INVESTIGATING THE EFFECT OF ANXIETY ON TEST TAKERS' PERFORMANCE ON A LARGE-SCALE HIGH-STAKES COMPUTER-BASED ENGLISH LISTENING AND SPEAKING TEST

By

Jinhong Chen

A thesis submitted to the Faculty of Education in conformity with the requirements for the degree of Master of Education

Queen's University

Kingston, Ontario, Canada

July, 2022

Abstract

Previous research have suggested that anxiety may affect test takers' performance on high-stakes tests, such as the National College Entrance Examination (Guangdong) Computer-based English Listening and Speaking Test (the CELST). The CELST test takers can potentially experience three types of anxiety, namely cognitive test anxiety (CTA), foreign language listening anxiety (FLLA) and foreign language speaking anxiety (FLSA). Despite the high-stakes impacts the CELST has on millions of test takers' future, scarce research has specifically investigated the effect of anxiety on test takers performance and the source of anxiety from the CELST test takers' perspectives. Therefore. the present study aimed to investigate how CTA, FLLA and FLSA affect CELST test performance.

Quantitative data were collected by administering questionnaires to 585 CELST test takers. Qualitative data were collected by individual semi-structured interviews with twelve test takers. The quantitative analysis showed that: (1) most test takers experienced a moderate degree of CTA, cognitive dimension of FLLA and FLSA; (2) females had a significantly higher level of CTA, cognitive dimensions of FLLA and FLSA than males. (3) self-perceived English competence did impact test takers' CTA, FLLA and FLSA; (4) the frequency of language use negatively correlated with CTA, FLLA and FLSA; (5) CTA, behavioral dimension of FLSA negatively predicted the CELST scores while the somatic dimension of FLSA positively predicted the CELST scores.

Second, the qualitative analysis confirmed the existence and interference of CTA, FLLA and FLSA among CELST test takers. Besides, the results revealed reasons that lead to anxiety in the CELST, and identified helpful strategies to alleviate anxiety from the test takers' perspectives.

The findings of the present study have important implications. Theoretically, the study bridged the gap by deepening the understanding of how CTA, FLLA and FLSA affect test takers' performance in high-stakes foreign language tests. Practically, the findings provide valuable references for the CELST test developers to design test tasks that consider the effect of anxiety such that the CELST can elicit test takers' best performance. Besides, the findings also provide insights for foreign language teachers to alleviate their students' anxiety so that students can better perform during the CELST.

Keywords: English listening and speaking test; test anxiety; foreign language anxiety; test performance

Acknowledgement

My academic journey of pursuing a dual master's degree at South China Normal University, Guangzhou, China and Queen's University, Kingston, Canada has been filled with rewards and challenges. Without the support and encouragement of so many people, I would not have gotten this far.

I would like to thank my supervisor Dr. Yuzhu Tang at South China Normal University for his thoughtful comments on this thesis. My deepest and sincerest gratitude goes to my supervisor Dr. Saad Chahine at Queen's University. Dr. Saad Chahine has offered me invaluable feedback on numerous revisions of this thesis. Specially, his expertise in quantitative research helped me navigate through the complex quantitative data analysis of the study. I especially want to thank my committee member Dr. Liying Cheng at Queen's University for her guidance. Her course on foreign language education and her great expertise have guided me into this wonderful research field. Her encouragement to write this thesis made me confident that the topic was worthy of investigation.

I would like to extend my thanks to the teaching staff in the School of Foreign Studies, South China Normal University and the Faculty of Education, Queen's University. Their courses have pushed me to sharpen my research skills and helped me become a better researcher and thinker.

I am really thankful to my friend Melinda Madison for her meticulous proofreading of the thesis. A special word of gratitude also goes to August Bradley. His PPV system methodology and its Notion implementation have given me alignment and efficiency during the process of completing this research project. Finally, I am extremely grateful to my friends' continuous encouragement and my family's unconditional support throughout the entire academic journey.

Contents

List of Tables

List of Figures

List of Abbreviations

NMET	National Matriculation English Test
The CELST	The Computer-based Listening and Speaking Test (Part of the NMET)
FL	Foreign Language
CTA	Cognitive Test Anxiety
FLLA	Foreign Language Listening Anxiety
FLSA	Foreign Language Speaking Anxiety
FLCAS	Foreign Language Classroom Anxiety Scale
FLLAS	Foreign Language Listening Anxiety Scale
FLSAS	Foreign Language Speaking Anxiety Scale
EFL	English as a Foreign Language
ESL	English as a Second Language
SD	Standard Deviation
ANOVA	Analysis of Variance
EFA	Exploratory Factor Analysis
N	Number

Chapter One Introduction

This chapter starts with the introduction to the research background of the present study. It then presents the research purposes and questions, followed by the significance of the study. Finally, the overview of the thesis is presented.

1.1 Research Background

The National Matriculation English Test (NMET) is the nationwide university entrance test of English for the People's Republic of China, which comprises 20% of the weight of the total college entrance examination scores. The test makes inferences about test takers' English proficiency, which is used in university admission decisions together with the scores from university entrance tests in another five secondary school subjects (Cheng, 2008). The NMET is a norm-referenced standardized test that aims to select secondary school graduates for institutions of higher education. It is a high-stakes English test which influences how society evaluates students and schools and how schools, parents and students evaluate the teachers (Cheng, 2008).

In 2011, the computer-based English listening and speaking test (hereafter the CELST) was incorporated into the Guangdong Province's version of NMET. The CELST comprises 13 % of the NMET scores. Every year, over half a million Guangdong senior high school Grade three students who aim to pursue tertiary education must take the CELST whose scores are used for university admission decisions. For the first time, foreign language speaking ability and listening ability have been tested on a large scale through high-stakes computer-based testing in China.

As shown in Table 1.1, the CELST consists of three parts that focus on assessing test takers' listening and speaking skills. In part one Reading Aloud, test takers are required to watch a one-minute English video clip with subtitles and read after the speaker in the video. Then, students have one minute to practice reading aloud the lines on the screen. Next, they listen to the speaker once again and try to read after the speaker. Finally, they read as the speaker in the video without hearing the original speaker and their responses are recorded for scoring. In part two Role Play, test takers are required to act as a role and communicate with another role played by the computer. First, they are required to watch a two-minute video clip about a dialogue to gain some basic information and take notes. Then they ask the role played by the computer three questions with Chinese prompts and take notes of the answers from the computer. Next, test takers listen to the rest of the dialogue in the previous video and answer five relevant questions from the computer to show their listening comprehension. In part three Story Retelling, test takers are required to listen to a two-minute monologue twice and are given one minute to retell it. A brief summary of the story plots and five keywords are presented during part three. After listening, test takers are given one minute to organize their speech before they start to retell the story in their own words.

Table 1. 1

Structure of the CELST. Source: Adapted from Zhan and Wan (2016, p.5)

Subtest	Test Purposes	Test Format	Full Marks	Testing Time
Part one	Pronunciation and intonation	Reading aloud	20 points	About 30 minutes
Part two	Abilities of receiving and comprehending information; and abilities of asking and answering interactive questions	Role play	16 points	
Part three	Comprehensive abilities of listening and speaking	Story retelling	24 points	
Total			60 points	

The CELST, as a required part of the NEMT for test takers in Guangdong Province, demonstrates its high-stakes nature where test takers must strive for higher scores to get admitted into their ideal universities. Since 2021, the weight of the CELST has increased from 10 % to 13.3 % of the total NMET score, which makes the CELST a more vital part of the NMET and exerts a larger impact on millions of test takers' future. As a high-stakes foreign language (FL) test, the CELST can potentially evoke three types of anxiety, namely cognitive test anxiety (CTA), foreign language listening anxiety (FLLA) and foreign language speaking anxiety (FLSA). Because of the high-stakes impacts of the CELST scores on university admission decisions, the CELST test takers may experience cognitive test anxiety, which represents their worry over their test performance and test outcomes. Previous empirical studies have indicated that cognitive test anxiety was significantly and negatively related to a wide range of educational performance outcomes, including standardized tests and university entrance exams (von der Embse et al., 2018). Besides, considering English is a foreign language to the CELST test takers, attending the CELST might also evoke FLLA and FLSA. Defined as language-skill-specific anxiety, FLLA and FLSA have also been shown to have a negative correlation with test performance on foreign language listening (Elkhafaifi, 2005;Kim, 2000; Zhang, 2013) and foreign language speaking (He, 2018; Woodrow, 2006) respectively. Despite CTA, FLLA and FLSA may negatively affect test takers' performance, there have been no previous empirical studies about the effects of CTA, FLLA and FLSA in the CELST context so far.

When adversely affecting test takers' performance, the CTA, FLLA and FLSA test takers experience may also reduce or threaten the construct validity of the CELST. As Messick (1989) suggested, construct-irrelevant variance such as anxiety poses a major threat to construct validity as such variance undermines the accuracy of test score interpretations and the legitimacy of decisions made based on test scores. For instance, when test takers suffer from anxiety during the CELST and in turn fail to perform well, the test scores they receive might not reflect their true language proficiency because such scores have been negatively affected by their anxiety(Heng Tsung D. Huang & Hung, 2013). Therefore, the adverse effect of anxiety on test performance can affect the interpretation of the scores and harm the test validity. When researchers seek ways to expose evidence that leads to anxiety which would invalidate test inferences, test takers' perspectives of the source of anxiety in the test should be included as the perceptions of test takers provide researchers with significant validation evidence (Cheng & Deluca, 2011; DeLuca et al., 2013; Moss et al., 2006). However, experiences and perspectives of test takers have usually been overlooked in validation studies (Cheng, 2008; Zhan & Wan, 2016). It is clearly problematic that test takers' opinions are undervalued in the validation inquiries of language assessment (Cheng & Deluca, 2011).

The present study is motivated by the following gaps. Despite abundant research on the effects of test anxiety and foreign language anxiety on test performance, most of the empirical studies have examined the effects of anxiety on test performance measured by low-stakes foreign language exams(Huang et al., 2016; In'nami, 2006; Kim, 2001; Salehi & Marefat, 2014; Tsai & Li, 2012). Relatively little research has investigated the effects of either cognitive test anxiety or foreign language anxiety on test takers' performance in large-scale high-stakes foreign language tests which may elicit more anxiety(Salehi & Marefat, 2014; Yoongoo, 2017; Zheng & Cheng, 2018). Additionally, research on foreign language anxiety has focused on the general foreign language classroom anxiety proposed by Horwitz et al (1986). There have been very few empirical studies concerning the effect of language-skill-specific anxiety on high-stakes test performance(Yoongoo, 2017). Moreover, there has been scarce research combining the effect of the three types of anxiety (CTA, FLLA and FLSA) on test performance in a single investigation. To the best of my knowledge, there has been no research specifically examining the effect of anxiety and the source of anxiety in the CELST from the test takers' perspectives.

1.2 Research Purposes and Research Questions

To bridge the research gaps, the present study intended to investigate the characteristics of CTA, FLLA and FLSA among CELST test takers, the effects of CTA, FLLA and FLSA on test takers' performance (Wu, 2021), the reasons leading to anxiety in the CELST as well as strategies to alleviate the anxiety from the test takers' perspectives. Therefore, to achieve the research purposes, the research questions were proposed as follows:

1. What are the characteristics of cognitive test anxiety, foreign language listening anxiety and foreign language speaking anxiety among CELST test takers?
2. What are the effects of cognitive test anxiety, foreign language listening anxiety and foreign language speaking anxiety on CELST test takers' overall scores?
3. From the test takers' perspectives, what reasons may lead to anxiety in the CELST and how do test takers alleviate the anxiety?

1.3 Significance of the Study

The study is of theoretical and practical significance to the field of anxiety research and the CELST development.

Theoretically, considering the fact that scarce research has examined the effect of both CTA and language-skill-specific anxiety such as FLLA and FLSA on large-scale high-stakes foreign language tests, this study hopes to shed light on the subsequent research concerning this particular area and further enrich the empirical studies of CTA, FLLA and FLSA in high-stakes foreign language testing. Second, this research probed into the effects of three types of anxiety altogether on senior high school Grade three students who have been scarcely studied in this field of research. Therefore, the study hopes to cast light on the subsequent anxiety research concerning this group of students.

Hopefully, the findings of this study will also provide significant practical implications for millions of stakeholders involved in the CELST, including policymakers, the CELST test developers, EFL teachers and future CELST test takers. On the one hand, by understanding the effect of anxiety on the CELST performance, the CELST test developers and policymakers can take account of the effect of anxiety when designing and administering the CELST. This way, the future versions of the CELST can reduce the interference of anxiety so that test takers' true language proficiency can be accurately assessed, and thus improve test validity. On the other hand, the findings would draw teachers' attention to the effect of anxiety on the CELST test performance and generate positive washback effect. By recognizing the effect of anxiety on test performance and reasons leading to anxiety in the CELST, English as a foreign language (EFL) teachers could provide helpful instructions to help students deal with anxiety, such as giving more positive feedback after the mock CELST exercises.

1.4 Overview of the Study

The thesis consists of the following five chapters.

Chapter One gives an introduction to the whole thesis. It introduces the research background, research questions and research purposes of the study. Then, the research significance is discussed from the theoretical and practical perspectives. Finally, the structure of the thesis is presented.

Chapter Two introduces the literature review which is broken down into the following sections: 1) defining anxiety, text anxiety and foreign language anxiety; 2) theoretical foundations; 3) studies of test anxiety; 4) studies of foreign language listening anxiety; 5) studies of foreign language speaking anxiety; 6) Summary of literature review.

Chapter Three illustrates the methodology of the present study. It first gives the rationale of the research design of the study. Then, it introduces the participants and instruments of the study. Finally, it elaborates on both the quantitative and qualitative data collection processes and data analyses.

Chapter Four presents the results of the quantitative and qualitative data analyses.

Chapter Five concludes with major findings. Besides, theoretical and practical implications are illustrated. Finally, limitations of the study are presented.

Chapter Two   Literature Review

This chapter first presents definitions of the key terms, followed by theoretical foundations that underpin this study. Then it introduces studies of the effects of test anxiety, foreign language listening anxiety and foreign language speaking anxiety in and out of China. The measures of test anxiety, foreign language listening anxiety, foreign language speaking anxiety are also demonstrated. This chapter ends with a summary of the literature review.

2.1 Definition of key terms

2.1.1 Anxiety

Anxiety, one of the most common and vital reactions to stressful situations, has been referred to the subjective feeling of tension, apprehension and worry related to autonomic nervous system arousal (Lushene et al., 1983). Generally speaking, anxiety is conceptualized into three types: state anxiety, trait anxiety and situation-specific anxiety. The distinction of state anxiety and trait anxiety was first proposed by Cattell and Scheier (1960). Trait anxiety is a stable part of an individual's personality with a more permanent tendency toward anxiety while state anxiety, a transient experience of anxiety and apprehension related to specific events or situations, is experienced at a specific time (Ehrman, 1996; Ellis, 1994; Tasnimi, 2009). Situation-specific anxiety can be regarded as trait anxiety which is limited to a given context (MacIntyre & Gardner, 1991). From this point of view, situation-specific anxiety includes anxiety experienced in well-defined situations such as speaking in public, taking foreign language tests or foreign language class.

A second distinction regarding anxiety is about its facilitating and debilitating effect. Alpert and Haber (1960) first proposed this distinction and differentiated between facilitating and debilitating anxiety. The facilitating anxiety motivates the learner to fight the new learning task while the debilitating anxiety does the opposite by driving the learner to escape from the new learning task (Scovel, 1978). Ellies (1994), as cited in Tasnimi (2009), claimed that a low-level anxiety state has a facilitating effect on the learning process and outcome while a high-level state seems to have a debilitating one. As shown in Figure 2.1, the facilitating and debilitating effect of anxiety on performance can be best described with the Yerkes-Dodson Law (MacIntyre, 1995). When a task is relatively simple, the anxiety level is low and seems to have no negative effect and may actually improve performance. As anxiety increases to an optimal level, performance gradually increases to a peak level. However, when anxiety goes beyond the optimal level, anxiety starts to have a negative effect and impair performance. This law has been proven by a number of studies that showed students with a high level of anxiety tended to perform poorly in language class (He, 2017; Horwitz, 2008).

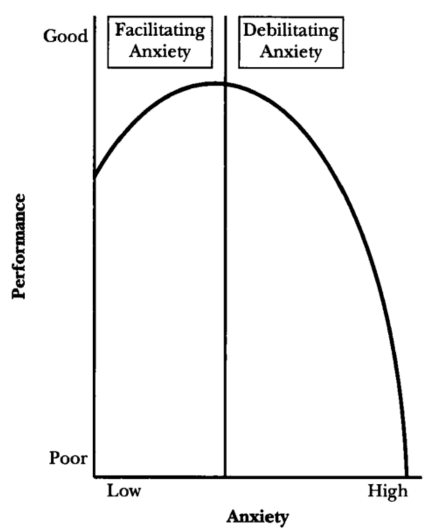

Figure 2. 1 Inverted "U" Relation between Anxiety and Performance. Source: from MacIntyre (1995, p.92)

2.1.2 Test Anxiety

As a scientific construct, test anxiety refers to " the set of cognitive, affective and behavioral reactions that accompany concern over possible negative consequences contingent upon performance in a test or evaluative situation" (Zeidner, 1998, p. 25). Although the question still remains whether test anxiety is best defined as a contemporary state of emotion or a stable personality trait, a widely accepted conceptualization proposed by Spielberger (1976) considers test anxiety as a situation-specific personality trait (Zeidner, 1998). It is widely accepted that test anxiety comprises two dimensions: emotionality and cognitive test anxiety (Cassady & Johnson, 2002).

Emotionality often manifests itself through a series of physiological responses, including: (1) dizziness, (2) nausea, (3) increased heart rate, or (4) feelings of panic (Cassady & Johnson, 2002; Hembree, 1988).

Cognitive test anxiety, traditionally referred to as "worry", consists of cognitive reactions before, during, and after the evaluative situations. Individuals with high levels of cognitive test anxiety commonly struggle with thoughts like: (1) comparing their ability to others, (2) Low self-perceived competence, (3) Thinking of the consequences of failure, (4) cause disappointment for their parents, (5) over worry about evaluation, (6) absence of self-worth. The present study mainly investigated the effect of the cognitive component of test anxiety, which has been shown to be more consistently and negatively correlated with test performance (Hembree, 1988; von der Embse et al., 2018).

2.1.3 Foreign Language Anxiety

Compared to the general anxiety research, the anxiety research on foreign language is a relatively new field (He, 2018). The 1980s was a turning point for foreign language anxiety research. Before 1980s, mixed and confusing results had been produced in the field of foreign language anxiety research (Scovel, 1978). Three kinds of relationship were found between anxiety and foreign language learning, that is, anxiety may exert a debilitating effect, facilitating effect or no effect on foreign language performance (He, 2018). Scovel (1978) attributed this confusing finding to the ambiguous construct of foreign language anxiety. In the 1980s, the landmark publication of Horwitz et al.'s (1986) work, for the first time, conceptualized foreign language anxiety as a distinct and independent construct specific to foreign language learning, that is, a kind of situation-specific anxiety different from state anxiety and trait anxiety. According to Horwitz et al. (1986, p.128), the definition of foreign language anxiety is best stated as: "a distinct complex of self-perceptions, beliefs, feelings, and behaviors related to classroom language learning arising from the uniqueness of the language learning process". As defined by Horwitz et al. (1986), there are three related components of foreign language anxiety, including communication anxiety, fear of negative evaluation and test anxiety. Based on this theoretical framework of foreign language anxiety, Hortwitz (1986) also developed the Foreign Language Classroom Anxiety Scale (FLCAS) to measure foreign language anxiety in a standardized way and the FLCAS has been proven to be a valid and reliable measure. The FLCAS contains 33 items that indicate communication anxiety, fear of negative evaluation and test anxiety.

The conceptualization of the foreign language anxiety construct, also known as foreign language classroom anxiety, and the FLCAS have advanced the empirical studies of foreign language anxiety and foreign language learning and generally the studies have supported foreign language anxiety as a different and separate construct from other types of anxiety (i.e., math anxiety) and can negatively influence foreign language learning (Yan & Horwitz, 2008).

However, Horwitz et al.'s (1986) foreign language anxiety construct measured by the FLCAS was not without its drawbacks. Some factor analysis in empirical studies had produced inconsistent results that suggested the component of foreign language anxiety might not include test anxiety. For example, with foreign language anxiety construct as the theoretical framework for the study, Aida (1994) conducted the FLCAS developed by Horwitz et al. (1986) to examine the relationship between foreign language anxiety and the learning of Japanese as a second language with 96 college students learning Japanese in the US. The factor solution in Aida's (1994) study partly supported the foreign language anxiety construct of Horwitz et al. (1986) by revealing the two factors (communication anxiety, fear of negative evaluation) as essential components of foreign language anxiety construct. However, Aida's (1994) study did not support Horwitz et al.' (1986) conceptualization of test anxiety as the important component of the foreign language anxiety construct as the three items indicative of test anxiety did not load on any factor. MacIntyre and Gardner' s (1989) finding, consistent with Aida's (1994), also indicated that test anxiety did not contribute to the communitive anxiety and not specific to the language classroom. Based on the findings, they suggested that test anxiety did not have a conceptual relation to other factors of the foreign language anxiety construct and the FLCAS could exclude the test anxiety items.

2.1.4 Foreign-language-skill-specific Anxiety

Foreign-language-skill-specific anxiety could be described as the negative attitudes and emotions developed during the process of acquiring each of the four language skills (Cheng et al., 1999). In an attempt to systematically and empirically examine the construct of foreign language classroom anxiety and foreign language writing anxiety, Cheng, Horwitz and Schallert (1999) concluded that foreign language classroom anxiety measured by the FLCAS is a more general kind of anxiety about learning a foreign language with a strong speaking element while foreign language writing anxiety appears to be a type of language-skill-specific anxiety, related but distinguished from the general foreign language anxiety. Cheng et al., (1999) also suggested that while some students might feel particularly nervous in foreign language speaking situations, listening, reading and writing in foreign language learning would provoke different levels of anxiety in foreign language learners. That is to say, the experience of foreign language learners acquiring each of the four language skills varies and their history of success or failure in performing the four language skills might lead to different emotions and attitudes toward each of the four language skills.

The turn of the 20th century also witnessed a new tendency in the field of foreign language anxiety research in which an increasing number of researchers investigated anxiety in specific language skills and developed scales to measure the anxiety specific to each language skill (He, 2018). Studies about the investigation of foreign language anxiety specific to listening (Kim, 2000), reading (Saito et al., 1999), writing (Cheng, 2004, 2017a) and speaking (He, 2018; Woodrow, 2006) have begun to surface.

2.1.5 Foreign Language Listening Anxiety

There is no existing definition of foreign language listening anxiety yet. However, researchers did acknowledge the existence of foreign language listening anxiety. For example, Meyer (1984) pointed out that the effect of task overload can cause apprehension in foreign language listening. Because listening is equally fast for students, and they have no control or little control over the listening speed, students simply cannot keep up and lag further behind the speaker. Even when they try desperately to decode, they miss the redundancies of real discourse that could help. “This word-to-word, beat-the-clock listening strategy squashes the listeners into a frustrating task overload failure (Meyer, 1984, p. 343)." Hortwitz et al. (1986) also discovered that listening could be as anxiety-provoking as speaking in testing situations. Hortwitz et al.'s (1986) study found that when students had trouble in differentiating the sounds and structures of the listening input and failed to comprehend the content of the extended foreign language utterances, they would be so apprehensive that they would not be able to comprehend all the input (Kim, 2001).

2.1.6 Foreign Language Speaking Anxiety

The one existing definition of the term "foreign language speaking anxiety" was first proposed by He (2018). In his book *Foreign Language Learning Anxiety in China*, He (2018) conceptualized FLSA as "an individual's fear or nervousness associated with either real or anticipated oral communication in foreign language with another person or persons" (He, 2018, p. 4).

2.2 Theoretical Foundations

2.2.1 Bachman and Palmer's Model

Bachman and Palmer (1996) in their book *Language Testing and Practice* Proposed a conceptual model for test development process. According to this model, test taker characteristics and test task characteristics affect and interact with each other. Consequently, test performance results from this interaction (In'nami, 2006). As shown in Figure 2.2, test takers' characteristics are composed of five components: (1) affective schemata, (2) strategic competence, (3) topical knowledge, (4) language knowledge, (5) personal characteristics. The former two interconnect and interact with the later three. According to Bachman and Palmer (1996), the affective schemata such as anxiety and motivation, combining with the characteristics of a specific task, to a large degree, determine the test takers' affective response to the task, and can “either facilitate or limit the flexibility with which he responds in a given context” (p. 65). Therefore, language test developers are advised to design tests to elicit test takers’ best performance by taking affective schemata into considerations and designing characteristics of test tasks that can improve comforting feelings and attitudes in test takers.

Affective schemata such as anxiety can occur during high-stakes tests such as the CELST and this negative emotion can, to a large extent, inhibit test performance. Realizing this, CELST test developers are advised to build test tasks and settings that can elicit test takers’ best performance by minimizing the effects of negative affective schemata.

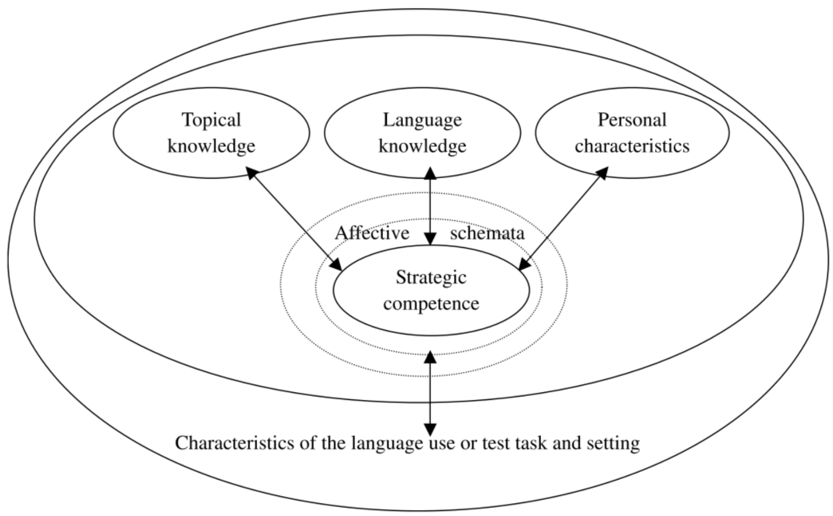

Figure 2. 2 Components of Language Use and Language Test Performance by Bachman and Palmer (1996). Source: from Bachman and Palmer (1996, p.63)

2.2.2 Tobias’ Model

In an effort to describe the effect of anxiety, Tobias (1986) proposed a model of how anxiety affects learning from instruction. The assumption of the model is that “since learning is a process that is essentially cognitively mediated, anxiety can affect learning only indirectly by impacting on the cognitive processes mediating learning at various stages”(Tobias, 1979, p.575). Tobias (1986) suggests that anxious persons tend to engage in derogatory cognition instead of focusing on the task at hand (MacIntyre & Gardner, 1989). The interference model, depicted in Figure 2.3, shows that interference may occur at three stages: input, processing, and output. Input represents presentation of instruction to students, such as display of content. Processing represents the process of students' encoding, organizing, and storing input. Output encompasses students' performance on any evaluative situations such as high-stakes tests.

According to Tobias (1986), at the input stage, highly anxious students divide their attention between task-related demands and ruminations stimulated by high anxiety. On the other hand, those lower in anxiety learn more since they allocate more attention to task demands and less to anxiety-related preoccupations. Interference at input suggests that highly anxious students tend to miss some proportion of input since they are occupied with task-irrelevant concerns such as worry. Interference at input is especially detrimental since the proportion of input not encoded cannot be subsequently processed. At the processing stage, three factors are likely to have important effects on cognitive processing of the instruction: difficulty, reliance on memory, and task organization. In short, the more difficult the task becomes, the more demand for calling for short-term memory and the less effective task organization can lead to a greater effect of anxiety on processing. At the output stage, anxiety may have interference in the retrieval of previously learned information. For example, a student might have studied diligently but "freeze up" on tests and fail to retrieve previously learned information due to test anxiety. Research studies have also provided empirical support for Tobias' model (MacIntyre & Gardner, 1994; Onwuegbuzie et al., 2000; Woodrow, 2006).

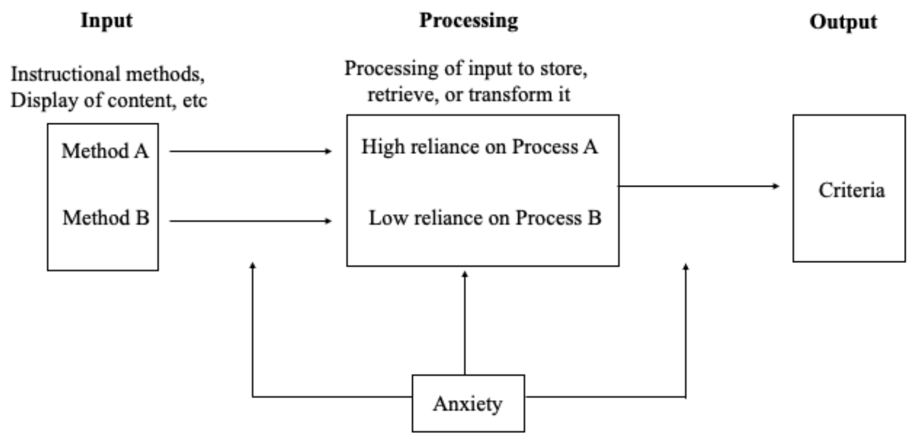

Figure 2. 3 Model Outlining the Effects of Anxiety on Learning from Instruction. Source: Adapted from Tobias (1979, p.575)

2.2.3 MacIntyre's Model

The theory of Tobias (1986) was used by MacIntyre (1989) to propose the mechanism by which anxiety may operate in foreign language context. MacIntyre (1989) developed three anxiety scales related to the three stages of cognitive processing and designed a series of experiments to verify Tobias's model. The results indicated that anxiety interfered with all three stages of cognitive processing of language learning, including decoding, storage and retrieval of language. Although anxiety can adversely affect cognitive processing of language learning and task performance, it should be noted that task performance can also cause anxiety. To give a better understanding of the effect of anxiety in foreign language learning, MacIntyre (1995) proposed that there is a recursive or cyclical relation among anxiety, cognition, and behavior.

As is depicted in Figure 2.4, anxiety, cognition, and behavior form recursive or cyclical relations where each influences the other (Levitt, 2015; MacIntyre, 1995). For example, the behavior of finishing a task in a foreign language class or test may cause anxiety which leads to task-irrelevant cognition such as worry and rumination. The task-irrelevant cognition causes a decline in cognitive performance and further causes a decline in task performance. The declines in performance may lead to more negative self-evaluation and anxiety which further impairs performance and so on. Therefore, the cyclical relation between anxiety and task performance suggests that the more failure in performance students experience, the more anxiety they experience, and the more their performance on foreign language suffers, which forms a vicious cycle.

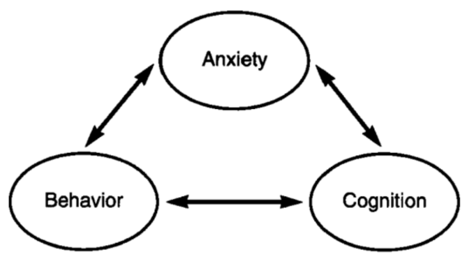

Figure 2. 4 Recursive Relations among Anxiety, Cognition, and Behavior. Source: MacIntyre (1995, p.93)

Relating the above three models to the current study which attempted to investigate the effect of anxiety on the CELST test takers, Bachman and Palmer's (1996) model suggests that it is crucial to take into account the affective schemata such as anxiety when designing the CELST so as to elicit test takers' best performance and improve construct validity. Tobias' model provides the theoretical foundation that anxiety can have an effect on the input and the processing of the CELST listening materials and the output of the information of the CELST test tasks. MacIntyre's model informs us that if the CELST test takers suffer from anxiety in the CELST, this may cause performance declines and in turn provoke more anxiety and create a vicious cycle. Therefore, it appears that it is indispensable to investigate the effect of anxiety in the CELST in order to facilitate students' foreign language learning and test performance. To give a more comprehensive understanding of the effect of anxiety, the following parts review previous empirical studies of test anxiety, foreign language listening anxiety and foreign language speaking anxiety.

2.3 Studies of Test Anxiety

2.3.1 Studies of Test Anxiety in General Test Performance

The society today is highly test-oriented and test-consuming. Considering the consequences and stakes of test performance might determine people's lives, such as admittance to a top university, it is not surprising that test anxiety is pervasive among students (Zeidner, 1998). Students with test anxiety tend to suffer from cognitive and affective behaviors in evaluative situations: such as self-derogatory thoughts, excessive worry over a test, dizziness, and nausea (Cassady & Johnson, 2002; Sarason & Stoops, 1978).

The primary focus of test anxiety research has been on the relationship between test anxiety and performance measures. More specifically, researchers have attempted to investigate the correlation between performance and different components of test anxiety (cognitive test anxiety and emotionality). Until now, researchers have generally identified the cognitive component of test anxiety as the primary factor that causes a decline in performance. For instance, two meta-analyses (Hembree, 1988; von der Embse et al., 2018) of test anxiety have identified the cognitive facet of test anxiety being the primary component that most consistently correlated with decrements in performance in important educational outcomes, including university entrance examinations, standardized tests and grade point average. In addition to meta-analysis, results from path analysis also confirmed that the cognitive component of test anxiety had the strongest relationship with performance decrements, whereas no significant influence has been found for the emotionality component (Cassady & Johnson, 2002; Hembree, 1988; Williams, 1991).

Although a negative correlation has been found between test anxiety and performance in many studies, it does not always remain true across the board (Tsai & Li, 2012). For example, Ruebush (1960) from Yale University administered the Yale Test Anxiety Scale to 280 elementary school students and found that those with higher levels of test anxiety have better performance than their lower anxiety level peers. A high level of emotionality component of test anxiety such as faster heartbeat was even shown to be facilitating for exams when the students have a high level of confidence in regard to test performance (Cassady & Johnson, 2002). McDonald (2001) also argued that an optimal level of anxiety is helpful to let students maintain focus on the task at hand and keep them feeling motivated and challenged. From this perspective, test anxiety might not always harm performance.

Researchers have also been interested in test anxiety and its relationship with gender and self-efficacy. Females have been frequently reported to experience more overall test anxiety than males (Hembree, 1988; von der Embse et al., 2018). There are different explanations for this gender difference. For example, Zeidner (1990) asserted that the gender difference might be due in large part to different scholastic ability, that is, females tended to have lower performance and thus became more test anxious. However, other researchers suggested that the reason for the gender difference may be that females tend to perceive testing situations as more threatening than males do. This perception of threat by females might come from self-doubt regarding their competence to perform well in a testing situations(Arch, 1987; Zohar, 1998). In addition, test anxiety has been reported to be negatively correlated with self-efficacy. For example, nine different studies in Germany reported a negative correlation ranging from -.30 to -. 66 between self-efficacy and test anxiety(Jerusalem & Schwarzer, 1992).

Therefore, research into the effect of test anxiety on general test performance has generally recognized that test anxiety, especially cognitive test anxiety has a negative effect on students' test performance while conflicting results have also been found that supported that test anxiety can facilitate performance. In addition, females have been reported to have higher levels of test anxiety than males and self-efficacy has been shown to negatively correlate with test anxiety.

2.3.2 Effect of Test Anxiety on Foreign Language Test Performance

Although research into test anxiety in general testing situations is abundant, empirical studies that specifically focus on test anxiety in foreign language learning and evaluative situations have been quite limited(Cheng et al., 2014; Huang & Hung, 2013; In'nami, 2006; Salehi & Marefat, 2014; Tsai & Li, 2012; Zheng & Cheng, 2018). Besides, the few studies conducted so far have produced conflicting findings. Most of the studies found a negative effect of test anxiety, while others found no significant effect of test anxiety on language performance.

In the early years of study, Horwitz (1986) adopted the FLCAS (containing only three test anxiety items) to measure introductory-level foreign language students' test anxiety and found that test anxiety formed a small and negative correlation with their final grades ($r = -.16$). However, another rigorous empirical study conducted by In'nami (2006) did not find a significant relationship between test anxiety and listening test performance. In'nami (2006) administered the Test Anxiety Scale (Sarason, 1975) and Test Influence Inventory (Fujii, 1993) to 79 Japanese university students. The results of the structural equation model revealed no relationship between test anxiety and listening test performance as measured by an adapted TOEFL practice listening test. In'nami (2006) attributed the non-relationship to three factors: (a) test takers' high English proficiency level, (b) strategic competence that deals with anxiety, (c) the low stakes nature of the listening test results. In two recent studies, test anxiety also has been shown to not significantly correlate with low-stakes test performance in foreign language reading (Amiryousefi & Tavakoli, 2011; Javanbakht & Hadian, 2014). In contrast, a recent study conducted by Salehi & Marefat (2014) showed that both test anxiety and foreign language anxiety negatively correlated with final exam scores of 200 Iranian EFL learners (193 participants were female). Test anxiety was also demonstrated to have a strong positive relationship with the foreign language classroom, indicating the more foreign language anxiety they have, the more students suffer from test anxiety.

Related Test Anxiety Studies in China. Studies that focus on test anxiety and foreign language performance started late in China. Although studies concerning test anxiety and foreign language test performance have been scarce, there appears to be a trend in the past decade that research in China started to investigate the effect of test anxiety and other variables such as foreign language anxiety or skill-specific anxiety in one single study(Cheng et al., 2014; Tsai & Li, 2012; Zheng & Cheng, 2018).

With 302 EFL university students as participants, Tsai & Li (2012) from Taiwan assessed the influence of test anxiety and foreign language reading anxiety on a simulated English proficiency test. The results revealed that test anxiety not only correlated with the test performance on a simulated test but also positively correlated with foreign language anxiety measured by the FLCAS. This suggested that test anxiety had a negative impact on test performance and more test-anxious students also tended to have higher levels of foreign language anxiety. Other Taiwanese researchers Huang & Hung (2013) investigated the effect of test anxiety on a simulated integrated listening and speaking test among 352 EFL university students in Taiwan. The results of path analysis data revealed that test anxiety significantly and negatively affected the integrated task.

In a large-scale international study, Cheng et al.(2014) reported skill anxiety, a factor of cognitive test anxiety, was a significant negative predictor of high-stakes test performance. In their studies, Cheng et al. (2014) attempted to investigate to what extent test takers' motivation and test anxiety affect test performance across a range of educational and social conditions in three high-stakes language tests: the Canadian Academic English Language (CAEL) Assessment in Canada, the General English Proficiency Test (GEPT) in Taiwan and the College English Test (CET) in mainland China. The results revealed that skill anxiety was a significant predictor across all three tests, indicating students with higher levels of anxiety tended to have lower performance. Their study also revealed that different test importance and test purposes across three tests resulted in different levels of test takers' anxiety and motivation which in turn influenced the test takers' performance to a varying degree. Another study by Zheng & Cheng (2018) also supported the finding that cognitive test anxiety was a significant negative predictor of CET scores. However, the interview results revealed that most students did not perceive foreign language classroom and testing situations to be very anxiety-provoking.

2.3.3 Measurement of Test Anxiety

The most used test anxiety scales in this field tend to be the Test Anxiety Scale (TAS) by Sarason (1975) and the Cognitive Test Anxiety Scale (CTAS) by Cassady and Johnson (2002). Although Sarason's (1975) TAS has been widely used and has high reliability, the present study adopted the CTAS to be the measure of test anxiety, specifically, the cognitive component of test anxiety.

As mentioned in section 2.3.1, the meta-analysis and path analysis studies both revealed that the cognitive facet of test anxiety has been shown to negatively affect test performance while the emotionality component of test anxiety generally did not exert any effect. Therefore, using a scale such as the TAS that does not differentiate cognitive component and emotionality component might attenuate the relationship between test anxiety and performance. In'nami (2006) also argued that "since test anxiety is a multi-componential construct, investigating how each component is related to performance measures may be more important than investigating how test anxiety overall is related to performance measures" (p. 321). Therefore, the present study adopted the CTAS as the measure of the cognitive component of test anxiety. The original CTAS has been shown to have high internal consistency (.91) and its concurrent validity has been proven by comparison with the well-researched test anxiety inventory: Sarason's (1984) Reaction to Tests Scale and Spielberger's (1983) Test Anxiety Inventory. The present study utilized the CTAS adapted by Zheng and Cheng (2018) who translated it into Chinese and administered it to 921 university students. The adapted version of CTA contained 20 items and was shown to be a valid measure of test anxiety in Zheng and Cheng's (2018) study.

2.3.4 Summary of Test Anxiety Studies

To sum up, empirical studies of test anxiety in foreign language test performance are relatively scarce with contradictory findings. Some studies reported test anxiety has a negative effect on test performance (Cheng et al. 2014; Huang & Hung 2013; Tsai & Li 2012; Salehi & Marefat 2014), while others showed that test anxiety exerted no effect on foreign language performance (Amiryousefi & Tavakoli, 2011; In'nami 2006; Javanbakht & Hadian, 2014). Thus, the relationship between test anxiety and foreign language test performance is still undefined. Besides, previous meta-analysis and path analysis studies revealed that cognitive test anxiety was the primary factor that caused declines in general test performance while the emotionality component generally did not exert any effect. (Cassady & Johnson, 2002; Hembree, 1988; von der Embse et al., 2018)

Even though the above-mentioned studies have contributed to advance research in this field, a few noticeable limitations should be pointed out:

(1) All of the above-mentioned studies of test anxiety in foreign language context used performance measures that are low-stakes in nature, except Cheng et al. (2014) and Zheng & Cheng (2018). The test results were not part of course grades, nor did they impact students' future by any means. Therefore, the weak correlation or no correlation between test anxiety and foreign language test performance might be due to the fact that students' performance was not negatively affected by test anxiety owing to the low-stakes and low-pressure testing situations. **(2)** The scale used (Horwitz, 1986) to assess test anxiety is short of items that indicate test anxiety. Using a scale with an insufficient number of items might not fully represent the construct of test anxiety and capture its true effect. **(3)** Very few studies have adopted qualitative methods to dig deeper into reasons leading to test anxiety and strategies to alleviate it **(4)** Participants in all of the studies are university students. Subjects like high school students who are most susceptible to test anxiety due to the high pressure of preparing for the college entrance examination have been scarcely examined.

2.4 Studies of Foreign Language Listening Anxiety

2.4.1 Effect of Foreign Language Listening Anxiety on Foreign Language Listening Performance

In the 1990s, studies that focused on the effect of foreign language listening anxiety were very scarce. However, recent years have seen an increasing number of researchers starting to explore how FLLA influences foreign language listening performance.

Kim (2001), one of the pioneer researchers in the FLLA research field, investigated the relationship between foreign language listening proficiency and FLLA with a sample of 253 Korean EFL university students. Participants' listening proficiency was assessed by a practice TOEFL listening test and the FLLA level was measured by a self-developed FLLA scale. The study revealed that a majority of participants did experience anxiety both in response to the listening test and in a foreign language classroom. Correlation analysis indicated a significant moderate negative relationship between foreign language proficiency and FLLA ($r = -.36$). Factor analysis revealed two factors of the FLLA scale: lack of self-confidence in listening, and tension and worry over English listening. The multiple regression analysis revealed that lack of confidence in listening was the best predictor of the listening proficiency. Kim (2001) also examined the correlation between FLLA and general foreign language anxiety as measured by the FLCAS. The squared correlation ($r^2 = .50$) demonstrated variation in the FLCAS only explained half of the variances of FLLA scale, which indicated that FLLA seemed to be a construct that was not only correlated but also relatively independent from general foreign language anxiety. The retrospective interview revealed factors that cause listening anxiety which were divided into three categories: characteristics of text, personal characteristics, and characteristics related to the listening process. Kim's (2001) finding echoed with Yoongoo's (2017) which found that FLLA not only significantly correlated ($r = -.44$) but predicted listening scores of 226 Korean EFL university students. Yoongoo (2017) also found that female students were slightly more anxious than males in foreign language listening.

In a non-EFL context, Elkhafaifi's (2005) found a high and significant correlation ($r = -.70$) between FLLA and the course grades of Arabic listening comprehension among 223 university students learning Arabic as a foreign language. No gender difference in FLLA was found in this study. It should be noted that the FLLA scale in this study was adapted from Saito et al.'s (1999) Foreign Language Reading Scale, which was not designed specifically to measure FLLA. Thus, whether the measure of FLLA was accurate in this study was uncertain. Elkhafaifi (2005) finding was also supported by Serraj & bt. Noordin's (2013) study which revealed that FLLA significantly and negatively correlated with IELTS listening comprehension of 210 Iranian students (r = -.41). In an attempt to investigate FLLA and its relationship with listening comprehension and strategy use, Golchi (2012) administered Kim's (2001) FLLA scale, Listening Comprehension Strategy Questionnaire and IELTS listening test to 63 Iranian EFL students. The results revealed that: (1) FLLA had a significant negative correlation with listening comprehension (r = -.63) and listening strategy use (r = -.32); (2) female students had a higher level of anxiety than male students; (3) highly anxious students adopted less metacognitive strategies than lowly anxious students; (4) students with more years of studying English tended to be less anxious.

Related Foreign Language Listening Anxiety Studies in China. Chinese researchers have also started to look into the effect of FLLA on listening performance in recent years. In the early stage of FLLA research in China, Yu (1999) used the FLCAS to measure 49 adult EFL learners' foreign language classroom anxiety and found that there was no significant relationship between students' foreign language classroom anxiety and their listening comprehension scores on the final exam. Zhou (2003) investigated the effect of FLLA on 96 English major students and found a significant and high correlation (r = -.55) between FLLA and a practice listening test. However, participants overall didn't experience too much anxiety in the classroom. Another researcher Zhou (2009) used a self-developed scale to measure 105 non- English major students' FLLA and found that FLLA negatively correlated with listening test scores (r = -.62). Afterward, Chen and Liu (2010) developed a scale based on Kim (2001) to measure 147 low-proficiency non-English majors' FLLA and found a negative correlation between FLLA and students' simulated CET-4 listening scores ($r = -.54$). Factor analysis of Chen and Liu (2010) revealed five factors that cause FLLA: insufficient listening skills, lack of background knowledge on unfamiliar topics, task difficulty, low listening proficiency and worry.

More rigorous empirical studies have been conducted recently by Zhang (2013) and Liu (2016). Zhang (2013) administered the FLLA scale, adapted from Elkhafaifi's (2005), to 300 first-year English majors and the result of the exploratory causal model revealed that FLLA could negatively impact students' performance on a simulated IELTS test. However, participants' FLLA was not affected by their listening performance systematically. In more recent years, Liu (2016) administered an adapted version of Elkhafaifi's (2005) FLLA scale, Foreign Language Listening Strategies Use Scale and a classroom English listening test to 1160 freshman non-English majors. The results revealed that students from the high-proficiency group were significantly less apprehensive and more confident in their English listening competence. Besides, FLLA had a significant negative correlation with foreign language listening strategies use for both high- and low-proficient groups. This indicated that for both high- and low-proficiency groups, the more FLLA they had, the fewer listening strategies they tended to use.

2.4.2 Measurement of Foreign Language Listening Anxiety

As the review in section 2.4.1 has shown, several FLLA scales have been developed and utilized to measure FLLA. Among those FLLA scales, the one developed by Kim (2001) and the one developed by Elkhafaifi (2005) appear to be the most used ones in the FLLA research field. The FLLA scale developed by Kim (2001) contains 33 items that specifically measure Korean university students' FLLA. It has been shown to have an acceptable level of internal consistency

(Cronbach's α = .90) in studies that it was used to assess Korean university students' FLLA. The FLLA scale developed by Elkhafaifi (2005) has 20 items. It was adapted for Arabic learners' listening anxiety from Saito et al.'s (1999) Foreign Language Reading Anxiety Scale (FLRAS). The word "reading" in the items of FLRAS was replaced by the word "listening". The word "Arabic" replaced the words "Japanese", "Russian", "French" in the FLRAS items. According to Elkhafaifi (2005), the FLLA scale has demonstrated a decent level of reliability and high internal consistency (Cronbach's α = .96). However, because the psychometric properties of the Elkhafaifi's (2005) scale were not checked before and the scale were adapted from a foreign language reading scale, it might not be an accurate measure of FLLA. It should also be noted that the two scales mentioned above were used to measure FLLA of Korean EFL learners (Kim, 2001) and American Arabic Learners (Elkhafaifi, 2005) in university. Therefore, the two scales may not be appropriate to measure the FLLA of Chinese EFL high school students, the participants in the present study.

In light of the need to measure Chinese students' language-skill-specific anxiety, Cheng Yuh show (2017), a well-known researcher in the field of foreign language anxiety research, developed four brief measures of language-skill-specific anxiety: foreign language listening, speaking, reading, and writing anxiety scales. For the present study, the FLLA scale and FLSA scale developed by Cheng Yuh show (2017) were adopted to measure test takers' FLLA and FLSA respectively.

I will elaborate on how Cheng (2017) developed the four brief scales. The rationale for adopting Cheng's (2017) FLLA scale and FLSA scale as the measurement tools for the present study is also illustrated.

In her paper published in *System,* Cheng (2017) reported the development and validation of these four 9-item scales. A total of 523 first-year university students from nine universities in Taiwan participated throughout the whole process of item generation and scale development and validation. Based on a review of relevant literature and results of a focus group interview, an initial pool of items was developed. Cheng (2017) conducted an exploratory factor analysis to select items for each anxiety scale, which was further validated by confirmatory factor analysis and correlation analysis. The final results provided sufficient evidence for the reliability, convergent validity, and discriminant validity of the scores of the four scales. Regarding the usefulness and multidimensional nature of anxiety and Cheng's (2004) success at translating Lang's three dimensional anxiety theory into the measurement of foreign language writing anxiety, Cheng's (2017) study adopted Lang's theory as the guiding principle to develop the four scales. Based on Lang's (1971) tripartite model, each of the four scales consists of three dimensions: cognitive, somatic and behavioral. As cited in Cheng (2017), it should be noted that Lang's theory resonates well with Horwitz's (1986) conceptualization of foreign language anxiety construct as "negative performance expectancies and social comparisons, psychophysiological symptoms, and avoidance behavior" (p.559).

As Cheng (2017) pointed out, the development of these four efficient scales allows researchers to conduct large-scale surveys and longitudinal studies on language-skill-specific anxiety with fewer resources and less time. These brief measures also allow researchers to extend the scope of their studies by incorporating more variables in one survey of reasonable length. Because the present study attempted to examine the effects of three anxiety variables (CTA, FLLA, FLSA), using the existing scales such as Kim's (2001) 33-item FLLA scale combined with the CTA scale and FLSA scale will require a questionnaire of approximately 100 items. A lengthy questionnaire containing around 100 items might sacrifice data quality "due to response burden and fatigue effect" (Cheng 2017). Besides, due to high pressure to prepare for the National College Entrance Examination, senior three students have a demanding schedule, and thus it was hard to invite a large group of senior three students to participate and fill in a lengthy questionnaire. Therefore, it appears that adopting Cheng's (2017) efficient scales (the FLLA scale and FLSA scale) with reasonable length and high validity as the measures of FLSA and FLLA would be the best option for the current study.

2.4.3 Summary of Foreign Language Listening Anxiety Studies

To conclude, most of the studies showed that FLLA had a moderate negative effect on students' test scores and their listening strategy use. Yu's (1999) study was the only one to have a non-significant effect. However, Yu's (1999) finding is not convincing because the sample size was too small (n = 49) and the use of the FLCAS as the measure of FLLA was not accurate. In addition, some studies found that females tended to be more anxious than males in listening (Golchi, 2012; Yoongoo, 2017), while others found no gender difference (Elkhafaifi, 2005). Besides, self-perception of listening ability also appeared to have an effect on FLLA (Kim, 2001).

While acknowledging their contributions to this relatively new research field, the review of the above studies of FLLA in and out of China also reveals some limitations:

(1) The measurement tool was not designed specifically to measure FLLA (Elkhafaifi, 2005; Liu, 2016; Yu, 1999; Zhang, 2013). This might reduce the validity of the measure of FLLA and make the findings less convincing. **(2)** All of the FLLA studies mentioned above used low-stakes course exams or adapted listening tests administered by the researchers. The validity and reliability of those tests may not be as valid as high-stakes standardized tests, and thus the test scores might not precisely reflect students' listening performance. Besides, whether those low-stakes practice listening tests can arouse a considerable amount of anxiety in the listening process and negatively affect performance is doubtful. Therefore, the correlation between listening performance and FLLA should be treated with caution. **(3)** Most of the studies only adopted correlation analysis to investigate the effect of FLLA. A simple correlation analysis might not be able to capture the true effect of FLLA. Additionally, very few studies except Kim (2001) have employed qualitative methods to probe deeper into the reasons that cause FLLA and strategies to alleviate it. **(4)** Almost all of the studies took English- or non-English major university students as participants. Therefore, the finding cannot be generalized to students at different stages in education such as high school students.

2.5 Studies of Foreign Language Speaking Anxiety

2.5.1 Effect of Foreign Language Speaking Anxiety on Foreign Language Oral Performance

It is always suggested that speaking is considered the most anxiety-provoking aspect in a foreign language situation (Cheng et al., 1999). According to Phillips (1992), the anxiety students experience may have a debilitating effect on their capacity to deliver a speech in a foreign language. Unfortunately, the ever-increasing use of communicative spoken testing will most likely escalate foreign language speaking anxiety (Phillips, 1992).

Attempting to determine the effect of anxiety on students' performance on a French spoken course exam, Phillips (1992) administered the FLCAS to 44 university students learning French as a foreign language. The results indicated a moderate negative correlation between FLSA and the performance measured by the exam grades (r=-.40, p,.01). That is to say, students who experienced less anxious feelings tended to receive higher oral exam grades than their highly anxious classmates. Additionally, the finding of Phillips (1992) also revealed a significant negative correlation between anxiety and the percent total words in communication units (CUs), which means students with a higher anxiety level tended to say less. Phillips (1992) further conducted interviews with six anxious students about their attitudes towards oral language tests and revealed that across ability boundaries, those students reported negative attitudes towards oral exams by using words such as “going blank”, “intimidated”, “worried” and “dumbfounded”.

In an English as a second language (ESL) context in Australia, Woodrow (2006) constructed the Second Language Speaking Anxiety Scale to investigate the effect of second language speaking anxiety on 275 advanced English for academic purposes (EAP) students with students from China being the largest group in the sample (34%, $n=92$). The results indicated that the oral performance measured by a simulated IELTS oral exam negatively correlated with in-class ($r=-.23$, $p = <.01$) and out-of-class ($r=-.24$, $p = <.01$) second language speaking anxiety.

Salem & Al Dyiar (2014) used a six-item speaking scale to examine the correlation between FLSA and EFL oral fluency with a sample of 121 Arab special education students with learning difficulties. The results showed that the oral fluency skill score measured by a semi-structured interview was found to be negatively correlated with students’ speaking anxiety ($r=-.261$, $p<.01$). Using a Personal Report of Public Speaking Anxiety Scale, Hasibuan & Irzawati (2020) also found a significant relationship ($r=-.358$, $P<.001$) between speaking anxiety and 87 Indonesian university students’ speaking performance measured by a role-play speaking activity. In a Turkey EFL context, Ça (2015) administered the FLSA scale adapted from the FLCAS to 147 university students and found that FLSA did not differ across four proficiency levels. Besides, females were shown to have significantly higher FLSA than males with a moderate effect size.

The above studies investigated the relationship between FLSA and the objective EFL/ESL achievement. However, researchers have also been interested in the relationship between FLSA and self-perceived achievement or competence. For example, it was reported that perceived English speaking competence strongly correlated with students' FLSA as measured by the FLCAS, which has a strong speaking element(Chiang, 2006; Truitt, 1995). Language anxiety has even been shown to have a stronger correlation with subjective proficiency or perceived competence than objective proficiency measures (MacIntyre et al., 1997). In MacIntyre et al.'s (1997) study, the regression analysis with the actual competence level controlled revealed that students with more anxiety tended to underestimate their competence compared to students with less anxiety.

Related Foreign Language Speaking Anxiety Studies in China. In the context of China, empirical studies that focused on the effect of FLSA on foreign language performance also started late and were rather limited. Among the few studies, Cheng, He and Han (2007) conducted an experiment to investigate whether foreign language classroom anxiety affected university students' English oral performance. During the experiment, 29 students from the experimental group were asked to fill out Horwitz et al. 's (1986) FLCAS and give an English public speech in front of the class for each month during the 4-month experiment period. Their speaking performance was assessed by the instructor based on the rubric: syntactic complexity, fluency, and lexical range. The results indicated that students' public speaking performance had a significant and negative correlation with both foreign language classroom anxiety (-.333) and the sub-scale of the FLCAS, speech anxiety ($r=-.354$). In addition, no gender difference was found between the foreign language classroom anxiety, speech anxiety and negative evaluation.

With a sample of 109 non-English majors, Lu (2010) administered the adapted FLCAS and found that the students experienced high levels of speaking anxiety, which was negatively correlated with their oral exam scores ($r = -.465$) and their speaking strategies use ($r = -.764$).

Another empirical study that systematically explored the effect of FLSA on Chinese university EFL students was conducted by He (2018). He (2018) developed his version of the FLSA scale and administered both FLSA scale and Horwitz et al. 's (1986) FLCAS to 302 non-English majors. The results indicated that the students' general foreign language anxiety measured by the FLCAS was not high while overall they had a comparatively high level of FLSA. The interview data were also congruent with the qualitative data because 15 out of 33 participants regarded speaking as the most anxiety-provoking activity. The results of correlation analyses revealed that students' FLSA levels significantly and negatively correlated with self-evaluated English-speaking proficiency (r=-.42). However, the FLSA level was not significantly correlated with the NMET scores as the exam did not test the speaking skill (r=.10). The interview data also revealed reasons that lead to FLSA, including frequency of language use, language testing and context of language acquisition etc.

2.5.2 Measurement of Foreign Language Speaking Anxiety

There are several major scales that have been adopted to assess students' foreign language speaking anxiety. The most widely used scale is the Foreign Language Classroom Anxiety Scale (FLCAS) developed by Horwitz et al. (1986). It was designed on the basis of the foreign language anxiety construct and aimed to measure three components of foreign language anxiety, that is, communication anxiety, fear of negative evaluation and test anxiety. The FLCAS contains 33 items that utilize a five-point Likert scale. Most of the empirical studies revealed that the FLCAS had a high internal consistency and supported it as a valid tool to measure general foreign language classroom anxiety. However, it should be noted that the FLCAS might not be able to measure students' FLSA accurately. Even though FLCAS has around one third of the items assessing students' FLSA, the other items are measuring the general foreign language anxiety in classroom. Researchers have also raised doubts about the adequacy of the FLCAS to assess the FLSA level. For instance, Aida (1994) attributed the result of a moderate relationship (r=-.38) between the foreign language anxiety performance and anxiety to the attenuating influence of performing an instrument that did not specifically assess oral skills. Similarly, the finding of Phillips (1992) also suggested that the FLCAS did not measure students' anxiety associated with the specific oral exam. Phillips (1992) also claimed that the more specific the anxiety measure is to the performance measure, the more likelihood there is of a correlation.

As justified in section 2.4.2, the FLSA scale developed by Cheng (2017) would be a better option for the present study. First, the reliability, discriminant validity, convergent validity of the scale have been proven to be high in Cheng's (2007) study published in *System.* Second, it was developed on the basis of a pool of Chinese EFL students to specifically measure students' FLSA. Third, it is a relatively efficient measure of FLSA which does not result in a lengthy questionnaire when combined with the CTA scale and FLLA scale. Concerning the reasons mentioned above, the current study adopted Cheng's (2017) FLSA scale as the measure of participants' FLSA level.

2.5.3 Summary of Foreign Language Speaking Anxiety Studies

To sum up, across different contexts, most of the studies revealed a moderate negative correlation between FLSA and oral performance and generally students regarded speaking as the most anxiety-arousing skill among the four language skills. Gender (Ça, 2015), self-perceived competence (Chiang, 2006; Truitt, 1995), frequency of language use (He, 2018) were also reported to impact FLSA. However, a few limitations of the above studies should be noted:

(1) Some of the scales such as the FLCAS were not designed specifically to measure FLSA, which raises questions about the validity of the measure of FLSA. **(2)** Similar to the limitations of test anxiety studies and FLLA studies, the speaking performance in FLSA studies was measured by some low-stakes tests such as course oral exams or speaking activities in class. It raises doubts about the validity of the performance measures such as course exams. Therefore, the correlation in the previous studies between FLSA and oral performance should be treated cautiously. **(3)** The participants of the studies above were mostly university students. High school students are scarcely researched. Therefore, rigorous empirical studies that investigate the effect of FLSA on high school students with a dedicated FLSA scale in high-stakes testing situations are highly needed across the globe.

2.6 Summary of Literature Review

The above literature review first defines the key terms of anxiety, test anxiety, foreign language anxiety, language-skill-specific anxiety, FLLA, FLSA. Then, the models that provide theoretical foundation for this study are illustrated to show how anxiety might impact the CELST test takers. The literature review then presents the empirical studies about the effects of test anxiety, FLSA and FLLA on performance and the scales to measure CTA, FLLA and FLSA. Results of most empirical studies in FLSA and FLLA generally revealed that FLSA and FLLA have a moderate negative effect on foreign language performance while studies on test anxiety showed conflicting findings with some reporting a significant negative effect of test anxiety on foreign language performance and others revealing a non-significant effect.

Even though the above reviewed empirical studies have contributed to a finer understanding of how test anxiety, FLSA and FLLA affect foreign language performance, a few limitations and literature gaps should be noted here.

(1) Among all of the reviewed empirical studies of test anxiety, FLLA and FLSA, very few studies, except Cheng et al. (2014) and Zheng and Cheng (2018), have investigated the effect of test anxiety, FLSA or FLLA in high-stakes testing situations. Using course exams or any

adapted version of standardized tests might not be as valid and reliable of a performance measure as real standardized tests, and thus the test scores might not fully capture participants' true proficiency. Besides, the low-stakes nature of the performance measure did not impact participants' life in any way, which calls into questions whether participants' performance is adversely affected by anxiety due to the low-stakes contexts. Therefore, the correlation between anxiety and foreign language performance in the previous studies should be interpreted with caution.

(2) The sample size of most of the studies is relatively small. The number of participants in some studies ranges from 44 to 150 (Cheng et. al, 2007; Chen & Liu, 2010; Hasibuan & Irzawati, 2020; In'nami, 2006; Phillips, 1992; Salem & Al Dyiar, 2014; Yu, 1999; Zhou, 2003; Zhou, 2009), others' range from 150 to 300 (Elkhafaifi, 2005; He, 2017; Kim, 2001; Salehi & Marefat, 2014; Serraj & bt. Noordin, 2013; Tsai & Li, 2012; Woodrow, 2006; Yoongoo, 2017; Zhang, 2013). Very few studies have a total of participants of more than 400. A small sample size may affect the reliability of the survey results and limits the generalization of the findings.

(3) Most of the studies in this field only adopted quantitative methods to investigate the effect of anxiety on performance. However, given the complex nature of anxiety, quantitative methods might not be able to give an in-depth understanding of the effect of anxiety. It appears that quantitative methods combined with qualitative methods would be a better option to fully explore how anxiety impact students' performance.

(4) Almost all of the published studies recruited university students as their participants. High school students, who might suffer most from high levels of anxiety to prepare for the NMET, have been scarcely examined.

(5) To the best of my knowledge, none of the studies have investigated test anxiety,

FLLA or FLSA in one single study. Although in recent years there has been a trend to investigate both general foreign language anxiety and test anxiety in one study (Salehi & Marefat, 2014; Zheng & Cheng, 2018), there have been very few studies examining the effects of the language-skill-specific anxiety and test anxiety in one study, let alone in the context of the CELST. There are two advantages when investigating the two types of anxiety. First, the effect of two types of anxiety on test performance can be compared to see which one is more prominent. Second, whether correlations exist between these two types of anxiety can be examined (Salehi & Marefat, 2014).

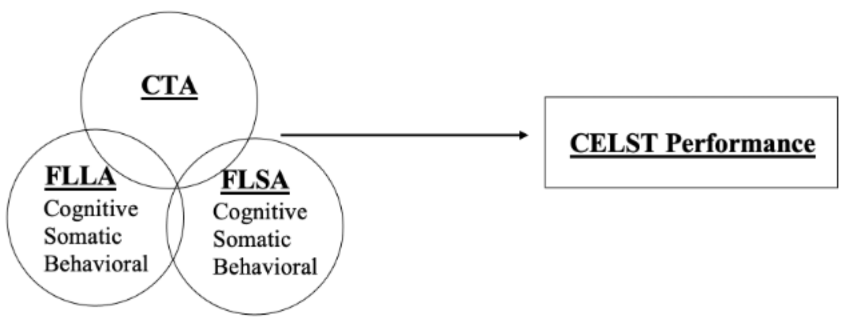

Figure 2. 5 Conceptual Framework of the Present study

In view of the above limitations, the main purpose of the current study is to bridge the literature gap by investigating the effects of test anxiety, FLLA and FLSA on a high-stakes foreign language listening and speaking test among 500 EFL senior three student test takers. Based on the literature review, a conceptual framework was established as above (see Figure 2.5). The independent variable is CTA, FLLA and FLSA, which are independent yet related to each other. As the study adopted Cheng's (2017) multi-dimensional FLLA and FLSA scales, FLLA and FLSA in this study both have three dimensions: cognitive, somatic and behavioral. The dependent variable is the test takers' performance on a large-scale high-stakes CELST, as measured by their CELST scores.

Chapter Three Methodology

This chapter introduces the research methodology. Firstly, it introduces the rationale of the research design and the two stages of this study. Secondly, it presents the information about the participants, instruments, data collection procedures and the quantitative and qualitative data analyses.

3.1 Research Design

The purpose of the present study was to investigate the characteristics of CTA, FLLA and FLSA among the CELST test takers and the effects of CTA, FLLA and FLSA on CELST test takers' performance. In addition, the study also attempted to explore possible reasons leading to anxiety in the CELST and strategies to alleviate it through the test takers' perspectives. Therefore, three research questions were formulated as follows:

1. What are the characteristics of cognitive test anxiety, foreign language listening anxiety and foreign language speaking anxiety among the CELST test takers?
2. What are the effects of cognitive test anxiety, foreign language listening anxiety and foreign language speaking anxiety on CELST test takers' overall scores?
3. From the test takers' perspectives, what reasons may lead to anxiety in the CELST and how do test takers alleviate the anxiety?

Previous anxiety research has mostly used a quantitative method. However, this study adopted a sequential mixed methods research design. The sequential mixed-method approach has advantages over other approaches in anxiety research. Using a combination of qualitative and quantitative research approaches provides a more comprehensive understanding of the effect of anxiety on the CLEST test takers. Given the complex nature of psychological factors like anxiety, using either a qualitative or quantitative approach might not be able to capture a comprehensive picture of the study phenomenon. A mixed method can offset weakness in a single method and provide stronger inferences. Second, a sequential mixed-method approach is useful when unanticipated quantitative findings emerge (Bryman, 2006). For instance, unusual findings from the quantitative survey can be explained by the follow-up interviews. Therefore, the sequential mixed-method approach is best suited to answer the research questions in this study.

The study consisted of two stages. In the first stage of the study, to collect quantitative data, the researcher constructed and administered a questionnaire which contains four sections: demographic information, Cognitive Test Anxiety Scale (CTAS), Foreign Language Listening Anxiety Scale (FLLAS) and Foreign Language Speaking Anxiety Scale (FLSAS). Quantitative data from stage one was used to examine the characteristics of CTA, FLLA and FLSA among test takers and the effects of the three types of anxiety on test takers' CELST performance. In the second stage of the study, the researcher collected qualitative data by conducting individual semi-structured interviews with twelve test takers. The qualitative data from stage two was used to complement the findings of the quantitative method. Additionally, the qualitative data also attempted to further probe into the reasons that caused test takers' anxiety in the CELST and seek strategies to alleviate the anxiety from the test takers' perspectives.

3.2 Participants

The questionnaires were distributed to a total of 585 CELST test takers who learned English as a foreign language. 510 out of 585 of the returned questionnaires were valid, which were used for the subsequent data analysis. Those participants were all high school Grade three students from an average level senior high school in Shenzhen city, Guangdong province, China. Due to the tight schedule of the students preparing for the College Entrance Examination, this study employed the convenience sampling methods in recruitment, which is a non-probability sampling that recruits participants based on proximity, willingness and accessibility to take part in the study (Larson & Csikszentmihalyi, 2014). According to the demographic information in the survey, among the 510 participants, 230 (45.1 %) were males and 280 (54.9 %) were females. In terms of major orientation, at Grade two in senior high schools, classes are divided into two types: history-oriented classes whose students choose history as their compulsory subject and physics-oriented classes whose students choose physics as their compulsory subject. Students from these two types of classes also learn Chinese, mathematics and English as compulsory subjects but they can randomly choose two of the following four subjects as additional compulsory subjects: chemistry, biology, geography and politics. In this study, 291 (57.1%) of the participants were from the history-oriented class, while 219 (42.9%) were from the physics-oriented class. As for the years of learning English, participants started learning English either in kindergarten or in primary school Grade one. The participants of the study were a representative sample of thes population of those who normally take the CELST.

Twelve of the participants who took the survey were invited to have a semi-structured interview with the researcher. The interviewees formed a balanced sampling in terms of gender (six males and six females) and anxiety level (four from each level: low, medium and high). The interviewees were divided into three anxiety levels based on their CTA scale scores, FLLA scale scores and FLSA scale scores with 25^{th} and 75^{th} percentiles used as the cut-off point (Hewitt & Stephenson, 2012; Phillips, 1992;). The test takers whose CTA scale, FLLA scale and FLSA scale scores fell into the top 25^{th} percentile, bottom 75^{th} to top 25^{th} percentile, and bottom 25^{th} percentile formed the high anxiety group, medium anxiety group and low anxiety group respectively. As presented in Table 3.1, pseudonyms were given to each of the twelve interviewees.

Table 3. 1
Interviewee Profile

	Anxiety Level					
Gender	High		Medium		Low	
Male	August	Rain	Tom	Zack	Ben	Sky
Female	Andy	Andrea	Jayda	Susan	Jenny	Lisa

3.3 Instruments

3.3.1 Questionnaires

Paper-based questionnaires (see Appendix B) were administered to the test takers to collect their personal information and measure their CTA, FLLA and FLSA. The questionnaires include four parts: demographic information, Cognitive Test Anxiety Scale (CTAS), Foreign Language Listening Anxiety Scale (FLLAS) and Foreign Language Speaking Anxiety Scale (FLSAS). To facilitate students' reading, the questionnaires were all in Chinese, the students' mother tongue.

Construction and Validation of the Questionnaires. To verify the construct validity and reliability of the questionnaires, the following method was applied during the construction and validation process of the questionnaires: literature review, qualitative input, peer evaluation, experts' judgment, and pilot study.

First of all, the content validity of the questionnaires was ensured by the previous relevant literature and interviews with EFL teachers and potential participants. As the literature in sections 2.3.3, 2.4.2 and 2.5.2 indicated, the CTAS (Cassady & Johnson, 2002), FLLAS and FLSAS (Cheng 2017) had been shown to have high reliability and validity and were the best suited scales to achieve the purposes of the study. Therefore, the study employed the CTAS, FLLAS, FLSAS for the measures of CTA, FLLA and FLSA. Previous studies also suggested that variables such as gender (Ça, 2015; Mohammadi Golchi, 2012; Yoongoo, 2017), self-perceived competence(Chiang, 2006; MacIntyre, 1992; Truitt, 1995), and frequency of language use (He, 2017) impacted students' anxiety levels in an EFL context. Therefore, gender, self-perceived English listening and speaking competence, frequency of language use related to English speaking and listening were included in the demographic information. Besides, the informal focus group interviews with two EFL teachers and six potential participants also suggested that students' anxiety levels might be varied between different major orientation students (history-oriented or physics-oriented). Thus, students' major orientation was included in the demographics as well.

Second, content validity was further confirmed by peer evaluation and experts' judgment. The researcher sent the questionnaires to four post-graduate students majoring in EFL teaching and two experts that specialized in foreign language anxiety research to collect their feedback on the item content. The expert suggested that the word "English test" as in CTAS items such as "I'm not good at English tests" was too broad and should be specified as "the CELST test". In addition, the expert also suggested that the situations that the FLLAS and FLSAS described should be narrowed down to the CELST. MacIntyre and Gardner (1991) also criticized that from the perspective of a situation, the researcher is free to define the situation, either very broadly or narrowly. Therefore, " it is the researcher's responsibility to define a situation that is sufficiently specific to be meaningful for the purpose at hand, yet to have reasonable generality to permit generalization" (MacIntyre & Gardner, 1991, p. 91). According to the experts' suggestions and the literature, the situation that the items in the CTAS, FLLAS and FLSAS described were defined more specifically to the context of the CELST.

Third, before piloting, the questionnaires were also reviewed by four potential participants to ensure the clarity and readability of the item wording. No problem was found regarding the item wording. Due to the demanding schedule of the senior three students, it was difficult to find a large group of students for a pilot study. Therefore, the questionnaires were only administered to 40 participants in the pilot study. The pilot data yielded Cronbach's alpha coefficients of .90 for the CTAS, .838 for the FLLAS and .854 for the FLSAS, indicating that the three scales had an excellent level of internal consistency and reliability. It should be noted here that the number of participants (N = 40) in the pilot study was too small to conduct exploratory factor analysis (EFA) to check the psychometric properties of the scales. Comrey and Lee (2013) suggested a range of minimum sample sizes for EFA to be from 50 (very poor) to 1,000 (excellent) and researchers are advised to acquire sample sizes larger than 500. Therefore, the EFA was conducted later in the main study with a sample size of 510 to ensure the construct validity of the questionnaires.

Content and Structure of the Questionnaires.

(1) Demographics

The demographic information includes students' ID, which was used to link to their correspondent CELST scores when the results of the CELST scores were announced later. Gender, perceived English speaking and listening competence, and major orientations were added as demographic items. The frequency of language use related to English speaking and listening includes five items: *How often did you speak English in English class in the past year? How often did you speak English outside of English class in the past year? How often did you do the mock CELST practice in the past year? How often did you watch English movies, TV shows and listen to English songs? How often did you communicate with native english speakers?* The items made use of a five-point Likert scale. Responses range from "rarely" to ""always".

(2) Cognitive Test Anxiety Scale (CTAS)

The CTAS was originally developed by Cassady and Johnson (2002) to measure students' cognitive test anxiety levels. This 27-item unidimensional scale of cognitive test anxiety has been proven to have high internal consistency (.91). To avoid redundancy, Zheng and Cheng (2018) trimmed the items that were termed very similar, resulting in a 20-item more concise Chinese version of the CTAS. The translated version of the CTAS by Zheng and Cheng (2018) was administered to 921 Chinese university students and was validated by them with an internal consistency of .87. I adapted their translated version of the CTAS to fit the specific context of CELST. The word "English test" in all of the CTAS items was replaced by the term "the CELST". Besides, three items that indicated anxious feelings in writing, reading and listening were deleted. The adapted version of CTAS contained 17 items with a five-point Likert scale, with responses ranging from "strongly disagree" to "strongly agree" (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, 5=strongly agree). Higher scores indicated a higher level of cognitive test anxiety. For positively worded items, "strongly disagree" received five points, and "strongly agree" received one point, whereas for negatively worded items, the inverse applied. The scores ranged from 17 to 85. The items were named CTAS1 to CTAS 17.

(3) Foreign Language Listening Anxiety Scale (FLLAS) and Foreign Language Speaking Anxiety Scale (FLSAS)

The FLLAS and FLSAS, developed by Cheng (2017), were to measure Chinese EFL students' FLLA and FLSA. As justified in section 2.4.2, given the length of the existing language-skill-specific anxiety scales, it is challenging to test the relationships among multiple anxiety constructs in one study due to limited time or item space. Therefore, the efficient FLLAS and FLSAS by Cheng (2017) with high reliability, convergent validity and discriminant validity would be the best scale to measure FLLA and FLSA in this study. The FLLAS and FLSAS are both composed of three dimensions: cognitive, somatic and behavioral. There are three items under each dimension. Each scale contained 9 items and was all originally written in Chinese. As with the CTAS, a five-point Likert-type scale was used as the response format. Higher scores indicated a higher level of FLLA and FLSA. The scores ranged from 9 to 45. The FLLAS items were named FLLAS 1 to FLLAS 9 while the FLSAS items were named FLSAS1 to FLSAS 9.

3.3.2 Interviews

The main aim of the semi-structured interview was twofold: first, to attempt to complement the findings from the questionnaires. Second, to further explore the reasons that lead to test takers' anxiety in the CELST and strategies helping to alleviate it. An interview outline (see Appendix C) containing a list of questions was developed based on the above-mentioned aims and the CTAS, FLLAS and FLSAS. The pre-established questions were used to inquire about test takers' feelings and attitudes towards the CELST, the effect of anxiety on the CELST performance, reasons that lead to anxiety in the CELST, the strategies helping to reduce anxiety and their suggestions for making the CELST less anxiety-provoking. The interviews were conducted in a semi-structured way which had a well-structured framework and meanwhile allowed flexibility during the interviews. For instance, the researcher could change the order of the questions and follow up unexpected answers if necessary to further explore test takers' ideas and elicit multiple perspectives.

In order to examine whether the interview outline was well designed or not, the researcher conducted pilot interviews with two senior three students first. From the pilot interviews, it was found that the original interview outline was well designed as all the questions were comprehensible to the students and could elicit enough responses from the interviewees. Therefore, the interview outline was expected to work well in the interviews in the main study.

3.4 Data Collection

3.4.1 Quantitative Data Collection

To get approval for conducting the surveys and interviews in the participating school, the researcher first approached the instructors and the chief director of the English courses in that school. After the researcher explained the research purposes and guaranteed the students' confidentiality, some instructors agreed to help me distribute the questionnaires to the class that they were teaching after the students took the CELST. However, two instructors rejected the distribution of the questionnaires as they were afraid that completing the questionnaires would occupy too much of their students' time. Nevertheless, the researcher also got the approval to access the file of the students' CELST scores two months later when the results of the CELST test scores were announced. The researcher then administered the questionnaires one day after students took the CELST. Before the distribution of the questionnaires, the researcher also explained the research purposes and the instructions of the questionnaires to the students and guaranteed the confidentiality of their personal information. The researcher then distributed the questionnaires with the help of the instructors. No time limit was imposed when students filled in the questionnaires. While answering the questionnaires, the students were encouraged to ask questions about anything in the questionnaires that was ambiguous to them. As this was an anonymous questionnaire, students were reminded to write down their student ID numbers so that the researcher could identify their scores using the ID numbers when obtaining their test scores from the participating school two months later. All together 585 questionnaires were acquired from the students. The researcher discarded the unusable questionnaires that had two or more items unanswered, resulting in 510 copies left for subsequent data analysis.

3.4.2 Qualitative Data Collection

From the questionnaires collected, the researcher identified twelve potential interviewees based on their gender (six males and six females) and anxiety level (four from each level: low, medium and high). The researcher then asked the instructors to help invite the ideal participants for the interviews. The instructors then asked for volunteer students based on the criteria the researcher requested and made a schedule for me to meet the participants individually over Tencent online meeting platform. The interviews with the twelve participants took place on the weekend after participants took the survey as this was the ideal time to interview due to their tight schedule. The researcher met with each of the twelve participants individually on the pre-scheduled time on Tencent online meeting platform. The whole interviews were conducted in Chinese, the mother tongue of the participants. During the interview, the researcher started with the lead-in section where he introduced himself and the research purpose to the participants, establishing a friendly atmosphere to make them feel comfortable sharing their thoughts. Then, the researcher asked for their verbal consent to participate in the study and their permission for audio-recording. Once participants felt well-informed, the researcher started to ask the questions in the interview outline. Each interview lasted approximately 25 minutes.

3.5 Data Analysis

Table 3. 2

Data Analysis for the Present Study

Purpose	**Analysis**
Validation of the questionnaire	Exploratory factor analysis; Reliability analysis
RQ1: Characteristics of CTA, FLLA and FLSA	Descriptive statistics; Independent samples t-test; One-way ANOVA; Pearson correlation analysis

RQ2: Effects of CTA, FLLA and FLSA on the CELST scores	One-way ANOVA; Pearson correlation analysis; Stepwise multiple regression analysis;
RQ3: Reasons leading to anxiety in the CELST and how to alleviate the anxiety	Thematic analysis

Data analysis includes quantitative data analysis and qualitative data analysis. A summary of the data analysis is shown in Table 3.2.

3.5.1 Quantitative Data Analysis

All the statistical analyses in this section were conducted using the Statistical Package for Social Sciences (SPSS) version 27. The questionnaire data were first entered into the SPSS for further statistical analysis. The responses in the negatively worded items were coded as "1=strongly disagree", "2=disagree", "3=neutral", "4=agree", "5=strongly agree" while the inverse applied for the positively worded item responses. To ensure accurate data coding, the data entry was completed in a two-person manner where one postgraduate student read out the numbers that represented the item responses and the researcher typed them into SPSS and read them back.

The statistical analyses were then conducted to achieve the following three purposes: (a) validate the questionnaire; (b) answer research question one: what are the characteristics of CTA, FLLA and FLSA among the CELST test takers; (c) answer research question two: what are the effects of CTA, FLLA and FLSA on the CELST test takers' performance.

Validation of the Questionnaire. As mentioned earlier, the questionnaire was validated by reliability analysis in a pilot study with 40 participants. However, as noted earlier, the sample size of the pilot study (N=40) was too small to conduct exploratory factor analysis (EFA). Therefore, EFA was conducted in the main study to further ensure the construct validity of the

CTAS, FLLAS and FLSAS and that all the three scales are reliable for further data analysis.

Prior to the extraction of the factors, the suitability of the sample data for conducting EFA was evaluated by the Kaiser-Meyer-Olkin (KMO) and Bartletts' Test of Sphericity. The KMO value of .60 was considered suitable for EFA (Kaiser, 1970) and the Bartletts' Test of Sphericity should be significant ($p<0.05$) for EFA to be suitable (Williams et al., 2010). Principal axis factoring was selected as the extraction method with orthogonal varimax chosen as the rotation method. For the multidimensional FLLAS and FLSAS, the eigenvalue > 1.0 was used as the criterion for extracting factors of the two scales (Fabrigar & Wegener, 2011). However, as the CTAS has been proven to be a unidimensional scale, using the criterion of eigenvalue greater than one is problematic for the unidimensional CTAS because SPSS uses Kaiser-Guttman criterion for deciding on the number of factors, which will result in too many factors. Therefore, to see whether the items load highly on the one factor of the CTAS, rather than using the criterion of eigenvalue > 1.0, the fixed number of factors extracted was employed and the extracted number of factors of the CTAS was restricted to one. The factor loading of .40 was chosen as the cut-off value for item loading and cross loading. Reliability analysis was performed on each scale to examine the internal consistency of the CTAS, FLLAS and FLSAS.

Data Analysis for Research Question One. To answer research question one concerning the characteristics of CTA, FLLA and FLSA among the CELST test takers, the following statistical analyses were conducted.

First, the descriptive analysis was conducted on participants' demographic information, which includes gender, major orientation, self-perceived English listening and speaking competence, frequency of language use.

Second, the descriptive statistics of the CTAS, FLLAS and FLSAS were analyzed at the item level to understand test takers' general condition of anxiety. Descriptive statistics that divide test takers into different levels of anxiety can give a more accurate understanding of the condition of anxiety. Therefore, the sample was divided into low, medium and high anxiety groups (Cassady & Johnson, 2002). The participants reporting the top 25 % of scores on the CTA formed the high CTA group, and those obtaining the lowest 25 % were considered the low CTA group. Participants in the top 25% to the bottom 25 % were classified as the medium CTA group (Wu, 2021). The same procedure was applied to divide participants into high, medium and low anxiety groups of FLLA and FLSA according to their FLLAS scores and FLSAS scores.

Third, an independent samples t-test was conducted to investigate whether significant differences exist in CTA, FLLA and its subdimensions, and FLSA and its subdimensions between different gender and different major orientation. T-tests are commonly used to examine whether the means of two variables are significantly different from each other. The common assumptions made when conducting a T test include normality of distribution, random sampling and equality of variances.

Fourth, one-way ANOVA was adopted to examine whether significant differences exist in CTA, FLLA and FLSA among participants from three different self-perceived competence groups (low, medium and high). Post hoc comparisons using the Tukey HSD test was also conducted to see which group significantly differed from others in regard to means. Conducting the one-way ANOVA also needs to meet the assumptions of homogeneity of variance and normally distributed data.

Fifth, Pearson product-moment correlation coefficient (r) was computed to investigate the correlation between the frequency of language use and CTA, FLLA and its subdimensions, and FLSA and its subdimensions. The coefficient values can range from -1 to 1, where +1 indicates a

100% positive correlation, -1 indicates 100% negative correlation and a 0 indicates no existing correlation. According to Cohen (1992), the strength of correlation coefficients can be interpreted as follows: small (.10), moderate (.30), and large (.50).

Data Analysis for Research Question Two. To answer research question two regarding the effects of CTA, FLLA and FLSA on the CELST test takers' performance, the following statistical analyses were performed.

First, one-way ANOVA was conducted to investigate whether participants' CELST scores differed from high, medium, low CTA groups. The one-way ANOVA also examined whether significant differences exist in the CELST scores among different levels of FLLA groups and FLSA groups.

Second, Pearson product-moment correlation coefficient was conducted to examine the CELST scores and its relationships with CTA, FLLA and its subdimensions, and FLSA and its subdimensions. In addition, the correlation between the three anxiety variables CTA, FLLA and FLSA was also examined using the Pearson correlation coefficient. The Pearson correlation analysis also served as the preliminary analysis for the subsequent multiple regression analysis. If CTA, FLLA and its subdimensions, FLSA and its subdimensions have significant correlations with the CELST scores, these variables would be entered into stepwise multiple regression for further analysis.

Third, Stepwise multiple regression was conducted to investigate which anxiety variable that had significant correlations with the CELST scores could best predict test takers' CELST performance. Stepwise regression is a method that iteratively investigates and selects the significant independent variables that can be used in the final regression model. The independent variables in the final model are considered the best predictors of the dependent variable, which is the CELST scores in this study.

3.5.2 Qualitative Data Analysis

The interviews were recorded by the Tencent online meeting platform and automatically transcribed verbatim into Chinese. The collected interview data were analyzed through thematic analysis (Phakiti, 2003), a method that examines the data to explore common themes, such as ideas or patterns of meaning that brought up repeatedly by interviewees. Through thematic analysis, no assumption was imposed during the data analysis process, and thus it helps to identify some unanticipated findings (Braun & Clarke, 2006). Following the guidelines of Braun and Clarke (2006). I took the following steps to analyze the data.

First, I checked the transcripts generated by Tencent online meeting platform application back against the audio recording repeatedly to ensure the accuracy of the transcripts. Second, I familiarized myself with the data by reading and re-reading the transcripts to gain a general understanding of the data and highlighting potentially relevant or interesting analytic materials. Third, I generated the initial codes by identifying interesting data items that might form the basis of themes. Fourth, I used Microsoft Word 2016 to code the data by tagging and naming data extracts. Fifth, I searched for themes by sorting the different codes into potential themes. Sixth, I reviewed the previously identified themes to see if the themes captured the contours of coded data. Seventh, I defined and named the themes by identifying what each theme was essentially about.

Chapter Four Results and Discussion

This chapter presents the results of both quantitative and qualitative analysis. The first section shows the results of the validation of the CTAS, FLLAS and FLSAS. The second and third sections present the descriptive and inferential results of the quantitative data analysis related to the first and second research questions. The fourth section demonstrates the results of the thematic analysis of the interview data to answer research question three.

4.1 Results of the Validation of the Questionnaire

The results of exploratory factor analyses and the reliability analyses conducted for each of the three scales (the CTAS, FLLAS and FLSAS) were presented below.

4.1.1 Exploratory Factor Analysis and Reliability Analysis of the CTAS

The Kaiser-Meyer-Olkin measure (KMO) of sampling adequacy value is .92, and the Bartlett's Test of Sphericity was statistically significant ($P<.001$), indicating that the data set was appropriate for EFA. EFA with principal axis factor analysis and varimax rotation was performed. According to the EFA results, the factor loadings of all the CTAS items (ranging from .510 to .704) loaded highly on one factor, with all the loadings above the .50 threshold. The one-factor structure retained all the 17 items on the CTAS and accounted for 44.22% of the variance, indicating the construct validity of CTAS was acceptable. Besides, the reliability analysis of the one-factor structure of the CTAS indicated an excellent level of internal consistency (Cronbach's alpha =0.92). As both EFA and reliability analysis supported the unidimensional CTAS as a valid scale, the CTAS with 17 items was qualified for subsequent data analysis in the current study.

Table 4. 1
Factor Loadings of Cognitive Test Anxiety Scale

	Factor
	1
CTAS1: I had less difficulty than average students in getting the English test instruction straight	.510
CTAS2: I was less nervous about the CELST than average students	.686
CTAS3: During the CELST, I found myself thinking of the consequences of failing.	.612
CTAS4: At the beginning of the CELST, I was so nervous that I often couldn't think straight.	.634
CTAS5: I was calmer in the CELST test situations than average students.	.700
CTAS6: My mind went blank when I was pressured for an answer on the CELST test.)	.626
CTAS7: During the CELST, the thought frequently occurred to me that I may not be too bright	.627
CTAS8: During the CELST, I got so nervous that I forgot facts that I really knew.	.704
CTAS9: I worried about doing well on the CELST than I should.	.685
CTAS10: Before taking the CELST, I felt confident and relaxed.	.696
CTAS11: While taking the CELST, I felt confident and relaxed.	.684
CTAS12: During the CELST, I had the feeling that I was not doing well.	.553
CTAS13: I worried that the CELST was too difficult and I felt defeated before I even started.	.693
CTAS14: I am not good at taking the CELST.	.641
CTAS15: When I first got my copy of the CELST, it took me a while to calm down to the point where I could begin to think straight.	.597
CTAS16: I felt under a lot of pressure to get good grades on the CELST.	.616
CTAS17: I did not perform well on the CELST.	.579

4.1.2 Exploratory Factor Analysis and Reliability Analysis of the FLLAS and FLSAS

The Kaiser-Meyer-Olkin measure (KMO) of sampling adequacy value for the FLLAS and FLSAS was .809 and .848 respectively. The Bartlett's Test of Sphericity for the FLLAS and FLSAS was statistically significant at the 0.001 level. This indicated that the dataset met the criteria for performing EFA. Thus, EFA with principal axis factor and varimax rotation was performed.

As seen in Table 4.2, the EFA results of the FLLAS suggested a three-factor structure of the scale. Although FLLA4 cross-loaded onto factor one and factor two, it had greater loading on factor two (.602) than factor one (.451). Therefore, FLLA4 was classified as the item of factor two. Factor one of the FLLAS comprises three items that measure the cognitive dimension of FLLA, and thus it was named foreign language (FL) listening cognitive anxiety, which refers to worry over bad performance and others' negative evaluation in listening to English (Cheng, 2017). Factor two contains three items that describe the somatic dimension of FLLA, and thus it was titled FL listening somatic anxiety, which refers to the physiological arousal stimulated by listening to English. Factor three is composed of three items that assess the behavioral dimension of FLLA, and thus it was termed FL listening behavioral anxiety, which refers to procrastination and avoidance behavior when listening to English is needed. Factor one, factor two and factor three accounted for 47.32%, 19.50 %, 12.44% of the variance respectively and altogether the three factors accounted for 79.26% of the total variance, suggesting that the FLLAS had fairly high construct validity. The result of the reliability analysis of the FLLAS also indicated excellent internal consistency (Cronbach's alpha =.86). Therefore, the results of both EFA and reliability analysis of the FLLAS suggested that the FLLAS could be a valid and reliable measure of FLLA.

Table 4. 2

Rotated Factor Matrix of Foreign Language Listening Anxiety Scale

	Factor		
	1	2	3
Factor 1 FL listening cognitive anxiety			
FLLAS1: When listening to English during the CELST, I often worried that I would miss information	.865		
FLLAS2: When listening to English during the CELST, I often worried that I would misunderstand something.	.907		
FLLAS3: When listening to English during the CELST, I often worried that my listening comprehension performance was worse than others.	.688		
Factor 2 FL listening somatic anxiety			
FLLAS4: When listening to English during the CELST, I often felt my heart pounding.	.451	.602	
FLLAS5: When listening to English during the CELST, I often got so nervous that I trembled.		.890	
FLLAS6: When listening to English during the CELST, I often sweated and perspired.		.839	
Factor 3 FL listening behavioral anxiety			
FLLAS7: When listening to English, I often give it up easily.			.689
FLLAS8: I often skimp over English listening comprehension exercises.			.817
FLLAS9: I usually do my best to avoid listening to English.			.720

As shown in Table 4.3, the EFA results of the FLSAS also indicated a three-factor structure of the scale, which aligned with the factor structure of the original FLSAS. As with the FLLAS, the FLSAS items loaded on factor one, factor two and factor three represented the cognitive, somatic and behavioral dimensions of the FLSAS, and thus they were termed FL speaking cognitive anxiety, FL speaking somatic anxiety and FL speaking behavioral anxiety respectively. FL speaking cognitive anxiety describes students' worry over poor performance and negative feedback from others when speaking English. FL speaking somatic anxiety means physiological arousal triggered by speaking English. FL speaking behavioral anxiety represents the withdrawal and avoidance behavior when speaking English is necessary. Factor one, factor two and factor three accounted for 55.43%, 15.24 %, 13.00 % of the variance and taken together they accounted for 83.67 % of the total variance, indicating that the construct validity of the FLSAS was relatively high. The reliability analysis of the FLSAS also indicated excellent internal consistency (Cronbach's alpha =.89). Therefore, the results of EFA and reliability analysis both indicated that the FLSAS were reliable for further statistical analysis.

Table 4. 3

Rotated Factor Matrix of Foreign Language Speaking Anxiety Scale

	Factor		
	1	2	3
Factor 1 FL speaking cognitive anxiety			
FLSAS1: When speaking English during the CELST, I often worried that I would make language mistakes.		.807	
FLSAS2: When speaking English during the CELST, I often worried that my speaking performance was worse than others.		.879	
FLSAS3: As soon as I started speaking English during the CELST, I begun to worry about not being able to express myself.		.731	
Factor 2 FL speaking somatic anxiety			
FLSAS4: When speaking in English during the CELST, I often felt my heart pounding.	.752		
FLSAS5: When speaking in English during the CELST, I often got so nervous that I trembled.	.925		
FLSAS6: When speaking in English during the CELST, I often sweated and perspired.	.868		
Factor 3 FL speaking behavioral anxiety			
FLSAS7: When practicing speaking English, I often give it up easily.			.795
FLSAS8: I often skimp over English-speaking exercises.			.770
FLSAS9: I usually do my best to avoid speaking English.			.713

In conclusion, the data above demonstrated that the CTAS, FLLAS and FLSAS could be reliable and valid measures of CTA, FLLA and FLSA respectively, and thus the three scales would be used for the subsequent data analysis.

4.2 Quantitative Results of Research Question One: Characteristics of CTA, FLLA and FLSA

The purpose of research question one was to examine the characteristics of CTA, FLLA and FLSA among the CELST test takers. The characteristics of CTA, FLLA and FLSA among the CELST test takers consist of: (1) the status quo of CTA, FLLA and FLSA among the CELST test takers; (2) the relationship between CTA, FLLA and FLSA and individual differences, that is, gender, major orientation, self-perceived English listening and speaking competence and frequency of language use related to English listening and speaking. **First**, the demographic information is presented to have an overall understanding of the participants. **Second**, the results of descriptive analyses of the CTAS, FLLAS and FLSAS at the item level were offered to illustrate the status quo of CTA, FLLA and FLSA among test takers. **Third**, the results of the independent samples t-test, one-way ANOVA are provided to examine whether CTA, FLLA and FLSA varied by test takers' gender, major orientation and self-perceived English competence. Besides, results of Pearson correlation is presented to examine the relationship between frequency of language use and CTA, FLLA and FLSA.

4.2.1 Demographic Information

As mentioned earlier, among the 510 participants, 230 (45.1 %) were males and 280 (54.9 %) were females. As for major orientation, 291 (57.1%) of the participants were from the history-oriented class, while 219 (42.9%) were from the physics-oriented class. The CELST scores of participants range from 5 to 20 with a mean score of 15.89 (SD=2.10). In terms of self-perceived English listening and speaking competence, the mean score is 1.67 (SD=0.61). Participants' actual English listening and speaking competence measured by CELST was above average with more than 65.8 % of participants scoring above 16 out of 20. However, only 6% (N=31) of the participants rated their listening and speaking competence as relatively high. 51.8% (N=264) of the participants rated their competence as medium and 42.2 % (N=215) of the participants rated their competence as relatively low. This indicated that the vast majority of participants underrated their actual competence and did not have much confidence in their English listening and speaking competence. There are five demographic items that describe the frequency of language use related to English listening and speaking. First, as for speaking English outside of and inside of the class, the mean score is 2.44 (SD=0.99) and 2.73 (SD=1.10) respectively, indicating most of the participants speak English "occasionally" or "sometimes" inside and outside of class. As for the frequency of practicing the mock CELST test, the mean score is 3.22 (SD=1.15), indicating most of the participants practiced the mock CELST sometimes. In terms of the frequency of communicating with native English speakers, the mean score is 1.82 (SD=0.88), indicating most of the participants talked with native English speakers occasionally. As for the frequency of them watching English movies, TV shows and listening to English songs, the mean score is 3.54 (SD=1.10), with 35.9 % and 20.8% of participants reporting they watch English TV shows or listening to English songs "often" and "always" respectively. Taken together, participants' frequency of language use related to English listening and speaking was at a moderate level.

4.2.2 Descriptive Statistics of the CTAS, FLLAS and FLSAS

To understand participants' status quo of CTA, FLLA and FLSA, descriptive statistics of the CTAS, FLLAS, FLSAS at the item level are presented. It should be noted that more detailed descriptive statistics containing participants at different anxiety levels would give a more in-depth understanding of the characteristics of CTA, FLLA and FLSA. Therefore, participants were divided into three levels of anxiety group (high, medium, and low group) based on their CTA scores, FLLA scores and FLSA scores with the 25th and 75th percentiles being employed as the cutoff point (Hewitt & Stephenson, 2012; Phillips, 1992; Wu, 2021). As presented in Table 4.4, the score range for the low, medium, and high CTA groups is 17-36, 36-53 and 53-85 respectively. For FLLA, the score range of low, medium, and high FLLA groups is 9-18, 18-26, and 26-41 respectively. As far as FLSA, the score range of low, medium, and high FLSA groups is 9-17, 17-26, and 26-45 respectively.

Table 4. 4
Participants Grouping According to Their CTA, FLLA and FLSA Scores.

		CTA	FLLA	FLSA
Low anxiety level (bottom 25%)	Score range	17-36	9-18	9-17
	M/SD	28.02/6.44	13.93/3.74	12.07/3.06
	N	126	126	106
Medium anxiety level (bottom 25% to top 25%)	Score range	36-53	18-26	17-26
	M/SD	45.38/4.70	22.87/2.27	21.61/2.68
	N	260	246	271
High anxiety level (top 25%)	Score range	53-85	26-41	26-45
	M/SD	59.77/5.51	29.70/2.80	30.61/3.65
	N	124	138	133

The CTAS, FLLAS, and FLSAS all use a five-point Likert scale format with “1= strongly disagree”, “2=disagree”, “3=neutral”, “4=agree” and “5=strongly agree”. Therefore, for the reporting of descriptive statistics in this study, mean scores within 1-2.4 indicate a low degree of agreement with the item statement and signal a low anxiety level; mean scores falling into the range of 2.5-3.4 connote a moderate degree of agreement and a medium anxiety level; mean scores ranging from 3.5 to 5 suggest a high degree of agreement and a high level of anxiety

(Oxford & Burry-Stock, 1995).

In terms of the CTAS descriptive statistics shown in Table 4.5, for all participants, the average mean scores of 17 CTAS items reached a moderate degree of agreement. When it comes to participants from the high CTA group (N=126), they showed a moderate to high degree of agreement with all the 17 items, with the items' mean scores ranging from 3 to 3.96. This indicated that participants with a high level of cognitive test anxiety experienced a considerable amount of stress and suffered from task-irrelevant thoughts in CELST. Besides, for participants from the medium CTA group (N=260), except item CTAS1(M=2.03, SD=.94), CTAS4 (M=2.02, SD=.686) and CTAS8 (M=2.34, SD=.79), the rest of CTAS items all show a moderate degree of agreement with the mean score ranging from 2.53 to 3.08, suggesting that participants with a medium level of CTA were, to some extent, test anxious and not very confident in the CELST. However, as for participants from the low CTA group (N=124), they had a low degree of agreement with the statements of all CTAS items, with the mean score ranging from 1.29 to 2.13, indicating CTA did not seem to interfere with this group of participants.

In regard to the FLLAS descriptive statistics shown in Table 4.6, for all participants, the mean score of FL listening cognitive anxiety indicated a moderate degree while the mean scores of FL listening somatic anxiety and behavioral anxiety suggested a low degree of agreement. For those from the high FLLA group (N=138), they had a high degree of agreement with items representing FL listening cognitive anxiety [FLLA1 (M=.4.04, SD=.66), FLLA2 (M=4.05, SD=.65) and FLLA3 (M=3.91, SD=.75)]. This suggests that participants with a high level of FLLA had often experienced negative expectations, preoccupation with performance and concerns about others' perceptions when they listened to English during the CELST. Those from the medium FLLA group (N=246) and low FLLA group (N=126) showed their agreement with

FL listening cognitive anxiety items at moderate and low levels respectively. FL listening cognitive anxiety also received the highest mean score among the three FLLA subdimensions for three anxiety groups. For items representing FL listening somatic anxiety and FL listening behavioral anxiety, those from the high FLLA group showed a moderate level of agreement, whereas those from the medium and low FLLA group both showed a low degree of agreement, indicating participants suffered less from FL listening somatic anxiety and behavioral anxiety than FL listening cognitive anxiety.

As is seen in Table 4.7, as for the FLSAS descriptive statistics, for all participants, the mean score of FL speaking cognitive anxiety indicated a moderate degree while the mean scores of FL speaking somatic anxiety and behavioral anxiety suggested a low degree of agreement. Items representing FL speaking cognitive anxiety reached a high, moderate, and low degree of agreement for participants from high (N=133), medium (N=271), and low (N=106) FLSA groups respectively, indicating that high and medium FLSA groups seemed to be interfered by FL speaking cognitive anxiety. For items representing FL speaking somatic anxiety and FL speaking behavioral anxiety, participants from the high FLSA group showed a moderate level of agreement while those from the medium FLSA group showed a low degree of agreement except item FLSAS4 “when speaking to English during the CLEST, I often feel my heart pounding” reached a moderate degree of agreement (M=2.78, SD=.672). The low FLSA group showed a low degree of agreement with all the items. This indicated that the high FLSA group experienced FL speaking somatic anxiety and behavioral anxiety while the medium and high FLSA group didn’t experience much of these two types of anxiety.

Table 4. 5
Descriptive Statistics of CTAS Items for Three CTA Groups

Item	CTA level			
	Low (N=126) M/SD	Medium (N=260) M/SD	High (N=124) M/SD	Total (N=510) M/SD
CTAS1	1.40/.622	2.03/.942	2.75/1.24	2.05/1.068
CTAS2	1.57/.686	2.56/.888	3.60/.891	2.57/1.103
CTAS3	1.48/.865	2.46/.897	3.53/.991	2.48/1.197
CTAS4	1.29/.646	2.02/.686	3.00/.902	2.08/.948
CTAS5	1.77/.792	2.88/.744	3.52/.715	2.76/.974
CTAS6	1.74/.931	2.82/.924	3.68/.851	2.76/1.135
CTAS7	1.48/.837	2.55/.983	3.58/.964	2.54/1.195
CTAS8	1.36/.572	2.34/.787	3.35/.894	2.34/1.039
CTAS9	1.44/.721	2.55/.914	3.56/.965	2.52/1.154
CTAS10	1.81/.827	2.98/.909	3.72/.669	2.87/1.075
CTAS11	1.88/.806	3.01/.892	3.70/.754	2.90/1.059
CTAS12	2.13/1.145	3.10/.960	3.74/.742	3.01/1.119
CTAS13	1.47/.701	2.57/.847	3.44/.757	2.51/1.050
CTAS14	1.79/1.040	2.96/1.014	3.96/.896	2.91/1.251
CTAS15	1.56/.863	2.53/.893	3.30/.816	2.48/1.059
CTAS16	1.96/1.031	3.08/.895	3.81/.932	2.98/1.145
CTAS17	1.88/1.078	2.93/.938	3.52/.821	2.82/1.114

Table 4. 6
Descriptive Statistics of FLLAS Items for Three FLLA Groups

Dimension	Item	FLLA level			
		Low (N=126) M/SD	Medium (N=246) M/SD	High (N=138) M/SD	Total (N=510) M/SD
Cognitive	FLLAS1	2.02/1.016	3.66/.800	4.04/.660	3.36/1.137
	FLLAS2	1.93/.905	3.46/.864	4.05/.654	3.24/1.141
	FLLAS3	1.63/.724	3.13/1.000	3.91/.753	2.97/1.208

Table 4. 6

Descriptive Statistics of FLLAS Items for Three FLLA Groups (Continued)

		FLLA level			
Dimension	Item	Low (N=126) M/SD	Medium (N=246) M/SD	High (N=138) M/SD	Total (N=510) M/SD
Somatic	FLLAS4	1.60/.739	2.57/.895	3.43/.801	2.56/1.063
	FLLAS5	1.33/.490	2.00/.538	3.07/.909	2.12/.910
	FLLAS6	1.31/.481	1.99/.619	3.05/.923	2.11/.936
Behavioral	FLLAS7	1.29/.504	1.87/.670	2.62/1.055	1.93/.901
	FLLAS8	1.50/.827	2.26/.963	2.84/.961	2.23/1.047
	FLLAS9	1.33/.655	1.93/.771	2.67/.998	1.98/.947

Table 4. 7

Descriptive Statistics of FLSAS Items for Three FLSA Groups

		FLSA level			
Dimension	Item	Low (N=106) M/SD	Medium (N=271) M/SD	High (N=133) M/SD	Total (N=510) M/SD
Cognitive	FLSAS1	1.95/1.133	3.21/.951	3.92/.794	3.13/1.168
	FLSAS2	1.59/.848	2.98/.977	3.99/.774	2.95/1.215
	FLSAS3	1.47/.693	2.97/.929	3.96/.732	2.92/1.191
Somatic	FLSAS4	1.25/.531	2.78/.672	3.33/.911	2.45/1.012
	FLSAS5	1.12/.330	2.00/.498	3.10/.912	2.10/.913
	FLSAS6	1.13/.367	1.97/.474	3.04/.924	2.07/.896
Behavioral	FLSAS7	1.06/.232	2.01/.626	3.05/.924	2.09/.953
	FLSAS8	1.34/.893	2.13/.802	2.91/.917	2.17/1.006
	FLSAS9	1.15/.644	2.12/.761	3.30/1.073	2.23/1.113

4.2.3 Results of the Relationships between Anxiety and Individual Differences

Gender and CTA, FLLA and FLSA. An independent samples t-test was conducted to compare the effect of gender on CTA, FLLA and FLSA. As shown in Table 6.1 in Appendix A, the males' CTA mean score (M=43.25, SD=12.24) is lower than that of females (M=45.69, SD=12.41). The independent samples t-test also suggested that there was a significant difference between males and females in CTA ($p< 0.05$). While the males' mean scores of FLLA and FLSA were lower than those of females, there were no significant differences between males and females in FLLA and FLSA ($p>0.05$). However, significant gender differences were found in the subdimensions of FLLA and FLSA. The T test suggested that males' mean scores of FL listening cognitive anxiety (M=9.02, SD=3.26) and FL speaking cognitive anxiety (M=8.57, SD=3.28) were significantly lower than those of females ($M_{\text{FL listening cognitive anxiety}}$=10.02, SD=3.03, $p<0.01$; $M_{\text{FL speaking cognitive anxiety}}$=9.37, SD=3.21, $p<0.01$). In addition, the males' mean scores of FL listening somatic anxiety and FL speaking somatic anxiety were lower than those of females, whereas their FL listening behavioral anxiety and FL speaking behavioral anxiety were higher than those of females. However, it should be noted that the gender differences in FL listening somatic anxiety, FL listening behavioral anxiety, FL speaking somatic anxiety and FL speaking behavioral anxiety were not significant ($p>0.05$). What is more, when conducting an independent-samples t-test on the CESLT scores, there were no significant differences ($p>0.05$) between males (M=15.74, SD=2.13) and females (M=16.01, SD=2.07) in terms of their CELST scores (see Table 6.2 in Appendix A).

Major Orientation and CTA, FLLA and FLSA. As shown in Table 6.3 in Appendix A, the results revealed that the history-oriented participants' mean scores of CTA, FLLA and its dimensions, FLSA and its dimensions are all slightly higher than those of the physics-oriented participants. However, the independent samples t-test revealed no significant differences between history-oriented participants and physics-oriented participants in terms of their CTA, FLLA and its dimensions, FLSA and its dimensions ($p>0.05$)

Self-perceived Competence and CTA, FLLA and FLSA. This study also explored the effect of self-perceived English listening and speaking competence on CTA, FLLA and FLSA. A one-way ANOVA was conducted to compare the effect of self-perceived English listening and speaking competence on participants' CTA, FLLA and FLSA. The results of tests of homogeneity of variances indicated the assumption of equal variances was not violated and was appropriate to conduct one-way ANOVA. As reflected in Table 4.8, there was a significant effect of self-perceived English listening and speaking competence on participants' CTA, FLLA and FLSA at the $p<.001$ level.

The scores of CTA significantly differed across participants' self-perceived English listening and speaking competence ($F=70.58$, $p<.001$). Post hoc comparisons using the Tukey HSD test indicated that the CTA mean score of the high self-perceived competence group ($M=28.42$, $SD=12.00$) was significantly lower than those of the medium ($M=41.86$, $SD=10.85$) and low ($M=50.26$, $SD=10.96$) self-perceived competence groups. The CTA mean score of the medium self-perceived competence group was significantly lower than that of the low self-perceived competence group.

The scores of FLLA were also significantly varied across participants' self-perceived English listening and speaking competence ($F=54.24$, $p<.001$). The results of Post hoc test indicated that the FLLA mean score of the high self-perceived competence group ($M=14.35$, $SD=6.79$) was significantly lower than that of the medium ($M=21.48$, $SD=5.83$) and low ($M=24.94$, $SD=5.54$) self-perceived competence groups. The FLLA score of the medium self-perceived group also significantly differed from that of low self-perceived competence group.

Besides, the mean score of FLSA also had a significant difference across participants' self-perceived English listening and speaking competence ($F=48.98$, $p<.001$). The results of Post hoc test suggested that the FLSA mean score of high self-perceived competence group ($M= 14.13$, $SD=6.24$) was significantly lower than the FLSA mean scores of those who perceived their English listening and speaking competence as medium (M=20.63, SD= 6.31) and low (M=24.76, SD=6.60). Additionally, the FLSA mean score of the medium self-perceived competence group significantly varied from that of the low self-perceived competence group.

Taken together, the results suggest that self-perceived English listening and speaking competence does have an effect on CELST test takers' CTA, FLLA and FLSA.

Table 4. 8

Results of One-way ANOVA of Three Self-perceived Competence Group

	High (N=31)	Medium (N=264)	Low (N=215)	F	Post hoc (Tukey HSD)
CTA	M/SD 28.42/12.00	M/SD 41.86/10.85	M/SD 50.26/10.96	70.58*	High<Medium<Low
FLLA	M/SD 14.35/6.79	M/SD 21.48/5.83	M/SD 24.94/5.54	54.24*	High<Medium<Low
FLSA	M/SD 14.13/6.24	M/SD 20.63/6.31	M/SD 24.76/6.60	48.98*	High<Medium<Low

*P<.001

(Note: High, medium, and low refer to high, medium, and low self-perceived English listening and speaking competence group respectively)

Frequency of Language Use and CTA, FLLA and FLSA. To investigate the relationship between the frequency of language use and CTA, FLLA and its subdimension, FLSA and its subdimensions, a Pearson product-moment correlation coefficient was computed. The results of correlation indicated that there was a moderate negative correlation between the frequency of language use and CTA ($r = -.345$, $p < .001$), FLLA ($r = -.329$, $p < .001$) and FLSA ($r = -.353$, $p < .001$). The results also revealed that there was a negative correlation between the frequency of language use and FL listening cognitive anxiety ($r = -.185$), FL listening somatic anxiety ($r = -.211$), FL listening behavioral anxiety ($r = -.390$), FL speaking cognitive anxiety (r

= -.236), FL speaking somatic anxiety (r = -.239), and FL speaking behavioral anxiety (r = -.397) at p<.001 level.

Table 4. 9

Correlation Analyses of Frequency of Language Use and CTA, FLLA and FLSA

Variables	1	2	3	4	5	6	7	8	9	10
1. Frequency of language use	--									
2. CTA	-.345**	--								
3. FLLA	-.329**	.755**	--							
4. FLSA	-.353**	.715**	.809**	--						
5. FL listening cognitive anxiety	-.185**	.673**	.843**	.659**	--					
6. FL listening somatic anxiety	-.211**	.610**	.808**	.626**	.574**	--				
7. FL listening behavioral anxiety	-.390**	.438**	.644**	.581**	.279**	.289**	--			
8. FL speaking cognitive anxiety	-.236**	.634**	.683**	.842**	.732**	.470**	.325**	--		
9. FL speaking somatic anxiety	-.239**	.569**	.675**	.822**	.486**	.719**	.357**	.549**	--	
10. FL speaking behavioral anxiety	-.397**	.533**	.614**	.774**	.351**	.350**	.767**	.441**	.488**	--

Note: **. Correlation is significant at the 0.01 level (2-tailed).

4.3 Quantitative Results of Research Question two: Effect of CTA, FLLA and FLSA

The purpose of research question two was to examine the effects of CTA, FLLA and FLSA on test takers' CELST performance, as reflected by their CELST scores. **First**, a one-way ANOVA was conducted to investigate whether the CELST scores were significantly varied across different anxiety level groups (low, medium, and high) of CTA, FLLA and FLSA. **Second**, the Pearson product-moment correlation coefficient was computed to examine the relationship between the CELST scores and CTA, FLLA and its subdimensions, FLSA and its subdimensions. Besides, the correlation analysis also examined the relationship between CTA, FLLA and FLSA.

Third, any of the anxiety variables that had a significant correlation with the CELST scores were analyzed in the stepwise multiple regression to further examine whether they were significant predictors of the CELST scores.

4.3.1 Results of One-way ANOVA

The Levene's test of homogeneity of variances for the CELST scores was not significant ($P>0.05$), indicating that the equal variances assumed were not violated and it was suitable to perform ANOVA. According to the ANOVA results, there was a significant difference among the three CTA anxiety level groups on the CELST scores, $F = 60.72$, $P<.001$. Post hoc testing using Tukey HSD also revealed significant differences among different CTA anxiety level groups, with the low CTA group having a higher CELST mean score ($M=17.14$, $SD=1.43$) than the CELST mean score by the medium CTA group ($M=15.94$, $SD=1.99$) and the CELST mean score by the high CTA group ($M=14.52$, $SD=2.07$).

There was also a significant difference among the three FLLA groups on the CELST scores, $F = 36.42$, $p<.001$. Post hoc testing revealed that the low FLLA group received the highest CELST mean score ($M=16.83$, $SD=1.94$), followed by the CELST mean score of the medium FLLA group ($M=16.02$, $SD=1.91$) and that of the high FLLA group ($M=14.80$, $SD=2.18$).

In terms of FLSA, participants' CELST scores also significantly differed across three FLSA groups, $F = 43.90$, $p<.001$. The results of the Post hoc test demonstrated that the highest CELST mean score was obtained by the low FLSA group ($M= 17.10$, $SD=1.77$), followed by the CELST mean score earned by the medium FLSA group ($M=15.97$, $SD=1.93$) and that of the high FLSA group ($M=14.75$, $SD=2.08$).

4.3.2 Results of Correlation Analysis

A Pearson product-moment correlation was conducted to examine the relationships among the CELST scores and CTA, FLLA and its subdimensions, FLSA and its subdimensions.

As seen in Table 4.10, the CELST score formed a significant negative correlation with CTA, FLSA and FLLA ($p<0.01$). The CELST scores were more strongly negatively related to CTA ($r=-.483$) than to FLSA ($r=-.413$) and FLLA($r=-.393$). The CELST scores also formed significant negative relationships with the three FLLA subdimensions and the three FLSA subdimensions ($p<0.01$). Among the significant correlations between six subdimensions and the CESLT scores, the strongest correlation occurred between FL speaking behavioral anxiety and the CELST scores ($r=-.419$). FL speaking cognitive anxiety and FL listening behavioral anxiety formed moderate correlations with the CELST scores ($r=-.335$, $r=-.378$ respectively). FL speaking somatic anxiety, FL listening cognitive anxiety and FL listening somatic anxiety formed small correlations with the CELST score ($r=-.254$, $r=-.286$, $r=-.255$ respectively).

The present study also attempted to investigate the relationship between CTA, FLLA and FLSA. As demonstrated in Table 4.10, CTA formed significant high correlations with FLLA ($r=.755$) and FLSA ($r=.715$) at the 0.01 level. FLLA and FLSA also highly correlated with each other ($r=.809$). In addition, as shown in Table 4.10, CTA, FLLA and its subdimensions, FLSA and its subdimensions also formed significant correlations with each other.

Table 4. 10

Results of Correlation Analysis of the CELST Scores and CTA, FLLA and FLSA

Variables	1	2	3	4	5	6	7	8	9	10
1. CELST scores	--									
2. CTA	-.483**	--								
3. FLLA	-.393**	.755**	--							
4. FLSA	-.413**	.715**	.809**	--						
5. FL listening cognitive anxiety	-.286**	.673**	.843**	.659**	--					
6. FL listening somatic anxiety	-.255**	.610**	.808**	.626**	.574**	--				
7. FL listening behavioral anxiety	-.378**	.438**	.644**	.581**	.279**	.289**	--			
8. FL speaking cognitive anxiety	-.235**	.634**	.683**	.842**	.732**	.470**	.325**	--		
9. FL speaking somatic anxiety	-.254**	.569**	.675**	.822**	.486**	.719**	.357**	.549**	--	
10. FL speaking behavioral anxiety	-.419**	.533**	.614**	.774**	.351**	.350**	.767**	.441**	.488**	--

Note: **. Correlation is significant at the 0.01 level (2-tailed).

4.3.3 Results of Multiple Regression Analysis

The Pearson correlation analysis revealed that CTA, FLLA subdimensions and FLSA subdimensions significantly and negatively correlated with the CELST scores. Therefore, a stepwise multiple regression was conducted to identify the possible best predictors of the CELST scores out of the following candidate variables: CTA, FL listening cognitive anxiety, FL listening somatic anxiety, FL listening behavioral anxiety, FL speaking cognitive anxiety, FL speaking somatic anxiety and FL speaking behavioral anxiety.

As presented in Table 4.11, three models were constructed through stepwise multiple regression. The adjusted R Square (R^2) in Table 4.11 indicates the percentage of variation explained by only the independent variables that actually affect the dependent variable. The higher the adjusted R^2, the better the model fit. As seen in Table 4.11, model three has the highest adjusted R^2, which is .272. This indicated that model three was the most satisfactory model, which accounted for 27.2 % of the variance in the CELST scores. From the R square change in Table 4.11, it can also be inferred that CTA explained the largest amount of variance in the CELST scores, which is 23.4%. FL speaking behavioral anxiety accounted for 3.6 % of the variance in the CELST scores. FL speaking somatic anxiety explained 0.6% of the variance, which is quite negligible.

Table 4. 11

Model Summary

				Std. Error	Change Statistics				
Model	R	R Square	Adjusted R Square	of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.483[a]	.234	.232	1.837	.234	154.829	1	508	.000
2	.519[b]	.270	.267	1.795	.036	25.135	1	507	.000
3	.525[c]	.276	.272	1.789	.006	4.337	1	506	.038

a. Predictors: (Constant), CTA
b. Predictors: (Constant), CTA, FL speaking behavioral anxiety
c. Predictors: (Constant), CTA, FL speaking behavioral anxiety, FL speaking somatic anxiety

As seen in Table 4.12, ANOVA was also performed to verify the model fit of model three. As shown in Table 4.12, the results of ANOVA revealed that model three significantly predicts the CELST scores, $F = 64.296$, $p <.001$.

Table 4. 12

ANOVA of Model Three

Model		Sum of Squares	df	Mean Square	F	Sig.
3	Regression	617.633	3	205.878	64.296	.000[d]
	Residual	1620.218	506	3.202		
	Total	2237.851	509			

a. Dependent Variable: CELST score

d. Predictors: (Constant), CTA, FL speaking behavioral anxiety, FL speaking somatic anxiety

As demonstrated in Table 4.13, all three independent variables significantly contributed to model three ($p<.05$). All the VIF values were smaller than 2 and tolerance values were beyond .20, indicating the model data was not highly affected by multicollinearity and could be accepted. As shown in Table 4.13, the constant of the model is 19.717, the regression coefficient B is -.069, -.195, .078 for CTA, FL speaking behavioral anxiety and FL speaking somatic anxiety respectively. Therefore, the predictive equation model of the estimated CELST score could be formulated as follows:

Estimated CELST score = 19.717 + (-.069 * CTA) + (-.195 * FL speaking behavioral anxiety) + .078 * FL speaking somatic anxiety. This model also suggests that the higher test takers' CTA and FL speaking behavioral anxiety, the lower their CELST scores, whereas the higher their FL speaking somatic anxiety, the higher their CELST scores.

Table 4. 13

Coefficient of the Regression

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
3	(Constant)	19.717	.301		65.471	.000		
	CTA	-.069	.008	-.406	-8.252	.000	.591	1.693
	FL speaking behavioral anxiety	-.195	.036	-.251	-5.402	.000	.665	1.504
	FL speaking somatic anxiety	.078	.038	.099	2.082	.038	.629	1.590

a. Dependent Variable: CELST score

4.4 Qualitative Results of the Research Question Three: Reason Leading to Anxiety and Strategies to Alleviate it

The results of the quantitative data analysis above indicated that CTA, FLLA and FLSA did exist among the CELST test takers and had a mainly debilitating effect on their CELST performance. To complement the quantitative findings, individual semi-structured interviews were conducted with twelve students who had taken the CELST. The interview data also attempted to answer the third research question: From the test takers' perspectives, what reasons may lead to anxiety in the CELST and how do test takers alleviate the anxiety? The thematic analysis was adopted to analyze the interview data. The themes were presented below with supporting excerpts from the transcripts. In data reporting, interviewees with high, medium and low anxiety levels were labelled "High", "Medium" and "Low" respectively in the bracket after the excerpts. The number in the bracket is the corresponding page number in the transcripts.

4.4.1 Feelings and Attitudes Towards the CELST

All interviewees, including those claiming to have a high English proficiency, reported that part two role play and part three story retelling were relatively difficult for them. Ten out of twelve interviewees claimed that they experienced anxiety before, during and after the CELST. They tended to be most anxious right before taking the CELST in the test waiting rooms and worry about their performance in the upcoming test. However, they became less anxious and more focused in the middle of the CELST. They also worried about their test scores after the test. For example, Jayda reported that:

> *Before taking the CELST, we needed to stay in the test waiting room for a really, really long time, which really affected my mood. I tended to overthink and get really nervous. But when I actually did the CELST, I tried to calm down and get focused and felt less anxious. After the test, I was also afraid that I would fail the test. (Jayda, Medium, 63)*

Besides, most interviewees considered the CELST as anxiety-provoking because it tested two of their weakest language skills, English speaking and listening ability. For interviewees from the high and medium anxiety groups, they perceived their English listening and speaking competence as low and felt intimidated participating in some speaking and listening tests or classroom activities. For example, Andy reported that:

> *I always got good scores on the paper-based tests involving reading and writing, but the CELST assessed speaking and listening, which were not my strong points. During the CELST, sometimes I would tremble and my mind would go blank when the listening was too difficult. (Andy, High, 7)*

One third of the interviewees claimed that the listening part was more difficult and anxiety-provoking than the speaking part of the CELST, as Tom commented:

I think I was more nervous during the listening because I worried about missing out on information, especially in part two and part three. But when I comprehended the listening, speaking wouldn't be a problem to me as I just need to speak out the notes I took from the listening. (Tom, Medium, 90)

Half of the interviewees also reported that during the speaking part, they worried about not being able to express themselves or making grammar mistakes, especially in part two and part three. However, interviewees from the low level anxiety groups claimed that they didn't experience too much anxiety when speaking and listening to English during the CELST as they were quite confident in their English proficiency. Ben mentioned that:

I did feel anxious after I took the CELST waiting for the test results coming out. But when I was right in the middle of the CELST, I was not anxious and focused on doing the test because I believed my English competence was good enough to handle it (Ben, Low, 124).

4.4.2 Effect of Anxiety on the CELST Performance

Interviewees from three anxiety level groups all agreed that anxiety had an impact on their CELST performance. When it comes to the effect of CTA on performance, interviewees from the high and medium anxiety groups reported that task-irrelevant thoughts would sometimes occur in their minds during the CELST and interfere with their performance. Andrea commented that:

When I got caught up in thoughts like 'I was not doing quite well in part one, 'what if I failed the CELST', I got distracted from the listening clip for almost half a minute and my subsequent task performance suffered as well. (Andrea, High, *34).*

In terms of the effect of FLLA and FLSA, interviewees with high anxiety levels reported that the FLLA and FLSA had a debilitating effect on their CELST performance. They mentioned that they were afraid of getting negative evaluations from their teachers and peers when they were asked by teachers to give speaking answers to some simulated CELST practice in front of the whole class. When their English speaking and listening ability was being laughed at and poorly judged by others, they tended to avoid speaking and listening to English. This avoidance behavior would in turn negatively impact their real CELST performance. Andy shared that:

> *One time I was asked by the teachers to retell the part three story, I remembered my peers laughed at me when I mispronounced 'R' as 'L'. From that point on, I tended not to speak English in class and when I had to speak in the CELST, my heart pounded very fast and I stuttered a lot. (Andy, High, 10)"*

During the CELST, interviewees from the high and medium anxiety groups also mentioned that they became anxious because when they heard speaking answers from those sitting next to them, they were very worried that their own answers were not as good and accurate as others' and this disturbing thought also interfered with their speaking and listening performance.

Surprisingly, two thirds of the interviewees tended to think that a moderate level of behavioral anxiety, such as physiological arousal, helped with their performance, as stated below.

> *I think a little bit of anxiety can help me to take the test more seriously. At the beginning of the test, my heart pounded faster than usual and I felt like I had more energy and tended to be more focused during the test.* (Sky, Low, 112).

4.4.3 Reasons that Lead to Anxiety in the CELST

High Stakes Nature of the CELST. Interviewees from three anxiety level groups all attached great importance to the CELST as the CELST scores impacted university admission decisions. They worried that they only had one test attempt and if they failed to perform well in the CELST, they wouldn't get a decent score to get into a good college and would disappoint their parents. For example, Rain mentioned that

> *"I worried that If I screwed up the CELST, I wouldn't have a second chance to take the test and I wouldn't be able to go to my dream university. I was also afraid that I was not able to live up to the expectations of my parents. (Rain, High, 21)."*

Interviewees from the high and medium anxiety groups also reported that the increasing weight of CELST this year let them attach more importance to the CELST and made them feel more anxious about the CELST. As Susan said:

> *This year, the CELST score has increased. The increasing score actually gave me more pressure to prepare it. And when the date of the CELST was approaching, I felt very nervous because I felt like I didn't prepare enough to take the CELST. (Susan, Medium, 83)*

Noisy Testing Environment. Ten out of twelve of the interviewees expressed that the testing environment was noisy, disturbing and anxiety-provoking. During the CELST, dozens of test takers were arranged in the same testing classroom taking the test at the same time. When all the test takers were speaking simultaneously, the testing classroom became clamorous and they could even hear what those sitting next to them were saying. Interviewees reported that they were often distracted by the noisy testing environment and got nervous when they were giving their speaking answers.

(a) During the test, we sat very close to each other and the guy sat next to me spoke really loudly, and I could hear clearly what he was saying even with my headphones on. In part A, everyone was reading loudly, I got affected by the noise and stuttered. In part B, I also got nervous when I heard that others' speaking answers were different from mine. (Andy, High,13)

(b) I want to say that the noisy testing environment interfered with my thinking process. I also got stuck when it was too noisy and became a bit nervous. (Zack, Medium, 104)

(c) When the test takers around me spoke loudly, I got irritated and for a few seconds, I tended to put my focus on what they were saying rather than organize my own answer. (Lisa, Low, 141)

Insufficient Test-taking Strategies. All the Interviewees mentioned that they relied on test-taking strategies, such as translating skills and note-taking skills as they processed the CLEST tasks. However, two thirds of the interviewees believed that their test-taking strategies were insufficient, which sometimes influenced their CLEST performance and provoked anxiety during speaking and listening. For example, August explained how his poor note-taking skill could arouse anxious feelings in both speaking and listening.

"When I did the part three story retelling, I didn't know how to take quick notes of the vocabulary words of the listening story and I couldn't take efficient notes of the key points. So I couldn't follow the speed of the listening story and got anxious. And when it came to retelling, I looked at the messy and incomplete notes, I couldn't recall what the notes meant and I stuttered a lot and felt quite nervous (August, High, 50)."

Inappropriate Test Format and Difficulty of Part Two. All interviewees reported that they experienced some degree of anxiety in the part two role play. Their concerns of the part two role play can be summarized into two aspects. First, it is the inappropriate test format of part two. Interviewees claimed that the video played at the beginning of part two was disturbing, which could lead to anxiety. The video clip was a conversation between two people which usually took place in a classroom or hallway. The audio of the video contained important information for the answers to two subsequent questions. However, the video scene did not contain any content-relevant visual cues.

Interviewees complained that it was hard for them to comprehend the listening if they were to simultaneously watch the distracting scenes of the two people talking and moving in the video clip. They suggested that the scenes of the video clip could just be replaced by a static photo of the two speakers, which would be less distracting. Besides, they also complained that the speakers in the audio of the video were not native English speakers and the pronunciation of the speakers in the video at some points was weird and not clear. Some of the interviewees also mentioned that the audio and the scenes of the video were not matching, which could be a bit disturbing.

(a) When the video started to play, I watch the video but people in the video scenes talking and moving would distract me from listening and taking notes. As a result, I got nervous as I couldn't comprehend the listening when watching the video. Besides, the audio didn't match the scene in the video, which made me feel the conversation was ingenuine and a bit weird. (Zack, Medium, 101)."

(b) When I watched the scenes of the video for a few seconds, I immediately got distracted by the scenes in the video. I think the video clip could be replaced by a

photo or something. And to be honest, I don't know if it was my problem. The pronunciation of the speakers sometimes was a bit blurred. I felt like the pronunciation was not very native and standard (August, High, 52).

The second concern was about the difficulty of the part two. Interviewees mentioned that they felt quite apprehensive about the video clip because the video was played once only and the speed was very fast. They felt worried that they couldn't follow the video and miss the important information, especially when they didn't have much background knowledge of the video content.

Because the video played only once and was very fast, when I missed the key information, I worried that I couldn't answer the subsequent questions. (Rain, High, 23)

Time Constraints in Part Three. Over half of the interviewees complained that the time constraints in the part three story retelling was too short and they panicked when they couldn't organize their answers within the short time limits. The following extract presented this concern:

In part three, after listening to the story, there were only 60 seconds given to organize all the notes I'd taken and put them into a story. I think 60 seconds was way too short and time was up when I organized only half of the notes. I got super nervous and got stuck in some part when I retold the story. I think If I had a little bit more time, I would be less nervous and have a better performance. (Jayda, medium, 71)

4.4.4 Suggestions to Alleviate Anxiety and Make the CELST Less Anxiety-provoking

The interviewees revealed some useful strategies they took to alleviate CTA, FLLA and FLSA. In terms of strategies to deal with CTA, half of the interviewees mentioned some relaxation techniques such as having a light-hearted small talk with their classmates before the

CELST, having a positive self-talk and taking some nice deep breaths when feeling nervous during the CELST. In addition, over two thirds of the interviewees reported that reading out the English test instructions at the beginning of the CELST helped warm them up for the real CELST task. For instance, Jenny mentioned that:

> *Reading out the English instructions kind of opened up my mouth and made me get used to the test circumstance and feel more at ease speaking English (Jenny, low, 162).*

Finally, two interviewees with high anxiety levels suggested that they would be less test-anxious if they were given multiple attempts to take the CELST. As August explained it:

> *When we take the CELST once, the score from the first attempt is our final result. That's pretty daunting to me, because that single one test result can determine my future. I think maybe the CLEST can learn from IELTS, giving test-takers multiple attempts to do the test and choose the best score as the final result. This way, I think I would feel less anxious during the test. (August, High, 60)*

In terms of strategies to deal with FLLA and FLSA, four interviewees from the low anxiety group mentioned that doing mock CELST practice regularly and having more English exposure would make them feel more confident and at ease speaking and listening to English. The following extract illustrates this point:

> *I did the CLEST practice every week. Although the mock practice was not as formal as the real CELST, it helped me to get more familiar with the CELST and make the real CELST less nerve-racking to me. I also watched tons of English movies and imitated the native speakers' accents since I was little, I think this also helped (Sky, low, 121).*

What's more, one third of the interviewees mentioned that encouraging feedback from teachers and peers would contribute to making them feel less intimidated and improve their confidence when speaking and listening to English.

> *"It always feels good when I receive praises from teachers, saying that my pronunciation was very good. I think this made me feel more confident about my English and more chilled in speaking English. (Ben, Low, 131)*

In summary, the findings from the thematic analysis of interviews are consistent with the results of the quantitative data analysis. It confirmed that CTA, FLLA and FLSA did exist among the CELST test takers and had an impact on their test performance. In addition, the interviews also revealed some reasons that caused anxiety in the CELST. Finally, suggestions to alleviate CTA, FLLA and FLSA and make the CELST less anxiety-provoking were provided.

Chapter Five Conclusion

The major findings of the present study are summarized below based on the sequence of the research questions, followed by the discussion of the findings. Then, it illustrates the implications from the theoretical and practical perspectives, followed by the possible limitations and suggestions for future research.

5.1 The Characteristics of CTA, FLLA and FLSA among CELST Test Takers

This section addresses the first research question: What are the characteristics of cognitive test anxiety, foreign language listening anxiety and foreign language speaking anxiety among CELST test takers?

The present study first explored the status quo of CTA, FLLA and FLSA among CELST test takers. According to the descriptive statistics, the majority of CELST test takers reported a moderate level of CTA, indicating that they worried a lot about their performance and struggled with task-irrelevant thoughts during the CELST. As for FLLA and FLSA, the majority of test takers also reported experiencing FL listening cognitive anxiety and FL speaking cognitive anxiety at a medium level, suggesting that test takers lacked confidence in their English listening and speaking ability and worried about poor performance and negative evaluations when speaking and listening to English. Besides, most of the test takers experienced less somatic and behavioral dimensions of FLLA and FLSA than the cognitive dimensions of FLLA and FLSA, indicating that the majority of test takers had a relatively low level of physiological arousal and avoidance behavior related to English speaking and listening.

The study also explored CTA, FLLA and FLSA and their relationships with gender, major orientation, self-perceived English listening and speaking competence and frequency of language use related to English speaking and listening. The result of the independent T test revealed that females had a significantly higher level of CTA than males while there were no significant gender differences in terms of the overall FLLA and FLSA. However, at the subdimension level, females had a significantly slightly higher level of FL listening cognitive anxiety and FL speaking cognitive anxiety than males. Taken together, this indicated that female test takers tended to be more anxious than males during the CELST. Besides, the independent samples t-test revealed no significant difference in terms of the major orientation, suggesting that test takers from physics-oriented classes may experience a similar level of anxiety as those from the history-oriented classes. The results of the one-way ANOVA revealed that self-perceived English listening and speaking competence had a significant effect on test takers' CTA, FLLA and FLSA, which indicated that the higher test takers perceived their English listening and speaking competence, the lower the levels of CTA, FLLA and FLSA they would have. The correlation analysis revealed that frequency of language use related to listening and speaking had a significant negative correlation with CTA, FLLA and its subdimensions and FLSA and its subdimensions, suggesting that the higher frequency of language use test takers had, the lower their CTA, FLLA and FLSA.

5.1.1 Discussion of The Characteristics of CTA, FLLA and FLSA among CELST Test Takers

The results of the independent T test revealed that female test takers' CTA levels were significantly higher than those of male test takers. The results are consistent with meta-analysis studies that revealed females reported significantly higher levels of CTA than males (Hembree, 1988; von der Embse et al., 2018). The possible reason for the females' higher CTA might be that females tend to perceive evaluative situations as more threatening than males do and this perceived threat might be due to the self-doubt concerning their competence to perform well in an evaluating situation(Arch, 1987; Zohar, 1998). However, the finding did not support

Zeidner's (1990) assertion that the gender difference in test anxiety might be due in large part to different scholastic ability between genders because the independent T test also revealed no significant gender difference in terms of the CELST scores. In addition, the results indicated that females tended to have a slightly higher level of FLLA and FLSA than males but the differences were not statistically significant. This finding is aligned with previous studies that found no gender difference in FLLA (Elkhafaifi, 2005) and FLSA (Cheng et. al, 2007). However, at the subdimension level, the only slight significant gender differences were found in FL listening cognitive anxiety and FL speaking cognitive anxiety, indicating that females might worry more about poor performance and negative evaluations when speaking and listening to English.

In terms of major orientation, the results of the independent T test revealed that there were no significant differences in CTA, FLLA and its subdimension, FLSA and its subdimension.

The results of ANOVA found that CTA, FLLA and FLSA all significantly varied across different levels of self-perceived English listening and speaking competence groups, which indicated that the test takers who perceived their competence as low tended to have a higher level of CTA, FLLA and FLSA and vice versus. The results affirmed studies on foreign language anxiety that stated self-perception of language proficiency was closely and negatively correlated with anxiety (Chiang, 2006; MacIntyre, 1992; Truitt, 1995). It should be noted that about 65.8 % of the participants' CELST scores were above 16 out of 20, which is above the average level. However, 51.8% and 42.2% of participants rated their competence as "medium" and "low", indicating most of the participants underrated their actual performance. This low perception of the actual competence might lead to higher levels of CTA, FLLA and FLSA. The results also indicated that the higher the test takers perceived their English listening and speaking competence to be, the less anxious they were towards speaking and listening to English, and thus less test anxious in the CELST which assessed listening and speaking ability.

Regarding the frequency of language use related to English speaking and listening, the Pearson correlation revealed that the frequency of language use formed a medium significant correlation with CTA, FLLA and its subdimension, FLSA and its subdimensions. This indicated that test takers with a higher frequency of language use tended to have less CTA, FLLA and FLSA. This finding supported Baker and MacIntyre's (2000) study which found that Anglophone learners of French in a French immersion program with a higher frequency of language use had a lower level of anxiety in French than those in a non-French immersion program even though the two groups had the nearly same level of communication apprehension in English. The reason behind this might be that the increased exposure to the foreign language appears to give the learners more time and occasions to improve their actual proficiency and their ability to predict and confirm expectations and beliefs of their competence, and thus enhance their self-perceived foreign language competence. This triggers a set of behavior in which the learners feel less apprehensive about communicating in foreign language situations and thus be more competent in speaking and listening to a foreign language (Baker & MacIntyre, 2000). Simply put, the higher frequency of language use test takers had, the more likely they had higher actual or perceived English listening and speaking competence and tended to be less anxious in an evaluative situation that assessed their speaking and listening ability. The interview finding also confirmed that test takers with low anxiety levels tended to have higher frequency of language use related to listening and speaking than test takers with high and moderate levels of anxiety.

5.2 The Effects of CTA, FLLA and FLSA on CELST Test Takers' Scores

This section addresses the first research question: What are the effects of cognitive test anxiety, foreign language listening anxiety and foreign language speaking anxiety on CELST test takers' overall scores?

The results of ANOVA revealed that the CELST mean scores received by the high CTA, FLLA and FLSA groups were significantly lower than those received by the low and medium CTA, FLLA and FLSA groups, indicating that highly anxious test takers tended to score lower in the CELST. The correlation analysis further confirmed that CTA, FLLA and its subdimensions, FLSA and its subdimensions negatively moderately correlated with the CELST scores. The regression analysis further identified CTA, FL speaking behavioral anxiety as the negative predictors of the CELST scores while FL speaking somatic anxiety as the positive predictor of the CELST scores. The above results indicated that CTA, FLLA and its dimensions, FLSA and its dimensions all appeared to have significant effects on the CELST performance. While CTA and FL speaking behavioral dimension seemed to exert the most negative effects on the CELST performance, certain levels of FL speaking somatic anxiety may have a slight positive effect on the CELST performance.

5.2.1 Discussion of the Effect of CTA on Test Takers' CELST Scores

CTA was shown to be a negative predictor of the test takers' CELST scores, which is consistent with the previous studies (Cheng et al., 2014; Huang & Hung, 2013; Yoongoo, 2017; Zheng & Cheng, 2018) that revealed that CTA negatively predicted test takers' performance in foreign language tests. The finding that CTA negatively predicts test performance also supports the interference model by MacIntyre (1995) that suggests that task-irrelevant cognition such as worry causes declines in cognitive performance and further impairs task performance.

However, the finding was not congruent with In'nami's (2006) finding based on a structural equation model that revealed test anxiety did not affect Japanese university students' listening test performance on a TOEFL practice listening test. One possible explanation for this inconsistent finding lies in the different stakes of the tests between In'nami's (2006) study and the present study. The stakes of a test refer to the extent the test results impact test takers' future life. In'nami (2006) administered the adapted TOEFL listening test to 79 Japanese university students as a practice test to assess students' listening proficiency. The practice TOEFL test was neither part of the course grade, nor did it have any impact on students' life. It is not surprising that participants in In'nami's (2006) study might perceive a negligible level of test anxiety, which were not strong enough to negatively affect students' performance. Studies of test anxiety conducted in a low-stakes context might fail to capture the true effect of test anxiety. However, the CELST in the present study was high-stakes in nature and, to a large degree, affected test takers' university admission and future. The descriptive statistics also reported that the majority of the test takers did experience a certain degree of test anxiety during the CELST. Therefore, the test takers' performance was more likely to be affected by test anxiety in the present study with a high-stakes context and the relationship between test anxiety and foreign language performance may be better illustrated in this study.

5.2.2 Discussion of the Effect of FLLA on Test Takers' CELST Scores

The correlation analysis revealed that the CELST scores formed a significant negative moderate correlation with FLLA ($r=-.393$) and FL listening behavioral anxiety ($r=-.378$). The CELST scores also formed a significant negative small correlation with FL listening somatic anxiety ($r=-.255$) and FL listening cognitive anxiety ($r=-.286$), indicating that higher levels of FLLA tended to be associated with lower levels of test performance. The finding is consistent with previous studies that found a negative correlation between FLLA and listening test performance (Elkhafaifi, 2005; Kim, 2001; Golchi, 2012; Zhou, 2009). However, when the FLLA subdimensions were entered into the stepwise regression model with CTA and FLSA subdimensions, none of the FLLA subdimensions were shown to be significant predictors of the CELST scores. One possible explanation for the absence of FLLA dimensions as significant predictors might be due to the fact that the construct of FLLA has conceptual overlapping with the construct of FLSA. The correlation analysis also supported that FLLA and FLSA conceptually overlapped with each other as FLLA and FLSA highly correlated with each other ($r=.809$). Besides, the correlation also indicated that FLSA ($r=-.413$) was more strongly correlated with the CELST scores than FLLA ($r=-.393$). Therefore, when FLLA subdimensions were entered together with the FLSA subdimensions in the regression model, the variance in the CELST scores explained by the FLLA might be replaced by FLSA which was more strongly correlated with the CELST scores, resulting in FLLA not accounting uniquely and significantly to the variance in the CELST scores.

In conclusion, FLSA might have a stronger effect on the CELST performance than FLLA. However, this did not indicate that FLLA exerted no effect on the CELST performance because: (a) FLLA moderately negatively correlated with the CELST scores; (b) the interview data revealed that the interviewees also perceived the listening part of the CELST as anxiety-provoking and FLLA negatively affected their performance.

5.2.3 Discussion of the Effect of FLSA on Test Takers' CELST Scores

The correlation analysis revealed that the CELST scores moderately negatively correlated with FLSA ($r=-.413$), FL speaking behavioral anxiety ($r=-.419$) and FL speaking cognitive anxiety ($r=-.335$). FL speaking somatic anxiety formed a small positive correlation with the CELST scores ($r=-.254$). The regression analysis further revealed that FL speaking behavioral anxiety was a negative predictor of the CELST score, indicating the more withdrawal and avoidance behavior related to speaking English test takers had, the lower they scored in the CELST.

However, the regression analysis revealed that FL speaking somatic anxiety was a positive predictor of the CELST scores, indicating that the more physiological arousal test takers had, the higher they scored in the CELST. This positive regression coefficient contradicted with the negative correlation coefficient between FL speaking somatic anxiety and the CELST score. According to Falk & Miller (1992), the inconsistent sign between the regression coefficient and the correlation coefficient might be due to the suppression effect. One reason for the suppression effect is that there are two or more variables that contain the same information and are therefore redundant. The redundancy might change the sign of the relationship, resulting in a negative correlation coefficient and a positive regression coefficient.

Another more likely explanation is that the positive regression coefficient was right and FL speaking somatic anxiety did have a positive effect on CELST test takers performance. According to Cassady & Johnson (2002), a moderate level of emotionality, similar to FL speaking somatic anxiety was considered benign for exams. The thematic analysis of the interview data also revealed that two thirds of the interviewees believed that a certain level of physiological arousal, such as faster heart rate actually helped them to be more focused on the task at hand and facilitated their performance. The Yerkes-Dodson Law of anxiety in section 2.1.1 also suggested that anxiety can be helpful when it is at its optimal level. The descriptive statistics revealed that test takers might have an optimal level of FL speaking somatic anxiety as the majority of test takers experienced low to moderate levels of FL speaking somatic anxiety. Therefore, it can be concluded that a certain level of FL speaking somatic anxiety may have a positive effect on the CELST test takers' performance. However, it should be noted that the positive effect of FL speaking somatic anxiety might be quite slight as FL speaking somatic anxiety only accounted for 0.6 % of the variance in the CELST scores, which is rather small.

Another point worth discussing here is that the cognitive dimension of FLSA didn't appear as a significant predictor of the CELST scores. The absence of FL speaking cognitive anxiety might be due to its conceptual overlapping with the CTA. That is, the construct of CTA may not only contain test takers' worry over the general performance of the CELST, but also cover worry over the specific speaking ability. The correlation analysis also supported that CTA and FL speaking cognitive anxiety formed a moderate correlation with each other ($r=.634$), suggesting that CTA and FL speaking cognitive anxiety might be conceptually overlapped yet also be independent from each other. Therefore, the amount of variance that FLLA explained in the CELST score might be replaced by the CTA, resulting in FL speaking cognitive anxiety not significantly and uniquely explaining the variance in the CELST score. Thus, it can be inferred that FL speaking cognitive anxiety still exerted an effect on the CELST performance but its effect may not be as strong as the CTA to make it a significant predictor.

5.3 Reasons that Lead to Anxiety in the CELST and Strategies to Alleviate it

The results of the thematic analysis of the interviews confirmed the existence and interference of the CTA, FLLA and FLSA among CELST test takers. Besides, the results revealed reasons that lead to anxiety in the CELST, including the high-stakes nature of the CELST, a noisy testing environment, insufficient test-taking strategies, the inappropriate test format and difficulty of part two, the time constraints in part three. The results also revealed helpful strategies to alleviate anxiety from the test takers' perspectives, including relaxation techniques, reading out the English test instructions at the beginning, multiple test attempts, higher frequency of language use, and encouraging feedback from teachers and peers.

5.3.1 Discussion of the Reasons that Lead to Anxiety in the CELST and Strategies to alleviate it

This section addresses the first research question: From the test takers' perspectives, what reasons may lead to anxiety in the CELST and how do test takers alleviate the anxiety?

The results of the thematic analysis also confirmed the findings of one of the very few validation studies of the CELST conducted by Zhan and Wan (2016). They qualitatively validated CELST through the test takers' perspectives by conducting retrospective interviews with 20 test takers. Even though the study did not focus specifically on the factors that caused anxiety, the content analysis of the interviews in Zhan and Wan's (2016) study revealed that students expressed similar concerns about the noisy physical testing conditions, the inappropriateness of test format and time constraints. However, differing from the test takers in Zhan and Wan's (2016) study who thought that the weight of CELST was comparatively low, the present study revealed that test takers believed the weight was relatively heavy and gave them a considerable amount of pressure preparing for the CELST. The reason might be that the weight of the CELST the test takers of the present study took has increased this year. In addition, the present study also revealed strategies to alleviate CTA, FLLA and FLSA, such as relaxation techniques, receiving encouraging feedback, allowing multiple test attempts and higher frequency of language use. To sum up, the participants provided valuable perceptions on what might cause anxiety from the test content, task properties, the process of preparing and taking the test and strategies to cope with anxiety. Their perceptions of the CELST can have implications for refining the CELST and making it less anxiety-provoking, thus improving the construct validity (Storey, 1997).

5.2 Implications of the Study

5.2.1 Theoretical Implications

To the best of my knowledge, the present study is not only one of the few studies that empirically and systematically investigated the effects of test anxiety and language-skill-specific-anxiety in high-stakes foreign language tests on Chinese EFL learners, but also the first ever empirical study that explored the effects of CTA, FLLA and FLSA together on Chinese EFL learners in one single investigation. Besides, among the empirical studies of the CELST, the present study is the first one that focused specifically on the effect of anxiety and revealed reasons that lead to anxiety in the CELST and strategies to alleviate it from the test takers' perspectives. Therefore, the findings generated from this study have greatly enriched the literature gap in this specific domain and significantly contributed to a more in-depth understanding of how CTA, FLLA and FLSA affect high school students in high-stakes foreign language tests.

5.2.2 Practical Implications

The results of qualitative analysis has provided evidence for the existence of CTA, FLLA and FLSA among the CELST test takers and their mainly debilitating effects on the test performance. However, the existence and interference of the anxiety may not have been given sufficient attention by the stakeholders. As the CELST has high-stakes impacts on millions of test takers' future, different stakeholders involved in the CELST, including policymakers, the CELST test developers, EFL teachers and the future CELST test takers are advised to make efforts to address the adverse effect of anxiety.

Advice for Policymakers and the CELST Test Developers. Policymakers and the CELST test developers are recommended to build test tasks and settings that can elicit test takers' best performance by minimizing the negative effect of anxiety (Bachman & Palmer, 1996). They could consider addressing the following factors that cause anxiety in the CELST, which were revealed by the interview data: a noisy testing environment, inappropriate test format and difficulty of part two, time constraints in part three and one test attempt.

In terms of the noisy testing environment, it has to be admitted that it is virtually impossible to eliminate the clamorous background noise because there are always test takers speaking simultaneously during the test. However, it is possible to reduce the effect of the loud background noise. One possible solution is to arrange fewer test takers in one testing classroom or assign students to sit far away from each other. Another possible solution is to equip test takers with better noise-canceling headphones to reduce loud noise, allowing them to pay more attention to the task.

Regarding the inappropriate format reported by the interviewees, test developers could consider replacing the video clip in the part two role play with a static photo of two people communicating. As the video in part two does not contain any content-relevant visual cues, it is not helping to foster test takers' understanding of the listening input in any way, but distracts their attention from listening. A static photo is not as distracting as a video and it also serves the same purpose of setting up a situation for the role play as a video does. Concerning the difficulty of the part two listening material, test developers are advised to select topics which test takers have enough background knowledge on so that it is fair to every test taker and does not necessarily evoke anxiety.

As for the time constraints in part three story retelling, it was reported that the time limit of sixty seconds for test takers to organize their notes for retelling the story was too short and anxiety-arousing. It is understandable that the speed of organizing notes into a speech can be one of the assessing criteria. However, the measure of test takers' real speaking ability is a more important assessing criterion. Insufficient time for organizing notes limits test takers to demonstrate their speaking ability when retelling the story. Therefore, test developers could consider extending the time limit to, say, ninety seconds such that test takers have more time to organize their notes and feel less anxious. This way, test takers' speaking ability can be adequately assessed.

It was also reported by the interviewees that the CELST would be less nerve recking if test takers were allowed to have two or more test attempts. It is admitted that more test attempts require much more resources and funding. However, due to the high-stakes impacts of the CELST, test takers deserve to have at least a second test trial and choose the best scores as their final scores. This way, the CELST is not only less anxiety-provoking but provides more learning opportunities for test takers to learn from past test attempts.

To conclude, the CELST test developers are well advised to take into account the construct-irrelevant variance such as factors that cause anxiety during the test development process. By diminishing the effect of construct-irrelevant variance that may undermine test performance, test takers' true language proficiency can be accurately reflected by their scores (Cheng et al., 2014), thus improving the construct validity of the test.

Advice for EFL Teachers and Future CELST Test Takers. Although the study has found that test takers experienced anxiety before, during and after the CELST, little effort has been put in EFL teaching to assist these test takers to alleviate such anxiety. EFL teachers' and test takers' awareness of the effect of anxiety and the relevant coping strategies has been rather limited (He, 2017). Therefore, in-service and pre-service EFL teachers as well as future test takers should be well informed of the role played by CTA, FLLA and FLSA. Their awareness and mastery of the anxiety-coping strategies should also be raised. They are also advised to take concrete action to address not only anxiety during the CELST, but also anxiety experienced in everyday EFL teaching and learning. EFL teachers may start to alleviate students' anxiety by referring to the reasons that cause anxiety and coping strategies in this study.

First, students should be guided with a sufficient number of practice exercises that are specifically tailored to improve their test-taking skills in the CELST. The qualitative results demonstrated that the majority of interviewees lacked confidence in their test-taking skills and believed that their insufficient skills such as inefficient note-taking skills could cause anxiety during the CELST. Therefore, more systematic instructions targeted at test-taking skills should be provided.

Second, teachers should provide encouraging feedback at every step of the listening and speaking instructions. Students' anxiety over foreign language speaking and listening does not form overnight. Recurring failures in the past can accumulate anxiety about English listening and speaking and fear of the CELST which assesses their listening and speaking ability. The adverse effect of anxiety can further impair students' performance and make them even more anxious in future foreign language learning and testing. Thus, teachers are advised to identify highly anxious students and give them more encouraging feedback and generous support after they finish each exercise, even when they do not perform very well. Reasons that lead to anxiety and coping strategies may vary among individual students. Therefore, it is the responsibility of teachers to identify the source of their students' anxiety and come up with strategies to alleviate it. Although providing such guidance and support requires more time and effort from teachers, it is worthwhile when it works to reduce the students' stress and apprehension in foreign language speaking and listening, and thus improve their confidence and performance in a foreign language situation.

The findings also have practical implications for future CELST test takers. The quantitative findings show that higher frequency of language use and self-perceived competence have a significant positive impact on reducing levels of anxiety. Thus, highly anxious students are advised to increase their frequency of language and not underestimate their own competence. Besides, they are encouraged to learn how to manage the anxiety on their own. For example, they can try some relaxation techniques revealed in this study such as deep breathing and positive self-talk.

5.3 Limitations and Suggestions for Future Research

A few limitations should be taken into consideration when interpreting the findings of the present study. Besides, suggestions are also provided for future research.

First, due to the self-reported measures of anxiety, the responses on the questionnaires depend totally on the participants' capabilities which might be subject to distortions. For example, Participants might be unwilling or unable to report accurately on their affective state.

Second, although the FLLAS and FLSAS are efficient and valid measures of FLLA and FLSA, the number of items in these two scales is relatively small and might not fully represent the constructs of FLLA and FLSA. Therefore, future research are recommended to employ scales with a more sufficient number of items when time and resources permit.

Third, the findings only indicate a single sample in Shenzhen city, a modern metropolis that enjoys rich educational resources. Therefore, caution is required to generalize the findings to samples in other cities or rural areas that might not have many educational resources. Future empirical studies are needed to investigate the effect of anxiety on CELST performance in other areas of Guangdong province, China.

Fourth, the findings from the correlation analysis cannot indicate any causal relationships because there are many unidentified variables that can influence the relationships between the scores of the CTAS, FLLAS and FLSAS and the CELST scores.

Fifth, as this study demonstrated that CTA, FLLA and FLSA can interfere test takers in the CELST, future research should investigate more in depth what interventions work to help test takers alleviate the anxiety, and thus minimize the negative effect of anxiety during the CELST and other high-stakes foreign language listening and speaking tests.

5.4 Conclusion

This study investigated the characteristics of CTA, FLLA and FLSA among CELST test takers and the effects of CTA, FLLA and FLSA on CELST test takers performance. The findings confirmed the existence and interference of anxiety. In addition, the present study explored the reasons that lead to anxiety in the CELST and strategies to alleviate it from the test takers' perspectives. This study served to fill the literature gap as no previous research has examined the effects of CTA, FLLA and FLSA on CELST test takers in one single investigation. The qualitative findings of the study also provide millions of CELST stakeholders a finer understanding on the source of anxiety in the CELST and some valuable coping strategies.

References

Alpert, R., & Haber, R. N. (1960). Anxiety in academic achievement situations. *The Journal of Abnormal and Social Psychology*, *61*(2), 207-215.

Amiryousefi, M., & Tavakoli, M. (2011). The relationship between test anxiety, motivation and MI and the TOEFL iBT reading, listening and writing scores. *Procedia - Social and Behavioral Sciences*, *15*, 210–214.

Arch, E. C. (1987). Differential responses of females and males to evaluative stress: Anxiety, self-esteem, efficacy and willingness to participate. In R. Schwarzer, H. M. van der Ploeg, & C. D. Spielberger (Eds.), *Advances in test anxiety research, 5,* 97-106.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.

Baker, S. C., & MacIntyre, P. D. (2000). The role of gender and immersion in communication and second language orientations. *Language Learning*, *50*(2), 311–341.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research*, *6*(1), 97–113.

Ça, S. (2015). *Examining EFL students' foreign language speaking anxiety: The case at a Turkish state university*. *199*, 648–656.

Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, *27*(2), 270–295.

Cattell, R. B., & Scheier, I. H. (1960). Stimuli related to stress, neuroticism, excitation, and anxiety response patterns: Illustrating a new multivariate experimental design. *The Journal* *of Abnormal and Social Psychology*, *60*(2), 195-204.

Cheng, L. (2008). The key to success: English language testing in China. *Language Testing*, *25*(1), 15–37.

Cheng, L. (2008). Washback, impact and consequences. In: Shohamy E, Hornberger NH (eds) *Encyclopedia of Language and Education, Vol. 7: Language Testing and Assessment.* New York: Springer, 349-364.

Cheng, L., Deluca, C (2011). Voices from test-takers: further evidence for language assessment validation and use. *Educational Assessment, 16*(2), 104-122.

Cheng, L., Klinger, D., Fox, J., Doe, C., Jin, Y., & Wu, J. (2014). Motivation and test anxiety in test performance across three testing contexts: The CAEL, CET, and GEPT. *TESOL Quarterly*, *48*(2), 300–330.

Cheng, Y. S. (2004). A measure of second language writing anxiety: Scale development and preliminary validation. *Journal of Second Language Writing*, *13*(4), 313–335.

Cheng, Y. S. (2017). Development and preliminary validation of four brief measures of L2 language-skill-specific anxiety. *System*, *68*, 15–25.

Cheng, Y. J., Liu, D. J. (2010). 英语听力焦虑成因探讨及应对策略实证研究 [An empirical study on the causes of English listening anxiety and coping strategies]. *Foreign Literature, 26*(1), 141-144.

Chiang, Y. show (2006). *Connecting two anxiety constructs: An interdisciplinary study of foreign language anxiety and interpretation anxiety* [Unpublished doctoral dissertation]. The University of Texas at Austin.

Cohen, S. (1992). Stress, social support, and disorder. In H.O.F. Veiel & U. Baumann (Eds.), *The Meaning and Measurement of Social Support* (pp. 109-124). New York: Hemisphere.

Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis*. Psychology press.

Ehrman, M. E. (1996). *Understanding second language learning difficulties*. Sage.

Elkhafaifi, H. (2005). Listening comprehension and anxiety in the Arabic language classroom. *Modern Language Journal*, *89*(2), 206–220.

Ellis, R. (1994). *The study of second language acquisition*. Oxford University.

Deluca, C., Cheng, L., Fox, J., Doe, C., & Li, M. (2013) Putting testing researchers to the test: An exploratory study on the TOEFL IBT. *System, 41*(3), 663-676.

Fabrigar, L. R., & Wegener, D. T. (2011). *Exploratory factor analysis*. Oxford University Press.

Falk, R. F., & Miller, N. B. (1992). *A primer for soft modeling.* University of Akron Press.

Fujii, Y. (1993). Construction of a test influence inventory (TII). *Shinrigaku Kenkyu: The Japanese Journal of Psychology*, *64*(2), 135–139.

Gebre, B., & Wako, G. (2017). Relationship Between Foreign Language Listening Anxiety and Listening Comprehension. *International Journal of Foreign Language Teaching & Research*, *5*(19), 11–23.

Haladyna, T. M., & Downing, S. M. (2004). Construct-Irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, *23*(1), 17–27.

Hasibuan, A. R., & Irzawati, I. (2016). Students' speaking anxiety on their speaking performance: A study of EFL learners. *Advances in Social Science, Education and Humannites Research, 394,* 101-106

He, D. (2017). How to cope with foreign language speaking anxiety effectively? The case of university students in China. *Electronic Journal of Foreign Language Teaching*, *14*(2), 159–174.

He, D. (2018). *Foreign language learning anxiety in China: Theories and applications in*

English language teaching. Singapore: Springer.

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, *58*(1), 47–77.

He, Q. G., Han, X. L., & Cheng, Y. P., (2007). 英语课堂焦虑对大学生口语表达的影响[The influence of English classroom anxiety on the oral expression of college students]. *Language Teaching and Research, 01*, 89-95.

Hewitt, E., & Stephenson, J. (2012). Foreign language anxiety and oral exam performance: A replication of Phillips's MLJ study. *Modern Language Journal*, *96*(2), 170-189.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*(2), 125–132.

Huang, H. T. D., & Hung, S. T. A. (2013). Comparing the effects of test anxiety on independent and integrated speaking test performance. *TESOL Quarterly*, *47*(2), 244-269.

In'nami, Y. (2006). The effects of test anxiety on listening test performance. *System*, *34*(3), 317–340.

Javanbakht, N., & Hadian, M. (2014). The effects of test anxiety on learners' reading test performance. *Procedia - Social and Behavioral Sciences*, *98*, 775–783.

Jerusalem, M., & Schwarzer, R. (1992). Self-efficacy as a resource factor in stress appraisal processes. In R. Schwarzer (Eds.), *Self-Efficacy: Thought Control of Action* (pp. 195-213). Hemisphere, Washington, DC.

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, *35*(4), 401–415.

Kim, J. H. (2001). Foreign language listening anxiety: A study of Korean students learning English. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, *62*(2), 497-509.

Kim, J. (2000). *Foreign language listening anxiety: A study of Korean students learning English* [Unpublished doctoral dissertation]. University of Texas at Austin.

Lang, P. J. (1971). The application of psychophysiological methods to the study of psychotherapy and behavior modification. In *Handbook of Psychotherapy and Behavior Change* (pp. 75–125). New York: Wiley.

Larson, R., & Csikszentmihalyi, M. (2014). The experience sampling method. In *Flow and the foundations of positive psychology* (pp. 21–34). Springer, Dordrecht.

Levitt, E. E. (2015). *The psychology of anxiety*. Routledge.

Lu, H. Y. (2010). 非英语专业大学生英语口语焦虑与英语学习策略的相关性[Correlation between English speaking anxiety and English learning strategies of non-English majors] . *Foreign Language Research, 05,* 65-71.

Lushene, S. G., Gorsuch, R. L., & Spielberger, C. D. (1983). *Manual for the state-trait anxiety inventory*. Palo Alto: Consulting psychologists Press.

MacIntyre, P. D. (1995). How does anxiety affect second language learning? A reply to Sparks and Ganschow. *The Modern Language Journal*, *79*(1), 90–99.

MacIntyre, P. D., & Gardner, R. C. (1991). Methods and results in the study of foreign language anxiety: A review of the literature. *Language Learning*, *41*(1), 85–117.

MacIntyre, P. D., Noels, K. A., & Clément, R. (1997). Biases in self-ratings of second language proficiency: The role of language anxiety. *Language Learning*, *47*(2), 265–287.

MacIntyre, P. D. (1992). *Anxiety and language learning and stages of processing perspective* [Unpublished doctoral dissertation]. University of West Ontario, London, Canada.

MacIntyre, P. D, & Gardner, R. C. (1989). Anxiety and second-language learning: Toward a theoretical clarification. *Language Learning*, *39*(2), 251–275.

MacIntyre, P. D, & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning*, *44*(2), 283–305.

McDonald, A. S. (2001). The prevalence and effects of test anxiety in school children. *Educational Psychology*, *21*(1), 89–101.

Mede, E. (2017). The predictor roles of speaking anxiety and English self efficacy on foreign language speaking anxiety. *Journal of Teacher Education and Educators*, *6*(1), 117–131.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, *18*(2), 5–11.

Meyer, R. (1984). Listen my children, and you shall hear... *Foreign Language Annals*, *17*(4), 343.

Mohammadi Golchi, M. (2012). Listening anxiety and its relationship with listening strategy use and listening comprehension among Iranian IELTS learners. *International Journal of English Linguistics*, *2*(4), 115-128.

Moss, P.A., Girard, B. J., & Haniford, L. C. (2006). Validity in educational assessment. *Review of Research in Education, 30,* 109-162.

Onwuegbuzie, A. J., Bailey, P., & Daley, C. E. (2000). The validation of three scales measuring anxiety at different stages of the foreign language learning process: The input anxiety scale, the processing anxiety scale, and the output anxiety scale. *Language Learning*, *50*(1), 87–117.

Oxford, R. L., & Burry-Stock, J. A. (1995). Assessing the use of language learning strategies worldwide with the ESL/EFL version of the Strategy Inventory for Language Learning (SILL). *System*, *23*(1), 1–23.

Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing*, *20*(1), 26–56.

Phillips, E. M. (1992). The effects of language anxiety on students' oral test performance and attitudes. *The Modern Language Journal*, *76*(1), 14–26.

Ruebush, B. K. (1960). Interfering and facilitating effects of test anxiety. *The Journal of Abnormal and Social Psychology*, *60*(2), 205-212.

Saito, Y., Garza, T. J., & Horwitz, E. K. (1999). Foreign language reading anxiety. *The Modern Language Journal*, *83*(2), 202–218.

Salehi, M., & Marefat, F. (2014). The effects of foreign language anxiety and test anxiety on foreign language test performance. *Theory and Practice in Language Studies*, *4*(5), 931–940.

Salem, A. A. M. S., & Al Dyiar, M. A. (2014). The relationship between speaking anxiety and oral fluency of special education Arab learners of English. *Asian Social Science*, *10*(12), 170-176.

Sarason, I. G. (1975). *The test anxiety scale: Concept and research.* In I. G. Sarason & C.D. Spielberger (Eds.), Stress and anxiety (pp. 27-44). Washington, DC: Hemisphere.

Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: reactions to tests. *Journal of Personality and Social Psychology*, *46*(4), 929-938.

Sarason, I. G., & Stoops, R. (1978). Test anxiety and the passage of time. *Journal of Consulting and Clinical Psychology*, *46*(1), 102-109.

Scovel, T. (1978). The effect of affect on foreign language learning: A review of the anxiety research. *Language Learning*, *28*(1), 129–142.

Serraj, S., & bt. Noordin, N. (2013). Relationship among Iranian EFL students' foreign language anxiety, foreign language listening anxiety and their listening comprehension. *English Language Teaching*, *6*(5), 1-12.

Spielberger, C. D. (1976). The nature and measurement of anxiety. In *Cross-Cultural Anxiety* (pp.3-27). Washington D.C.: Hemisphere.

Spielberger, C. D. (1983). *State-trait anxiety inventory for adults*. Palo Alto, CA: Mind Garden.

Storey, P. (1997). Examining the test-taking process: A cognitive perspective on the discourse cloze test. *Language Testing*, *14*(2), 214–231.

Tasnimi, M. (2009). Affective factors: Anxiety. *Journal of Pan-Pacific Association of Applied Linguistics*, *13*(2), 117–124.

Tobias, Sheila. (1986). Anxiety and cognitive processing of instruction. *Self-Related Cognition in Anxiety and Motivation* (pp. 35–54). Psychology Press.

Tobias, Sigmund. (1979). Anxiety research in educational psychology. *Journal of Educational Psychology*, *71*(5), 573-582.

Truitt, S. N. (1995). *Anxiety and beliefs about language learning: A study of Korean university students learning English* [Unpublished doctoral dissertation]. The University of Texas at Austin.

Tsai, Y.-C., & Li, Y.-C. (2012). Test anxiety and foreign language reading anxiety in a reading-proficiency test. *Journal of Social Sciences*, *8*(1), 95–103.

von der Embse, N., Jester, D., Roy, D., & Post, J. (2018). Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. In *Journal of Affective Disorders* (pp. 483–493). Elsevier B.V.

Wang, S. Y., & Cha, K. W. (2019). Foreign language listening anxiety factors affecting listening performance of Chinese EFL learners. *Journal of Asia TEFL*, *16*(1), 121-134.

Williams, B., Onsman, A., & Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine*, *8*(3), 12-19.

Williams, J. E. (1991). Modeling test anxiety, self concept and high school students' academic achievement. *Journal of Research & Development in Education, 25, 51-57*.

Woodrow, L. (2006). Anxiety and speaking English as a second language. *RELC Journal*, *37*(3), 308–328.

Wu, Z. Y. (2021). Investigating the effect of English anxiety on senior high school students' writing anxiety [Unpublished master thesis]. South China Normal University.

Yan, J. X., & Horwitz, E. K. (2008). Learners' perceptions of how anxiety interacts with personal and instructional factors to influence their achievement in english: A qualitative analysis of EFL learners in China. *Language Learning*, *58*(1), 151–183.

Yoongoo Jason Nam. (2017). Exploring relations between Korean test takers' anxiety and performance on the test of English for international communication in listening and reading comprehension (TOEIC LR) [Unpublished master thesis]. Queen's University.

Yuh, S. C., Horwitz, E. K., & Schallert, D. L. (1999). Language anxiety: differentiating writing and speaking components. *Language Learning*, *49*(3), 417–446.

Yukie Aida. (1994). Examination of Horwitz, Horwitz, and Cope's construct of foreign language anxiety: The case of students of Japanese. *The Modern Language Journal*, *78*(2), 155–168.

Yu, X. L. (1999). 成人学生英语课堂焦虑感与听力理解成绩的关系[The relationship between adult students' English classroom anxiety and listening comprehension performance]. *Language Teaching and Research, 02,* 132-143.

Zeidner, M. (1998). *Test anxiety: The state of the art*. New York, NY: Plenum Press.

Zhan, Y., & Wan, Z. H. (2016). Test takers' beliefs and experiences of a high-stakes computer-based English listening and speaking test. *RELC Journal*, *47*(3), 363–376.

Zheng, Y., & Cheng, L. (2018). How does anxiety influence language performance? From the perspectives of foreign language classroom anxiety and cognitive test anxiety. *Language*

Testing in Asia, *8*(1), 1-19.

Zhou, D. D. (2003). 二语课堂中的听力焦虑感与情感策略[Listening anxiety and emotional strategies in the second language classroom]. *Foreign Language Teaching and Learning, 03,* 22-29.

Zhou, X. L. (2009). 英语课堂听力焦虑的成因及其与四级听力成绩的关系[The causes of listening anxiety in English classroom and its relationship with CET-4 listening scores]. *Chongqing University Social Science Journal, 9*(01), 137-141.

Zohar, D. (1998). An additive model of test anxiety: Role of exam-specific expectations. *Journal of Educational Psychology*, *90*(2), 330-340.

Appendix A: Results for T Tests

Table 6. 1

Results of Independent T test between Gender in CTA, FLLA and FLSA

	Gender	N	Mean	Std. Deviation	Sig.	t
CTA	Male	230	43.2478	12.23512	.027	-.2.222
	Female	280	45.6857	12.40568		
FLLA	Male	230	21.9043	6.49013	.052	-1.944
	Female	280	23.0000	6.20036		
FLSA	Male	230	21.5391	6.99131	.206	-1.266
	Female	280	22.3286	7.02473		
FL listening cognitive anxiety	Male	230	9.0174	3.26282	.000	-3.581
	Female	280	10.0179	3.03440		
FL listening somatic anxiety	Male	230	6.5826	2.46719	.090	-1.700
	Female	280	6.9750	2.69263		
FL listening behavioral anxiety	Male	230	6.3043	2.57354	.173	1.366
	Female	280	6.0071	2.33563		
FL speaking cognitive anxiety	Male	230	8.5696	3.27781	.006	-2.770
	Female	280	9.3679	3.20680		
FL speaking somatic anxiety	Male	230	6.2652	2.41973	.089	-1.702
	Female	280	6.6679	2.83988		
FL speaking behavioral anxiety	Male	230	6.7043	2.77850	.085	1.724
	Female	280	6.2929	2.60134		

Table 6. 2

Results of Independent T test between Gender on CELST scores

	Gender	N	Mean	Std. Deviation	Sig.	t
CELST Score	Male	230	15.74	2.127	.698	- 1.434
	Female	280	16.01	2.068		

Table 6. 3

Results of T test between Major and the CELST Scores

	Major	N	Mean	Std. Deviation	Sig.	t
CTA	History	291	45.0000	12.37991	.385	.870
	Physics	219	44.0365	12.37927		
FLLA	History	291	22.6942	6.23886	.441	.772
	Physics	219	22.2557	6.50042		
FLSA	History	291	22.3711	7.03928	.139	1.481
	Physics	219	21.4429	6.96058		
FL listening cognitive anxiety	History	291	9.6873	3.07971	.323	.989
	Physics	219	9.4064	3.29917		
FL listening somatic anxiety	History	291	6.9491	2.48491	.625	-.490
	Physics	219	6.8630	2.74605		
FL listening behavioral anxiety	History	291	6.2577	2.43797	.215	1.240
	Physics	219	5.9863	2.45786		
FL speaking cognitive anxiety	History	291	9.1821	3.21277	.164	1.393
	Physics	219	8.7763	3.31527		
FL speaking somatic anxiety	History	291	6.5842	2.66049	.339	.957
	Physics	219	6.3562	2.66838		
FL speaking behavioral anxiety	History	291	6.6048	2.67398	.221	1.225
	Physics	219	6.3105	2.70311		

Appendix B: Questionnaire

Note: The questionnaires distributed to the participants were in Chinese only. The English translation of the questionnaire is provided here for reference.

此问卷旨在研究焦虑对广东高考英语听说考试表现的影响。此项研究有重大研究价值，您对此调查问卷的作答将有助于该领域的研究，并帮助未来的考生更好地为听说考试做准备。请您按照实际情况填写该问卷，谢谢您的合作。(The purpose of this questionnaire is to investigate the effect of anxiety on the CELST performance. This research has important implications. Your responses to this questionnaire will facilitate research in the field and help future test takers better prepare for the CELST. Please fill in the questionnaire according to the actual situation. Thank you for your cooperation.)

个人基本信息(Demographic Information)

1. 性别(Gender):

   男性(Male) □ 女性(Female) □

2. 您的学生证号码是(Your student ID number is)___

3. 您是历史类学生还是物理类学生？(Are you a history-oriented student or a physics-oriented student?)

   历史类(History-oriented) □ 物理类(Physics-oriented)□

4. 您什么时候开始学英语的？（When did you start to learn English?）

   幼儿园(Kindergarten) □ 一年级(Grade 1) □ 二年级(Grade 2) □ 三年级(Grade 3) □ 三年级后(After Grade 3) □

5. 过去一年您在英语课堂上说英语的频率是？(In the past year, how often did you speak English in the English class?)

从不(rarely)□ 偶尔(occasionally)□ 有时(sometimes)□ 经常(often)□ 总是(always)□

6. 过去一年您在课余时间说英语的频率是？(In the past year, how often did you speak English outside of English class?)

从不(rarely)□ 偶尔(occasionally)□ 有时(sometimes)□ 经常(often)□ 总是(always)□

7. 过去一年您平常自己练习模拟英语听说考试的频率是？（In the past year, how often did you practice the mock CELST? ）

从不(rarely)□ 偶尔(occasionally)□ 有时(sometimes)□ 经常(often)□ 总是(always)□

8. 您看英语电影、电视节目或听英文歌曲的频率？(How often did you watch English movies, TV shows, and listen to English songs?)

从不(rarely)□ 偶尔(occasionally)□ 有时(sometimes)□ 经常(often)□ 总是(always)□

9. 您曾经跟英语母语者交流的频率是？ (How often did you communicate with native English speakers?)

从不(rarely)□ 偶尔(occasionally)□ 有时(sometimes)□ 经常(often)□ 总是(always)□

10. 与您同学相比，您认为您的英语听说能力在哪个层次？(Compared to your peers, how would you rate your English listening and speaking competence?)

较低(relatively low)□ 中等(medium)□ 较高(relatively high)□

考试认知焦虑量表 （Cognitive Test Anxiety Scale）

本节的目的是测量您对广东高考英语听说考试的感受。请从1到5划圈选择您认为最符合您情况的选项。(The purpose of this part is to measure how you felt in the CELST. Please circle one of the numbers from 1 to 5 that best describes your situation.)

1=非常不同意(strongly disagree)；2=不同意(disagree)；3=不确定(neutral/not sure)；4=同意(agree)；5=非常同意(strongly agree)						
1.	相比其他学生，看懂英语听说考试指令，对我来说并不难(I had less difficulty than average students in getting the English test instruction straight)	1	2	3	4	5
2.	相比其他学生，我对高考英语听说考试不太紧张。(I was less nervous about the CELST than average students)	1	2	3	4	5
3.	高考英语听说考试中，我会不由自主地想不及格的后果。(During the CELST, I found myself thinking of the consequences of failing.)	1	2	3	4	5
4.	高考英语听说考试刚开始时，我会很紧张，以至于不能思考。(At the beginning of the CELST, I was so nervous that I often couldn't think straight.)	1	2	3	4	5
5.	相比其他学生，我在高考英语听说考试考场更加镇定。(I was calmer in the CELST test situations than average students.)	1	2	3	4	5
6.	回答高考英语听说考试中某一问题有压力时，我的脑子会一片空白。(My mind went blank when I was pressured for an answer on the CELST test.)	1	2	3	4	5
7.	高考英语听说考试中，我经常会想我也许不够聪明。(During the CELST, the thought frequently occurred to me that I may not be too bright.)	1	2	3	4	5
8.	高考英语听说考试中，我会很紧张，以至于忘了我所知道的内容。(During the CELST, I got so nervous that I forgot facts that I really knew.)	1	2	3	4	5
9.	我过度担心我是否能考好高考英语听说考试。(I worried about doing well on the CELST than I should.)	1	2	3	4	5

10.	在考高考英语听说考试之前，我感到自信和放松。(Before taking the CELST, I felt confident and relaxed.)	1	2	3	4	5
11.	在考高考英语听说考试之时，我感到自信和放松。(While taking the CELST, I felt confident and relaxed.)	1	2	3	4	5
12.	在高考英语听说考试中，我会感到我做得不太好。(During the CELST, I had the feeling that I was not doing well.)	1	2	3	4	5
13.	我担心高考英语听说考试很难，还没开考，我就有受挫感。(I worried that the CELST was too difficult and I felt defeated before I even started.)	1	2	3	4	5
14.	我不擅长英语听说考试。(I am not good at taking the CELST.)	1	2	3	4	5
15.	看到高考英语听说考试试题时，我要花一点时间才能镇定下来，开始思考。(When I first got my copy of the CELST, it took me a while to calm down to the point where I could begin to think straight.)	1	2	3	4	5
16.	我想在高考英语听说考试中取得好成绩，所以感到压力很大。(I felt under a lot of pressure to get good grades on the CELST.)	1	2	3	4	5
17.	我英语听说考试通常发挥不好。(I did not perform well on the CELST.)	1	2	3	4	5

外语听力焦虑量表（**Foreign Language Listening Anxiety Scale**）

此量表是为了测量您在广东高考英语听说考试以及平时听英语时的感受。(This scale is to measure how you felt when you listened to English in the CELST and when you listened to English in daily life)

1=非常不同意(strongly disagree)；2=不同意(disagree)；3=不确定(neutral/not sure)；4=同意(agree)；5=非常同意(strongly agree)					
认知维度(cognitive dimension)					
高考英语听说考试中，在听英语时，我常常担心自己会听漏信息。(When listening to English during the CELST, I often worried that I would miss information.)	1	2	3	4	5
高考英语听说考试中，在听英语时，我常常担心自己会误解听错。(When listening to English during the CELST, I often worried that I would misunderstand something.)	1	2	3	4	5
高考英语听说考试中，在听英语时，我常常担心自己的听力表现会比别人差。(When listening to English during the CELST, I often worried that my listening comprehension performance was worse than others.)	1	2	3	4	5
躯体维度(somatic dimension)					
高考英语听说考试中，在听英语时，我经常会感到心跳加速。(When listening to English during the CELST, I often felt my heart pounding.)	1	2	3	4	5
高考英语听说考试中，在听英语时，我经常紧张到发抖。(When listening to English during the CELST, I often got so nervous that I trembled.)	1	2	3	4	5
高考英语听说考试中，在听英语时，我经常会冒冷汗。(When listening to English during the CELST, I often sweated and perspired.)	1	2	3	4	5
行为维度(behavioral dimension)					
平时听英语或做听力练习时，我常常会听一下就放弃，不听了。(When listening to English, I often give it up easily.)	1	2	3	4	5

做英语听力练习时，我经常会敷衍了事。(I often skimp over English listening comprehension exercises.)	1	2	3	4	5
我通常会尽可能避免听英语。(I usually do my best to avoid listening to English.)	1	2	3	4	5

外语口语焦虑量表（**Foreign Language Speaking Anxiety Scale**）

此量表是为了测量您在高考英语听说考试以及平时说英语时的感受。(This scale is to measure how you felt when you spoke English in the CELST and when you spoke English in daily life)

1=非常不同意(strongly disagree)；2=不同意(disagree)；3=不确定(neutral/not sure)；4=同意(agree)；5=非常同意(strongly agree)					
认知维度(cognitive dimension)					
高考英语听说考试中，在说英语时，我常常担心自己会犯语言错误。(When speaking English during the CELST, I often worried that I would make language mistakes.)	1	2	3	4	5
高考英语听说考试中，在说英语时，我常常担心自己的口语表现比别人差。(When speaking English during the CELST, I often worried that my speaking performance was worse than others.)	1	2	3	4	5
高考英语听说考试中，一开始说英语，我就开始担心自己会无法表达自己的想法。(As soon as I started speaking English during the CELST, I begun to worry about not being able to express myself.)	1	2	3	4	5
躯体维度(somatic dimension)					
高考英语听说考试中，在说英语时，我经常会感到心跳加速。(When speaking in English during the CELST, I often felt my heart pounding.)	1	2	3	4	5
高考英语听说考试中，在说英语时，我经常紧张到发抖。(When speaking in English	1	2	3	4	5

during the CELST, I often got so nervous that I trembled.)					
高考英语听说考试中，在说英语时，我经常会冒冷汗。(When speaking in English during the CELST, I often sweated and perspired.)	1	2	3	4	5
行为维度(behavioral dimension)					
平时说英语或做口语练习时，我常常会讲一下就放弃，不讲了。(When practiceing speaking English, I often give it up easily.)	1	2	3	4	5
做英语口语练习时，我经常会敷衍了事。(I often skimp over English-speaking exercises.)	1	2	3	4	5
我通常会尽可能避免说英语。(I usually do my best to avoid speaking English.)	1	2	3	4	5

Appendix C Interview Outline

Note: The questions were asked in Chinese in the semi-structured interviews. The English translations of the interview questions are also provided for reference.

考生对高考英语听说考试的态度(test takers' attitudes towards the CELST)
1. 你认为这次高考听说考试难吗？哪一部分对你来说有难度？为什么？ (Do you think the CELST you took was difficult? Which part was the most difficult for you? Why?) 2. 在高考听说考试前，考试中，跟考试后，你是紧张不安还是自信镇定？ (Did you feel confident or nervous before, during and after the CELST?) 3. 你是否经常担心自己高考听说考试考不好？考试当中是否出现跟考试试题无关的想法？(Did you often worry that you fail to perform well on the CELST? Did task-irrelevant thoughts come to your mind during the CELST?) 4. 在高考听说考试中，你是否经常担心听错或者听漏信息，是否担心听力表现比别人差？听英语时是否会有什么身体反应？平常是否会尽量避免听英语？ (While listening to English during the CELST, did you often worry that you would miss or misunderstand information or perform worse than others? Did you have any physiological arousal while listening to English during the CELST? Did you avoid listening to English in your daily life?) 5. 在高考听说考试中，你是否经常担心自己犯语言错误或无法表达自己，是否担心口语表现比别人差？考试中说英语是是否会有什么身体反应？平常是否会尽量避免说英语？(While speaking English during the CELST, did you often worry that you would make language mistakes or unable to express yourself or perform worse than others? Did you have any physiological arousal while speaking English during the CELST? Did you avoid speaking English in your daily life?)

焦虑对高考英语听说考试表现的影响(the effect of anxiety on the CELST performance)
6. 如果你在英语听说考试中感到焦虑，你认为这种焦虑感对你来说是有益还是有害的，还是没有影响？为什么？ (If you did experience anxiety during the CELST, do you think the anxiety has a beneficial or harmful effect, or no effect on your CELST performance? Why?)
导致考生在高考英语听说考试中焦虑的因素(reasons leading to anxiety in the CELST)
7. 你认为高考英语听说考试中有什么因素会导致焦虑感？ (What do you think would lead to anxiety in the CELST?)
缓解焦虑的策略以及对高考英语听说考试的建议(strategies to alleviate anxiety and suggestions for the CELST)
8. 如果你在听说考试前或考试中感到焦虑，你会采取什么策略来缓解这种焦虑？ (If you experienced anxiety before or during the CELST, what strategies would you take to alleviate the anxiety?) 9. 你有什么建议可以使高考听说考试变得不那么令人焦虑，从而发挥你的最佳水平？ (What suggestions do you have to make the CELST less anxiety-provoking so that you can have your best performance?)

Appendix D Ethic Clearance of the Study

Queen's University

May 25, 2021

Jinhong Chen
Master's Student
Faculty of Education
Queen's University
Duncan McArthur Hall
511 Union Street West
Kingston, ON, K7M 5R7

Title: "GEDUC-1058-21 Exploring the Relationship Between Chinese Test Takers' Anxiety and Their Performance on a High-stakes Computer-based English Listening and Speaking Test;" TRAQ # 6032736

Dear Jinhong Chen:

The General Research Ethics Board (GREB), by means of a delegated board review, has cleared your proposal entitled **"GEDUC-1058-21 Exploring the Relationship Between Chinese Test Takers' Anxiety and Their Performance on a High-stakes Computer-based English Listening and Speaking Test"** for ethical compliance with the Tri-Council Guidelines (TCPS 2) and Queen's ethics policies. In accordance with the Tri-Council Guidelines (Article 6.14) and Standard Operating Procedures (405), your project has been cleared for one year.

You are reminded of your obligation to submit an annual renewal form prior to the annual renewal due date (access this form at http://www.queensu.ca/traq/signon.html/; click on "Events;" under "Create New Event" click on "General Research Ethics Board Annual Renewal/Closure Form for Cleared Studies"). Please note that when your research project is completed, you need to submit an Annual Renewal/Closure Form in Romeo/traq indicating that the project is 'completed' so that the file can be closed. This should be submitted at the time of completion; there is no need to wait until the annual renewal due date.

You are reminded of your obligation to advise the GREB of any adverse event(s) that occur during this one-year period (access this form at http://www.queensu.ca/traq/signon.html/; click on "Events;" under "Create New Event" click on "General Research Ethics Board Adverse Event Form"). An adverse event includes, but is not limited to, a complaint, a change or unexpected event that alters the level of risk for the researcher or participants or situation that requires a substantial change in approach to a participant(s). You are also advised that all adverse events must be reported to the GREB within 48 hours.

You are also reminded that all changes that might affect human participants must be cleared by the GREB. For example, you must report changes to the level of risk, applicant characteristics, and implementation of new procedures. To submit an amendment form, access the application by at http://www.queensu.ca/traq/signon.html; click on "Events;" under "Create New Event" click on "General Research Ethics Board Request for the Amendment of Approved Studies." Once submitted, these changes will automatically be sent to the Ethics Coordinator, GREB, at University Research Services for further review and clearance by GREB or the Chair, GREB.

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

Sincerely,

Chair, General Research Ethics Board (GREB)
Professor Dean A. Tripp, PhD
Departments of Psychology, Anesthesiology & Urology Queen's University

c: Dr. Saad Chahine, Supervisor
Dr. Saad Chahine, Chair, Unit REB
Kyle Cummings-Bentley, Dept. Admin.