dickins, P., Kiely, R., Lu, Y., Zhang, J., ... & Fang, L. (2017). Preparing for the Speaking Tasks of the TOEFL iBT® Test: An Investigation of the Journeys of Chinese Test Takers. *ETS Research Report Series*, 2017(1), 1-59.
- Yu, G., & Jin, Y. (2014). English language assessment in China: *Policies, practices and impacts*. *Assessment in Education: Principles, Policy & Practice*, 21(3), 245–250.
<https://doi.org/10.1080/0969594x.2014.937936>
- Zhang, Y. (2008). *Repeater analysis for TOEFL iBT* (Research Memorandum No. RM-08-05). Princeton, NJ: Educational Testing Service.

APPENDIX: AN EMAIL FROM ETS TO TOEFL IBT TEST-TAKERS

Re: Ready to register to retake the test?

Congratulations on taking the TOEFL® test! If you're like many students, you probably did better on some sections than others. To help boost your TOEFL scores in the areas you need it most, check out this page on improving your skills. It offers helpful tips in Reading, Listening, Speaking and Writing for all skill levels. There's advice to help you:

Expand your vocabulary

Enhance your listening skills

Perfect your ability to summarize an article

Pronounce words clearly and speak with good intonation

Once you feel more confident in your skills, register to retake the test. Then you can feel good about sending your best scores on the test that helps more than 90% of TOEFL test takers get into their 1st- or 2nd-choice university.

Ready to register to retake the test?

Retrieved from <http://bbs.gter.net/thread-1775529-1-1.html>