EXPLORING THE WASHBACK EFFECTS OF THE CHINESE NATIONAL MATRICULATION ENGLISH TEST (NMET) ON STUDENT LEARNING AND THEIR TEST ANXIETY

by

Xing Zheng

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

Queen’s University
Kingston, Ontario, Canada
August, 2019

Copyright ©Xing Zheng, 2019
Abstract

Washback refers to the effects of language testing on teaching and learning (Alderson & Wall, 1993). Washback studies have been frequently conducted in educational fields in the past several decades (Alderson & Wall, 1993; Cheng, 1997; 2005; Green, 2006; 2007; Qi, 2007). And those studies focus more on teaching and less on learning, so do the washback studies in the Chinese context (Gu, 2007; Gu & Peng, 2010; Qiao, 2006; Qin, 2012). The National Matriculation English Test (NMET) is the most populated large-scale high-stakes test in China with approximately 10 million test takers involved annually for admission to higher education. Thus, it offers a meaningful context to conduct research on washback as tests that have important consequences will have more intensive washback. Also, such important consequences as not being admitted to universities could lead to a potential source of heightened test anxiety, which was less researched. This study investigated the effects of the NMET in Hangzhou, Zhejiang province [hereafter referred to NMET (ZJ)], China on secondary school students’ learning and their test anxiety.

A quantitative data analysis (descriptive statistics, factor analysis and regression analysis) was conducted on 406 NMET (ZJ) test takers in Hangzhou in 2017. Results from the study are as follows: (1) Negative washback effects were observed on what and how students learned about English.  (2) There existed a negative effect on students’ test anxiety. (3) Both a social context aspect of test anxiety (peer-related test anxiety) and a cognitive aspect of test anxiety (worry) were observed. Additionally, one of the two test anxiety factors (test unconfident) negatively predicted students’ English level. As students’ English level decreased, their confidence on the test also decreased.

The results indicate that the NMET reform in Zhejiang did not achieve its intended purpose of decreasing students’ test anxiety and promote their English learning.
Both teachers and students should be aware of the negative effects of the test. Learning focus should be on genuine English learning instead of just learning for the test. Additionally, results on how changes of the test format negatively affect students also provide empirical evidence to test designers for future test development.
Acknowledgments

I would like to first express my sincere thanks to my supervisors, Dr. Liying, Cheng, Faculty of Education, Queen’s University, Kingston, Canada and Dr. Liyan, Huang, School of Foreign Studies, South China Normal University (SCNU), Guangzhou, China. During the past three years, as a lucky dual-degree student between those two prestigious institutions, I have received tremendous guidance, academic help and inspiring encouragements from those two supervisors, without either of whom, this thesis would never build itself up to what it is now. Not only did they two teach me how to do research, they also set a vivid example for me in how to become a terrific teacher by their guiding me during the process. I owe them too much.

I would like to extend my special gratitude to my committee member Dr. Kristy, Timmons at Queen’s for her insightful suggestions when I was categorizing the questionnaire used in the thesis as well as her supports in organizing my work.

My sincere thanks to all the teaching staff in two departments in both Canada and China can also not be ignored. Thanks to Dr. Lynda, Colgan at Queen’s, for her patience in helping me initiate my sense of writing a Master thesis especially; for Dr. Manfei, Xu at SCNU, for her critical comments on my research questions in the colloquium.

Special thanks go to South China Research Center for Elementary English Education and Assessment at SCNU which supports my data in the study. It has been such a great honor for me to be a member of it during the past three years.

Also, I feel more than just grateful to my family for their selfish love and support during my master process. They love me so much and I feel I owe them too much as well. I also feel thankful to my fiancé for his love and support especially during my periods in Canada.
Table of Contents

Abstract .................................................................................................................................................. I

Acknowledgments .......................................................................................................................... III

Table of Contents ........................................................................................................................ IV

Lists of Tables ................................................................................................................................ VII

Lists of Figures ................................................................................................................................ VIII

Lists of Abbreviations .................................................................................................................. IX

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

The Research Context .................................................................................................................. 2

National College Entrance Exam (NCEE) ................................................................................ 2

National Matriculation English Test (NMET) ........................................................................ 2

Rationale ........................................................................................................................................ 4

Washback and Consequences .................................................................................................... 4

Test Anxiety ................................................................................................................................... 7

Purpose and Research Questions ............................................................................................. 8

Overall Structure of the Thesis ............................................................................................... 8

Chapter 2 Literature Review ...................................................................................................... 10

Washback ...................................................................................................................................... 10

The Definition of Washback ...................................................................................................... 10

Alderson and Wall's (1993) 15 Washback Hypotheses .......................................................... 12

Empirical Studies of Washback in Various Contexts .............................................................. 15

Empirical Studies of Washback in the NMET Context ........................................................... 18

Test Anxiety ................................................................................................................................. 19
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Methodology</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Research Design</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Participants</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Instruments</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>Data Collection</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Data Analysis</td>
<td>31</td>
</tr>
<tr>
<td>4</td>
<td>Results</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Descriptive Statistics</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Factor Analysis</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>Regression Analysis</td>
<td>46</td>
</tr>
<tr>
<td>5</td>
<td>Discussions and Conclusions</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>The Washback Effects of the NMET (ZJ) on Student Learning</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Washback Effects on What to Learn</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Washback Effects on How to Learn</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Washback Effects on Rate and Sequence of Learning</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>Washback Effects on Attitudes towards Content, Method, etc. of Learning</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>The Washback Effects of the NMET (ZJ) on Test Anxiety</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>Relationship between test anxiety and self-rated English proficiency</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Revisit the 15 Washback Hypotheses</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Implications and Limitations</td>
<td>57</td>
</tr>
</tbody>
</table>
References.................................................................................................................................................61

Appendix A A Student Questionnaire (In Chinese)....................................................................................83

Appendix B A Student Questionnaire (In English)......................................................................................86

Appendix C Ethics Approval Letter...........................................................................................................90

Appendix D Outline for Interview (In Chinese).........................................................................................91

Appendix E Outline for Interview (In English).........................................................................................92
Lists of Tables

Table 1.1 Test Format of the NMET (ZJ) in 2017.................................................................04
Table 4.1 Frequency Distribution for the Highest Response on “H5 what to learn”
Question Items......................................................................................................................36
Table 4.2 Frequency Distribution for the Highest Response on “H6 how to learn”
Question items.......................................................................................................................37
Table 4.3 Frequency Distribution for the Highest Response on “H8 rate and sequence of
learning” Question Items......................................................................................................38
Table 4.4 Frequency distribution for the Highest Response on “H11 attitudes towards
content, method, etc. of learning” Question Items.................................................................40
Table 4.5 Reliability Analysis for Test Anxiety Question Items..............................................41
Table 4.6 Descriptive Statistics for Test Anxiety Question Item 35 and Item 39..............42
Table 4.7 Factor Loadings of 14 Test Anxiety Question Items................................................45
Table 4.8 Correlation between Dependent Variables and Independent Variables............46
Table 4.9 Multiple Regression Analysis of Self-rated English Proficiency......................47
Lists of Figures

Figure 4.1 English Tutor for Students With Different English Levels..................37
Figure 4.2 Private Tutor Time Students with Different English Levels Spent..........39
Figure 4.3 Peer-related Test Anxiety of Students with Different English Levels......43
Figure 4.4 Worry of Students with Different English Levels.............................43
Figure 4.5 Scree Plot.........................................................................................44
### Lists of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMET</td>
<td>National Matriculation English Test</td>
</tr>
<tr>
<td>NMET (ZJ)</td>
<td>National Matriculation English Test in Zhejiang province</td>
</tr>
<tr>
<td>SCNU</td>
<td>South China Normal University</td>
</tr>
<tr>
<td>EFL</td>
<td>English as a Foreign Language</td>
</tr>
<tr>
<td>ESL</td>
<td>English as a Second Language</td>
</tr>
<tr>
<td>NCEE</td>
<td>National College Entrance Exam</td>
</tr>
<tr>
<td>MOE</td>
<td>Ministry of Education</td>
</tr>
<tr>
<td>CET</td>
<td>College English Test</td>
</tr>
<tr>
<td>SPSS</td>
<td>Statistical Package for Social Science</td>
</tr>
</tbody>
</table>
Chapter 1 – Introduction

The test researched in this study is the NMET (ZJ), which is a part of the National College Entrance Examination (NCEE) for the English subject. On the national scale, the learning of English as a Foreign Language (EFL) in China has become a nationwide endeavor pursued at all academic levels. As education in China is selection biased, each educational level selects the best-qualified graduates from the former level. Students’ performance in English testing is one of the most used criteria (Cheng, 2008; Cheng & Qi, 2006; Qi, 2007). Consequently, secondary schools are concerned with these test scores as higher admission rates into universities attract more promising students for a new academic term annually (Chen, 2010). One of the most valued examinations in China is the National College Entrance Exam (NCEE), known as Gaokao, which is developed to select secondary candidates for higher education and is considered to be the crucial step for upward social mobility (He, 1998; Yu & Suen, 2005). A more detailed introduction of the two tests is laid out in the following section of this chapter. The phenomenon that I was exploring in this thesis is generally known as washback. It was defined as the influence of language testing on teaching and learning (Alderson & Wall, 1993).

This thesis reported a survey study and explored how the NMET (ZJ), one of the most large-scale high-stakes tests in the Chinese context influences its test takers’ learning and test anxiety. This study aimed to provide insights into how this newly reformed test worked within the school system. It also intended to give suggestions on how the stakeholders including students and teachers should react appropriately to those changes. Therefore, three research questions were addressed (see this Chapter below).
The Research Context

National College Entrance Exam (NCEE)

NCEE is a highly centralized examination system in China. It is a large-scale high-stakes test with approximately 10 million secondary students taking the test across the country every year. Secondary students in their final year take the NCEE on June 6, 7 and 8 each year before they apply for a university in July. Test scores in the NCEE determine university admission. Subjects tested on the NCEE have been modified several times since it was resumed in 1978 (Chen, 1997, He, 2000). The latest NCEE examination system is called a “3+1” format which was first introduced in Guangdong province in 1999 and was subsequently implemented into a national scale within three years (Ministry of Education, 2001). This current system requires all test takers to take the examinations on three common subjects: Chinese language, mathematics, and a foreign language.

Additionally, students are also required to choose another comprehensive test in either arts or sciences depending on their personal choices. “Comprehensive arts” [文科综合] includes history, civics/political sciences, and geography while “Comprehensive sciences” [理科综合] includes chemistry, physics, and biology. Although the foreign language test in the NCEE provides a long list of choices (e.g., English, Japanese, French, and others), English remains a major and the only choice for over 95% of test takers as English education dominates China’s foreign language education market, so the majority of all educational schools provides English courses to their students. And the English test of the NCEE is referred to as the National Matriculation English Test (NMET).

National Matriculation English Test (NMET)

The NMET is a standardized norm-referenced proficiency test which is developed by the National Education Examinations Authority directly under the Chinese Ministry of
The NMET was introduced in 1985 as a replacement for the old national university entrance English test which was criticized for its low validity and reliability (Li, 1990, also cited by Qi, 2007). Currently, the test format of the NMET in every province or municipality is uniformly decided by the MOE. Only five provinces or cities (Beijing, Shanghai, Tianjin, Jiangsu province and Zhejiang province) have free propositional rights in designing the test format of the NMET in their areas. Like the NCEE, the NMET has also been through several reforms as well.

Recently, after its long notorious once-in-a-lifetime examination reputation, the MOE announced in late 2013 that the NMET would adopt a “two tests a year” policy (Yang, 2014). The new NMET reform was first piloted in Zhejiang province and Shanghai municipality by the MOE in 2016. It will be subsequently implemented in the whole country by 2020 (State Council, 2014). It requires that, including the test in June discussed above, senior 3 students can choose to take their first NMET in October and take their second in the next June. Additionally, only the higher score of these two tests they obtain will remain for university application in July. Having been modified several times, the specific test format of the latest NMET in Zhejiang province is in Table 1.1.

Like its old version, the NMET (ZJ) does not measure speaking abilities. Compared with the old NMET, there are two changes worth noting in this 2017 version: 1) the decreasing weighing of multiple choice questions and 2) the adding of a new writing test task called “continuation task.” Results showed that this new writing “task” was a challenge for both teachers and students: test takers were required to write two compositions instead of one as usual within a limited time (Hou, 2017).
Table 1.1

Test Format of the NMET (ZJ) in 2017

<table>
<thead>
<tr>
<th>Content</th>
<th>Item Type</th>
<th>Item No.</th>
<th>Marks</th>
<th>Weighing (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Listening</td>
<td>Multiple Choice</td>
<td>20</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>II. Reading</td>
<td>Multiple Choice</td>
<td>12</td>
<td>35</td>
<td>23</td>
</tr>
<tr>
<td>III. Cloze</td>
<td>Multiple Choice</td>
<td>20</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>IV. Grammar/Vocab</td>
<td>Gap Filling</td>
<td>10</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>V. Writing</td>
<td>Guiding Writing</td>
<td>1</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Continuation Task</td>
<td>1</td>
<td>25</td>
<td>17</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>62+2</td>
<td>150</td>
<td>100</td>
</tr>
</tbody>
</table>

Research on the validity of this NMET (ZJ) revealed that tested items were more reasonable than the older version as there was an increasing weighing of non-multiple choice questions as well as reinforcement of writing abilities and thinking methods (Chen & Luo, 2016; Lu, 2017; Wang & Yang, 2017). However, Hu (2015) expressed concern that this new reform in Zhejiang brought a negative washback to English teaching and learning in secondary schools when English instruction tended to be more test-driven as learners wanted to obtain a higher score in their second test.

Rationale

Washback and Consequences

Washback is the influence of language testing on language teaching and learning, which has been discussed extensively in the general education and the language education literature (Alderson & Wall, 1993; Chapman & Snyder, 2000; Frederiksen & Collins, 2016; Heyneman, 1987; Madaus, 1988; Morris, 1985; Swain, 1984; Wall, 1997, 2000; Wall & Alderson, 1993). Washback is a term that can also be found frequently in
many official documents and research in China (Chen, 2008; Gu & Xiao, 2013; Huang & John, 2013; Huang & Yang, 2002; Li, 1990; Ye, 1998). According to Alderson and Wall (1993), there were 15 areas of teaching and learning that can be influenced by testing. Over the past 30 years, many of these areas of washback have been empirically studied (see detailed review by Cheng, Sun, & Ma, 2015). Like empirical studies in other contexts, washback studies of the NMET in China have also focused more on teaching and teachers instead of learning and learners. Results in China reported both teaching content and teaching materials were more “testlike” (Cao, 2010; Dai, 2016; Dong, 2014; 2017; Hu, 2014; Song, 2013) and the reasons behind such unintended washback effects were mixed (Qi, 2005). Limited as it is, research on learning or learners suggested that the NMET brought about a negative impact on test takers’ English learning sequence and strategies (Hu, 2015); as well as test anxiety (Chen, 2008). Despite an increasing number of empirical washback studies in China, only 3.7% of the empirical washback studies focused on the NMET while the majority of studies (66.7%) focused on the College English Test (CET) in China over the past two decades (Zhou & Dong, 2014). There were few washback studies on learning and learners; and even fewer on the NMET context. Thus, the current study focused on the washback effects on student learning and test anxiety.

The function and importance of the NCEE system in modern China cannot be fully understood without a historical context (Hafsteinsson, 2016). Its historical counterpart, Keju  [科举] was similarly of great importance for the education system of the imperial China. Keju, this old civil service examination system of China had been used as a governmental promotion of education as well as a way to upward social mobility for centuries from the year 606 to 1905 (Suen & Yu, 2006). Both ancient and contemporary education in China has been created, promoted, and maintained by
state-orchestrated systems of high-stakes and extremely competitive examinations (Cheng, 2008; Feng, 1994; Guo, 1994; Yu & Suen, 2005). The modern examination system of China developed directly from the old imperial examination system (Kirkpatrick & Zang, 2014). The imperial idea of “万般皆下品，唯有读书高” (All pursuits are of low value; Only studying the books is of high value) is still the dominant view of education in China (He, 2000). The NCEE, as the contemporary Keju, still serves as a crucial step for upward social mobility today (Cheng, 2008; He, 1998; Yu & Suen, 2005).

The NCEE was initially established in 1952 and was resumed in 1977 after the 1966 Cultural Revolution [文化大革命] ended. After having been suspended for a decade, when the NCEE was first held in 1977, 5.7 million test takers attended the NCEE, but only 300,000 went to universities (Hsu & Wu, 2015). Therefore, the NCEE, from the beginning, has been fiercely competitive. Higher education in China has not been accessible to everyone due to space limitation, and only 20% to 40% of students can obtain the opportunity to enter universities each year (Liu & Dunne, 2009). Failure in any of the NCEE tested subjects can result in losing entrance to a dream university which interferes with obtaining satisfactory jobs. This highly competitive environment thus has brought about significant consequences and influences to millions of test takers’ life.

Therefore, it is not surprising that even though the NMET has been implemented for a relatively short period of time, the idea of obtaining a higher test score during the NMET has been rooted in millions of test takers’ minds. For example, a survey involving over 100,000 parents revealed that during the year when students take the NCEE, 43% of the parents would spend 20 thousand yuan (equally approximately 20% of their annual family incomes) on extracurricular expenditure to support studying of the NCEE tested subjects (www.sina.com, 2017). Among them, English ranked the third of all the nine
tested subjects, firmly behind physics and mathematics, with 52.4% parents revealing that they would spend extracurricular money on English for a higher score in the NCEE.

**Test Anxiety**

Test anxiety is an unpleasant state characterized by feelings of tension and apprehension, worrisome thoughts and the activation of the autonomic nervous system when an individual is faced with evaluative achievement-demanding situations (Speilberger, 1972; also cited by Ergene, 2003). Test anxiety has been researched for decades. Currently, the increasing prevalence and extensive use of large-scale, high-stakes testing to measure students’ achievement and schools’ effectiveness have made test anxiety one of the most confronting issues in modern times (Cheng et al., 2014; Datta, 2013; Embse et al., 2013).

The NCEE in China is the most critical examination (Feng, 1995) as test performance in the NCEE is crucial for seeking admission to universities. Consequently, more pressure and test anxiety have been put on students particularly for higher scores in the NCEE. Empirical investigations into test anxiety in the NCEE revealed that test anxiety in the NCEE became a serious social problem (Zhang et al., 2012; Zhou, 1990). Learning was considered by Chinese students as a ‘joyless’ process due to ‘endless,’ ‘stressful’ and ‘demanding’ academic pursuit. Up to 66.67% of the NCEE test takers reported having test anxiety (Sun & Chen, 2012) and female test takers were more anxious than their male counterparts (Huang et al., 2003).

Despite its prevalence, test anxiety on the NMET is a less researched area. Wang (2007) reported that 45.9% of 799 secondary school students were test-anxious about a mid-term English test and 19.3% of them were in a high test-anxious level. The most significant contributing factor to their English test anxiety was students’ fear of negative evaluation from others (e.g., peers, teachers) due to failure in the English test. A recent
study showed that the evaluative and selective uses of the NMET scores for high-stakes decisions in China brought about a negative washback effect on test anxiety in the NMET of Shandong province (Chen, 2008). However, how the latest NMET reform in such an important national English test might affect students’ learning and test anxiety is unknown. Further, there are scare washback studies focusing on testing consequences with test anxiety, which is what the current study has focused on.

**Purpose and Research Questions**

The purpose of the current study was to examine the washback effects of the NMET (ZJ) on student learning and their test anxiety. This study aimed to fill the research gap in examining the relationship between washback effects on student learning and their test anxiety. The following research questions were addressed:

1. What was the washback effect of the NMET (ZJ) on student learning?
2. To what extent might the NMET (ZJ) bring a washback effect on test takers’ test anxiety?
3. How did student factors (such as students’ school backgrounds and self-rated English proficiency) affect their test anxiety?

**Overall Structure of the Thesis**

This thesis consists of five chapters as follows:

Chapter One is the introduction part of this thesis. It contains the research context of this study, my rationale of conducting this study, and purpose and research questions of this study as well as overall of the thesis.

Chapter Two is the literature review section of this thesis. This Chapter reviews literature about washback and test anxiety respectively. Both relevantly theoretical and empirical studies are included.

Chapter Three describes the methodology of this thesis. It includes a description
of the participants and instruments, and a view of the data collection and data analysis.

Chapter Four reports the results of the thesis. Reports of the study results are matched with the research questions.

Chapter Five draws a conclusion of the thesis in relation to the literature. Discussions of the study results are followed. Additionally, limitations and implications of the thesis are described as well.
Chapter 2 - Literature Review

This chapter presents the theoretical and empirical underpinnings that guide my study. It consists of two interrelated parts. The first part defines washback; the conceptual framework of washback and the empirical evidence helps justify the existence of washback in various contexts including the NMET. The following part reviews literature about test anxiety. It includes a definition of test anxiety and review of the empirical studies focusing on test anxiety in various context including the NMET. The primary purposes of this chapter are to identify both important washback and test anxiety research findings relevant to my study; the linking areas between washback and test anxiety research where further studies are needed and present how my study helps devote to these areas.

Washback

The Definition of Washback

The phenomenon of washback has received increasing attention in various educational contexts over the past three decades. The definition that language testing influence language teaching and learning has various labels: backwash (Hughes, 1993); washback (Alderson & Wall, 1993) and impact (Wall, 1997); test impact (Bachman & Palmer, 1996; Baker, 1991); as well as consequential validity (Messick, 1989). In the simple definition, “backwash” or “washback” refers to the influence of testing on teaching and learning (e.g., Alderson & Wall, 1993; Bailey, 1996; Cheng & Curtis, 2004; Hughes, 1989; Saville, 2000). Alderson (2004) believes that there is no difference between these two terms.

Below are definitions cited chronologically:
Hughes (1989:1) uses the term “backwash” for test influence to describe “the effect of testing on teaching and learning” and it can be beneficial or harmful. Later (1994), he further emphasizes that “backwash” includes a test’s effect on any aspect of teaching and learning.

Alderson and Wall (1993, p. 117) describe washback as a “complex phenomenon” and refer it to the extent to which language testing influence how “teachers and learners do things they would not necessarily otherwise do because of the test.” Following their definitions, Messick (1996: 241) further expands that “washback refers to the extent to which the introduction and the use of a test influence language teachers and learners to do things they would not necessarily otherwise do that promote or inhibit language learning”.

Bailey (1996, p. 259) summarizes washback as “the influence of testing on teaching and learning,” and that it is widely held to exist and to be important. As there are little empirical studies to document its exact nature, she further comments that “there is also concern what constitutes both positive and negative washback, as well as how to promote the former and inhibit the latter” (p. 259).

Bachman and Palmer (1996) place washback within the scope of test impact. They have noted that washback can be viewed as various aspects of test impact oriented at a macro and micro level. Wall (1997) refers explicitly to impact as “any of the effects that a test may have on individuals, policies or practices, within the classroom, the school, the educational system or society as a whole.” She explains that both impact and washback are sometimes used interchangeably but the latter is more frequently used to refer to test effects on teaching and learning in the classroom.

In summary, tests drive the curriculum, teaching methods and students’ approaches to learning (Biggs, 1995); tests can either reinforce certain attitudes or
behaviors, make students study hard, or even bring about pressures which could lead to abnormal behaviors (Tsagari, 2007). The definitions of washback can potentially be defined at various levels, from the classroom to the whole social and educational systems. For the study, the term “washback” refers to the influence of testing on teaching and learning which can be harmful or beneficial as defined by Wall (2000).

**Alderson and Wall’s (1993) 15 Washback Hypotheses**

The origin of washback study can be traced to a significant landmark when Alderson and Wall (1993) published a seminal paper entitled “Does washback exist?” This excellent article provided a conclusive summary of less researched issues in language education including the concept of washback and the research methods associated with washback as well. This original piece called for empirical exploration into washback phenomena which also inspired many subsequent studies. The most salient part of the article was the presentation of 15 washback hypotheses, which were continuously cited by almost all subsequent researchers in the field:

| 1. A test will influence teaching. |
| 2. A test will influence learning. |
| 3. A test will influence what teachers teach. |
| 4. A test will influence how teachers teach. |
| 5. A test will influence what learners learn. |
| 6. A test will influence how learners learn. |
| 7. A test will influence the rate and sequence of teaching. |
| 8. A test will influence the rate and sequence of learning. |
| 9. A test will influence the degree and depth of teaching. |
| 10. A test will influence the degree and depth of learning. |
| 11. A test will influence attitude towards the content, method, etc., of teaching and learning. |
| 12. Tests that have important consequences will have washback; conversely, |
| 13. Tests that do not have important consequences will have no washback. |
| 14. Tests will have washback on all learners and teachers. |
| 15. Tests will have washback effects for some learners and some teachers, but not for others. |

By proposing the 15 Washback Hypotheses, Alderson and Wall focused on the potential areas of washback research to be conducted, i.e., to lay out the territory where
studies of washback might address in the future. Eight of those 15 hypotheses were about learners or learning, which suggested predictions regarding content (what), methods (how), rate, sequence, degree, and depth of learning as potential process variable for empirical investigation (see hypothesis 2,5,6,8,10,11,14,15) (Green, 2013). As it is easier to observe what and how students learn in comparison to the rate, sequence, degree or depth of their learning, subsequent research should be precise in deciding which aspects of student learning should be categorized into each heading.

Despite the fact that those 15 hypotheses are widely acknowledged by the field, both Alderson and Wall hold an open mind for potential expansion to the original 15 hypotheses. In 1996, they added another hypothesis “tests will have different amounts and types of washback on some teachers and learners than on other teachers and learners” and called for empirical justification. What Alderson and Wall intend to say is that washback may vary as context and other conditions such as stakeholders change. Therefore, it is imperative for future research to embrace any new possibility.

Following Alderson and Wall, Hughes (1993) was the first to provide an analytical framework of washback (Hughes calling it backwash) by drawing a useful distinction between effects on participants; process and products. Hughes identified three aspects of backwash:

1. participants: students, classroom teachers, administrators, material developers and publishers, whose perceptions and attitudes towards their work may be affected by a test
2. process: any actions taken by the participants which may contribute to the process of learning
3. product: what is learned and the quality of learning

This distinction offered a theoretical basis for deciding which aspect of washback should be the main territory for empirical investigation. Based on the work of Hughes (1994), Bailey (1996) further expanded the scope of both participants and products. She
categorized the washback effects of a test into two dimensions: first, “washback to the learners”, which was the result of providing “test-derived information” to the test takers and second, “washback to the programme”, which was the result of providing information to other participants in the process.

From Alderson and Wall (1993), to Hughes (1994) and then even to Bailey (1996), the washback models in the 1990s extended to all stake-holders, and society at large (Hamp-Lyons, 1997). The main context in those washback models remains within the school and classroom. However, washback effects on cross-school settings have not been given enough attention, which is to be a focus of this current study. How washback across schools (e.g., public and private; urban and regional) functions still calls for a more dynamic and comprehensive model to explain (Savillie, 2009). For the current study, how washback across the top and non-top secondary schools (including both average and low-level schools) operates is of particular interest. Another shortcoming of the previous washback models is the lack of a possibility that a language test might influence students’ test anxiety. Many researchers have suggested that high-stakes language testing bring students into an anxious situation (Ferman, 2004; Shohamy, 1993; Shohamy et al., 1996). However, students as a key participant and test anxiety as an unavoidable feeling during the process still call for more future research in various contexts to merit (Tsagari, 2007).

Washback models by Alderson and Wall (1993) and Bailey (1996) were the most frequently cited ones in the field; most of the washback studies in the 20th century followed those traditions. Building on the foundational work of Alderson, Wall and Bailey, Green (2007) provided a predictive model that provides new insights into those traditions by taking account of inter alia participant variables (e.g., previous experience of learning English) and process variables (e.g., course length, use of English outside
class) at the same time. Green (2007) located washback within a model of second language learning based on Skehan (1989), incorporating measures of test anxiety (Horwitz, Horwitzs & Cope, 1986) in that study (Green, 2013). His findings did not lay much territory into the washback on test anxiety, but it did point out the insight that washback was shaped by participants and that a fuller understanding of participants’ engagement in test preparation should be valued.

Based on both Hugh (1993) and Bailey (1996)’s work on washback, Green (2007) expanded on their model by outlining the relationships between different washback determinants among which Green pointed out that individual differences of participants could result in that participants in the same test context may be affected in a different way. Further, Green (2013) summarized washback research since Hugh’s publication of *Testing for Language Teachers* (1989). Green expressly referred to washback as the impact during the test preparation process. But just as Green revealed, the roles of learners who were perhaps the most important participants of all, were still less understood in that process.

**Empirical Studies of Washback in Various Contexts**

Theoretical discussions after the 1990s trigger a series of empirical work on the washback effects of different types of tests in various contexts. By the end of the third millennium, washback had become a more-widely researched topic in language education, more specifically in EFL and ESL contexts (Alderson & Hamp-Lyons, 1996; Andrews, 2014; Cheng, 1997; Ferman, 2004; Li, 1990; Qi, 2005, 2007; Shohamy et al., 1996; Tayeb et al., 2014; Wall, 1996; Wall & Alderson, 1993; Wall & Horak, 2006; Wang, 2010; Watanabe, 1996). Empirical studies after Alderson and Wall’s 1993 seminar paper mostly focus on the teaching or teachers’ perspectives (Alderson & Hamp- Lyons, 1996; Burrows, 2004; Glover, 2006; Noble & Smith, 1994a; Shohamy et al, 1996; Wall &
Alderson, 1993; Wall & Horak, 2006). For example, Wall and Horak (2006) provided information on how teachers reacted when a major change was introduced into the existing testing system. Cheng (1997) added that teachers quickly and effectively changed the teaching materials and reluctantly changed the teaching methods they employed when a revised English examination syllabus was first adopted in Hong Kong. Both revisions and reforms into the educational system were intended for positive effects on teaching and learning. Qi (2005) revealed the reasons behind the failure of those intended effects by pointing out that the selecting function and the change-promoting function of a high-stakes test are in many ways in conflict with each other. Under the pressure of the high-stakes consequences of a test, both the curriculum and the instruction for learners are somehow manipulated by the test rather than serving the students’ specific needs (Menken, 2006; Scott; 2005). But there were also studies claiming teachers changed their teaching methods directly or indirectly as a result of the test (Burrows, 2004; Lam, 1993; Shohamy, 1993; Stecher et al., 2004). Reasons behind those mixed results were involved, and studies reviewed above demonstrate future studies need to be clear about what kinds of classroom behaviors should be relevant to how teachers teach before they concluded there was a washback on that (Tsagari, 2007).

Comparatively, research in the 1990s into the washback of tests on learning or learners remained limited. It was in the 21th century that research started to merit the washback on learning and learners in large numbers (Andrews, Fullilove & Wong, 2002; Ferman, 2004; Gosa, 2004; Green, 2006; Lumley & Stoneman, 2002; Stoneman, 2006). Some findings showed that students’ learning performance after the test preparation process did not witness many improvements (Green, 2007; Perrone, 2011; Xie, 2010). Despite the findings, some research findings on washback effects on student learning were disparate and too mixed to draw a definite conclusion (Andrews et al., 2002; Rao et
al., 2003; Read & Hayes, 2003; Safi, 2006) as it was not very clear by defining student learning when some researchers look at both student performance and achievement, and some researchers look at motivation, learning strategies, and other learning related variables. Therefore, it is essential to decide which aspects of student learning among washback hypotheses should be studied.

However, washback is a complex system as there is a range of factors mediating the influence of a test on learners (Alderson & Wall, 1993; Messick, 1996). Other than the test preparation program, Rao et al. (2003) also highlighted the influence of several factors (e.g., anxiety and motivation). Murray, Raizi, and Cross (2012) considered the role of learner attitudes in terms of opinions, beliefs and emotions in shaping washback. However, even if previous studies were acknowledging the importance of learners’ views, the number was still relatively limited, and the number of washback studies on test anxiety was even smaller.

Smith (1991a) reported that young children suffered in the pressurized environment of high-stakes testing with children being reported to experience frustration and low self-esteem. Ferman (2004) shared a similar result with Smith (1991a), 77.5% of participants (N= 93) in his study reporting that the test aroused in them feelings of anxiety to quite a great extent. Paris et al. (1991) further added that it was not just low-achieving students experiencing test anxiety, but students of all achievement levels suffering from different levels of worry and stress. Shohamy et al. (1996) revealed that the level of anxiety was even higher in a high-stakes testing context.

All those findings indicated that test anxiety in high-stakes tests has been a confronting issue for all students. Despite the findings, it remains unclear whether those levels of anxiety are brought about by the test, or by the usual side-effects of any high-stakes assessments. In other words, the existence of washback of tests on test
anxiety requires further empirical evidence. Therefore, studies of test anxiety and its facilitating and debilitating effects on student learning during the test-preparing process would merit future washback research (Tsagari, 2007).

Although they are two separate and independent factors, student learning and test anxiety interact tightly in the learning process and should be given a combined view in their role of shaping washback on learners.

Empirical Studies of Washback in the NMET Context

As mentioned in Chapter 1, washback is a common term in many official documents as well as in many research publications on language testing in China. Most of the previous washback studies in the Chinese context, however, tended to focus more on examinations at the tertiary level (Gu, 2007; Gu & Peng, 2010; Kong & Nie, 2002; Qiao, 2006; Qin, 2012; Tang & Peng, 2004), reporting different degrees of washback on teaching and learning and teachers and learners in both quantitative and qualitative ways. Despite the significance of the NMET discussed in Chapter 1, it was surprising to note that for such a large-scale, high-stakes test, there were very few empirical studies accessible that examines its washback effects (Zhou & Dong, 2014). A search of Google Scholar, Baidu Scholar (Chinese version of Google Scholar), and CNKI using keywords ‘NMET washback’ ‘高考英语反拨效应’ revealed a limited number of relevant empirical studies. Findings of those studies, however, cast great light upon the current study.

Dong (2014) found that the NMET preparation started as long as a student just entered into secondary schools and the rate and sequence of preparation increased with the NMET approaching. Xu (2010) compared the NMET with IELTS and reported that the strong negative washback effects of the NMET on students’ English writing to some extent illustrated the Chinese test takers’ poor performance in the IELTS writing module. Also, the NMET was intended for a positive washback on teaching and learning by the
MOE but results showed that after 15 years, students were still ignorant of communicative context of writing and the development of learning strategies was overshadowed by the NMET’s high-stakes nature (Kennedy & Liu, 2013; Qi, 2004; 2005; 2007; Xiao, Sharpling & Liu, 2011). Many results suggested that such a large-scale, high-stakes test induced degrees of test anxiety (Chen, 2006; Xiao, 2003; Ye et al., 1999). Results reported the existence of negative washback effects on test anxiety, but the degree of test anxiety remained further exploration (Chen, 2008). As for the NMET (ZJ) in the current study, test takers reported that they were even more test-anxious when they had more than one chance to participate in the NMET as they were always eager to get a higher score in the second test in June (Hu, 2015). However, Chen et al. (2018) reported a positive washback effect of the NMET (ZJ) on test anxiety from students' perspectives and English writing was enforced due to the adding of ‘Continuation task’ in the NMET (ZJ) (see Chapter 1).

Studies reviewed above presents a mixed picture on the washback effects of the NMET (ZJ) mainly. How washback operates on students in the NMET or NMET (ZJ) contexts still remains blurred. The findings demonstrated a need for more empirical evidence into the washback effects of on student learning and on test anxiety if policy makers intended for better use of the NMET (ZJ) as a lever for functional changes.

**Test Anxiety**

**The Definition of Test Anxiety**

The definition of test anxiety has been long unsettled during the past decades. Some researchers believe it is the tendency to view the consequence of poor performance with alarm (Sarason, 1978), which considers test anxiety as an emotional response to testing situations. Some insist that individuals show differences in an enduring, trait-like, tendency to view examination situations as threatening (Spielberger & Vagg, 1995),
which views test anxiety as a human trait. Despite the debates, test anxiety is one of the constructs located at the disciplinary boundaries of both psychology and education (Putwain, 2008b). With its negative influences reported globally, this topic has long been well researched. Research on test anxiety reached its nadir from 1980 to 1984 and has been steadily declining since (Zeidner, 1998). Although test anxiety has been studied for many decades, there is still no consensus on its definition and construct. The searching for its definition and construct has been mainly through three stages as follows:

The work of Liebert and Morris (1967) shifted test anxiety theory towards a cognitive orientation by conceptualizing test anxiety as a two-dimensional construct consisting of two components worry and emotionality. Worry is the cognitive side of test anxiety which shows an individual’s beliefs or concerns about his or her performance (i.e., cognitive test anxiety), while emotionality is the heightened psychological symptoms which involve autonomic reactions manifested in test-taking situations (Cassady & Johnson, 2002; Kieffer & Reese, 2009; Liebert & Morris, 1967; Wine, 1971). Liebert and Morris’s theory first discussed test anxiety from cognitive psychology and explained individuals’ emotional responses to evaluative situations. Therefore, although this two-dimensional construct has continuously been challenged by a later multidimensional conceptualization of test anxiety (Sarason, 1984; Segool et al., 2014), both worry and emotionality have been persistently included in theoretical conceptualizations of test anxiety (Seipp, 1991).

Based on Liebert and Morris, Sarason (1984) further suggested that test anxiety is a multidimensional construct consisting of cognitive, emotional, behavioral and physiological components. He believes that Liebert and Morris did not distinguish worry and emotionality well and that the cognitive side of test anxiety should include both worry and test-irrelevant thinking; emotionality should include tension and bodily
reaction. Thus, test anxiety is a multidimensional concept and individuals’ behavioral and psychological responses to presumed test failure should also be considered. Sarason’s another contribution to the understanding of test anxiety is his development of Test Anxiety Scale in 1978. Some question items in this scale have been constantly adopted and localized to measure individuals’ test anxiety in different contexts as I did in this study.

After Sarason’s work, it seems to reach at least one consensus that test anxiety is a multidimensional concept. And building on the work of Liebert, Morris, and Sarason, the searching for the multidimensional construct of test anxiety in the 21st century started to include a social aspect (Lowe, 2014; Lowe, Grumbein & Raad, 2011; Segool et al., 2014). The nature of test anxiety has a social dimension (Putwain, 2008b) as individuals care about how their performance will be judged by other people. Reasons behind this consideration are obvious. In the 21st century, the use of tests as tools of social and educational policy is increasing. Test anxiety is gradually becoming a confronting social problem. Segool et al. (2014), for example, elaborated a five-factor model (cognitive processes, learning experiences, demographic characteristics, social context, and environmental contingencies) which considered the social context a critical factor that inevitably influenced students’ test anxiety in a high-stakes testing context. They observed the negative effects of social context factor (such as peer reference groups) on test anxiety. When discussing social factor (peers in their study), the they revealed that students judged themselves as less capable when placed in an environment with highly capable students; however, more capable when placed in an environment with less capable students (Segool et al., 2014). Students tended to feel unconfident about tests when they compared themselves to their peers. And students in the lowest level suffered most from this peer test anxiety, or feeling test unconfident.
In summary, despite no agreement on the definition of test anxiety, cognitive sides, psychological responses and social factors are all inevitable components when discussing test anxiety. Additionally, a well-established finding is that the cognitive aspect of test anxiety (worry) is negatively associated with test performance (Chapell et al., 2005; Hembree, 1988; Putwain & Aveyard, 2016). Some learners who could be good learners in other learning situations may claim to have a mental block against learning a foreign language (Horwitz, 1986). Peers have played an important role in developing those negative feelings about tests. Identifying the role and effects of language testing on test anxiety is, therefore, of both practical and theoretical value. For example, such findings can be used to instruct the development of an intervention to minimize the influence of test anxiety on student learning (Putwain & Aveyard, 2016).

In this study, I refer to both cognitive and social aspects of test anxiety to better explain test anxiety in Chapter 4 and Chapter 5 just as Segools et al. (2014) did in their study. I did this for two main reasons. First, their model incorporated both the classic cognitive side (such as worry) of test anxiety as most previous works did and the social context factor (such as feeling less confident compared to peers) which was necessary for my study as I discussed the selection biased educational environment in China (see Chapter 1); second, they also explored this model in a high-stakes context.

**Empirical Studies of Test Anxiety in Various Contexts**

Although there has been no consensus on the constructs of test anxiety over the past decades, researchers in both psychological and educational fields uniformly focused on the role and effects of test anxiety in student learning: test-anxious students were reported to have more unsatisfactory test performance (Angus, 2001; Cassady & Johnson, 2002; Chapell et al., 2005), lower academic achievement (Cassady, 2004; Pekrun, 2001; Yousefi et. al., 2010), a higher rate of school dropout (Cizek & Burg, 2006; Gerwing et
al., 2015), reduced motivation (Cizek & Burg, 2006; Zeidner, 1998), and more mental (LeBeau et al., 2010) and physical health problems (Wadee, Kuschke, Kometz, & Berk, 2001).

Test anxiety was much researched in many countries; and levels of test anxiety were reported to vary across cultures (Bodas & Ollendick, 2005; Seipp & Schwarzer, 1996). The concept of test anxiety was much investigated in America (Sarason, 1984), Egypt (El-Safty, 1995), Germany (Hodapp & Benson, 1997), India (Sud, 2001), Israel (Zeidner & Nevo, 1993) and Korea (Schwarzer & Kim, 1984), suggesting that this was a widespread problem in the world. Test anxiety occurred among students of all ages (Chapell et al., 2005; Majidifar, 2015; Putwain & Daniels, 2010). For example, approximately 15.1% of high school students in the UK reported moderate to high levels of test anxiety (Putwain & Daly, 2014) and more shockingly, even children as young as seven reported experiencing high test anxiety (Connor, 2003). But the highest levels of test anxiety were typically reported on middle school students because performance pressures and concerns about learning achievement increased a lot compared to those in their childhood (Ahtola et al., 2011; Wampler, Munsch, & Adams, 2002). Test anxiety and its relationship with learning achievement in schools were the most frequently reported emotion among high school and university students (Pekrun, 2000) as they reported to be pressured and overwhelmed by the continuous evaluation of their academic performance (Embse, Barterian, & Segool, 2013).

In particular, the use of high-stakes testing as performance and accountability measures renewed interest in test anxiety (Putwain, 2008b). Many students were undergoing severe stress when they were asked to participate in a high-stakes test (Embsel et al, 2013; Jones et al., 1999; Putwain, 2008). Student test anxiety was higher on high-stakes tests when compared to that on regular classroom tests (Maria-Cecilia,
Test-induced stress can evolve into a never-ending cycle for low-performing students as tests became prevalent and frequent. As students aged and their experience with testing increased, their test anxiety was heightened by successive poor performance (Embse & Hasson, 2012), which may debilitate their future developments in various aspects. Increasing emphasis on high-stakes testing was often accompanied by increased levels of test anxiety. Also, high-stakes tests and test anxiety held differential significance for students with disabilities, women, and students of minority backgrounds as well (Putwain, 2007; 2008). Students who scored high on measures of test anxiety scored lower on tests than did their peers with low anxiety scores (Hurren, Rutledge, & Garvin, 2006). Many studies showed that test anxiety was related to test performance: test anxiety was even considered as one of the most disruptive factors in test performance (Cizek & Burg, 2006). For students, test performance in a high-stakes test can affect their access to the job market or academic fields (Bradley & Lenton, 2007; Heath, Rothon & Kilpi, 2008); and influence their upward social mobility (He, 1998; Yu & Suen, 2005).

**Empirical Studies of Test Anxiety in the NMET Context**

Chinese students tended to experience as much test anxiety as students in other contexts did. Research showed that 30.9% (N=508) of secondary school students in China suffered from severe test anxiety (Zhang, Chen & Liu, 2013). Like other educational contexts discussed above, studies in the Chinese settings focused more on the negative relationship between test anxiety and test performance (Chen, 2006; Jiang & Zhang, 2006; Li, Liu, Lan, & Ma, 2009). There were also investigations about how test anxiety was explicitly related to test performance on the NCEE (Jiang et al., 2005; Ye et al., 1999). The NCEE (Gaokao), as a Chinese-specific examination, was considered as the most important test of students’ life; for that reason, Chinese students tended to
experience rather high degrees of test anxiety when facing this examination (Gregor, 2005; Zhang, Chen & Liu, 2013). Gu et al. (2001) did a comparative study on students’ test anxiety levels three months before the NCEE with those levels two days before the NCEE and found the levels of test anxiety increased substantially as the NCEE approached. As a critical component of the NCEE, test anxiety in the NMET should also be valued but research on that was comparatively limited.

Wang (2007) found 45.9% (N=799) of students reported to be test anxious and 19.3% of them were even in a highly anxious level about a final English examination. Huang (2003) revealed that students’ anxiety in the English classroom was negatively related to their test scores in the NMET; and the NMET was one of the major factors to anxiety. Besides, secondary school students seemed to experience more test anxiety in an English examination than in a Chinese one; interestingly male test takers shared more of the test anxiety in English than their female peers did (Chen, 2015). Chen (2008) conducted her MA thesis on the washback effects of the NMET in Shandong province on students’ test anxiety and reported a negative washback on that. But this study did not step on how the NMET influenced student learning. Bai (2019) reviewed test anxiety studies in China between 2008 and 2018 and she found that NCEE test takers were researchers’ most concerned participants but studies with NMET test takers were not well researched. Educational reforms have taken place a lot during the last decades (e.g., the NMET (ZJ) in my study). How the washback effects of the NMET operate on learners, learning and test anxiety needs more empirical evidence. Based on questionnaire and interview data, Chen et al. (2018) suggested that the NMET (ZJ) reforms to some extent helped relieve students’ study pressure and test anxiety because students could take a second test. However, Hu (2015) shared a different opinion with them, suggesting that the NMET (ZJ) reform had not achieved its intended washback effects on reducing test
anxiety. Because students in her study reported that no matter what scores they obtained for their first NMET, they continued to prepare for the second one for a higher score and that English learning in the test preparation was test-driven.

These findings suggest that in China, educational policy initiatives and reforms have increased the significance and frequency of testing. In particular, the testing results of secondary school students for use as effective indicators of both schools and classrooms; as selective determinants of students for higher education have increased test anxiety among students. However, mixed results suggest that how the washback effects of such an important test reform function on test anxiety call for more future empirical studies to merit. The NMET (ZJ) in the current study is such a typical case.

**Summary**

In this chapter, I reviewed previous studies done in the field of both washback and test anxiety to see what so far has been done and what methodologies have those previous studies adopted. Further, I tried to build the link between those two topics in Chinese context to see what key issues and gaps existing here to allow my study to continue. As this study aimed to examine potential NMET effects on student learning and test anxiety, only eight test-affected areas on learning from Alderson and Wall’s 15 Washback Hypothesis (Alderson & Wall, 1993) serve as the main theoretical framework of my washback part in the thesis as follows:

<table>
<thead>
<tr>
<th>Number</th>
<th>Washback Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.</td>
<td>A test will influence learning.</td>
</tr>
<tr>
<td>5.</td>
<td>A test will influence what learners learn.</td>
</tr>
<tr>
<td>6.</td>
<td>A test will influence how learners learn.</td>
</tr>
<tr>
<td>8.</td>
<td>A test will influence the rate and sequence of learning.</td>
</tr>
<tr>
<td>10.</td>
<td>A test will influence the degree and depth of learning.</td>
</tr>
<tr>
<td>11.</td>
<td>A test will influence attitude towards the content, method, etc., of teaching and learning.</td>
</tr>
<tr>
<td>14.</td>
<td>Tests will have washback on all learners and teachers.</td>
</tr>
<tr>
<td>15.</td>
<td>Tests will have washback effects for some learners and some teachers, but not for others.</td>
</tr>
</tbody>
</table>
These eight hypotheses help me understand how test might influence students and their learning in a general sense. Additionally, they also help me understand the student learning question items in my study (see discussion in Chapter 3). However, as those eight learning aspects are generalized from a non-Chinese test context, a localized version of them is presented and discussed in details in Chapter 3. Sarason’s Test Anxiety Scale (Sarason, 1978) was used in the study in order to measure test takers’ test anxiety level.
Chapter 3 - Methodology

This chapter presents the research design in general, a description of participants, the instruments used in the study, the procedures in data collection and data analyses used to answer my research questions.

Research Design

The purpose of the current study was to investigate the washback effects of the NMET (ZJ) on student learning and their test anxiety. Specifically, it aimed to address the following three research questions as discussed in Chapter 1:

1. What is the washback effect of the NMET (ZJ) on student learning?
2. To what extent might the NMET (ZJ) bring a washback effect on test takers’ test anxiety?
3. How do student factors (such as students’ school backgrounds and self-rated English proficiency) affect their test anxiety?

The current study adopted both qualitative and quantitative research methods using the interview and a questionnaire to explore the washback effects of the NMET (ZJ) on student learning and test anxiety. To be specific, the interview helped build the construct of the questionnaire.

The study received the approval from both the Educational Research Ethics Board and General Research Ethics Board at Queen’s University and at South China Normal University to ensure minimal risk to both the researches and the participants (see Appendix C).

Participants

Three secondary schools were selected for the current study in Hangzhou. Among
the schools, three different levels of school settings were considered (one top school, one average school, and one low-level school respectively). Different levels of schools are of considerable importance in China because tests might have different amounts and types of washback on test takers across different school settings. For this current study, 24.8% of the participants were from a top secondary school, 39.8% from an average one and 35.6% a low-level one. The participants in the study included 406 secondary school students. Of the 406 students, 46.9% were male, and 52.8% were female aging from 16 to 19. Detailed information of the participants was presented at the beginning of Chapter 4.

Before being instructed to complete the questionnaire, participants were informed of the necessary information about the study, such as voluntary participation, confidentiality policy, and withdrawal without penalty. All participants were voluntary for the study.

**Instruments**

On the first stage of the study, the instrument of the interview was first adopted in order to design a student questionnaire. Open questions about effects of the reform on teaching, student learning and test anxiety were provided to allow altogether five interview participants to brainstorm ideas about test preparation. Those open questions were outlined in Chinese and I translate them into English for the study (see Appendix D and Appendix E respectively). Interviews were conducted through semi-structured individual interviews with two NMET(ZJ) test constructors and three English teachers; a further group interview with the teachers. Altogether six interviews were audio-recorded and transcribed into Chinese. Themes about student learning and test anxiety were extracted to design the student questionnaire in the second stage. Next, with the eight learners or learning areas out of 15 Washback hypotheses (see Chapter 2); Sarason’s
1978 Test Anxiety Scale (Sarason, 1978) and qualitative data from interviews, a student questionnaire was designed. On the second stage, a student questionnaire with three sections was used in the study: 1) demographics; 2) students’ perceptions of the washback of the NMET on their learning, and 3) students’ test anxiety during the NMET. The questionnaire was designed, piloted, and administered before it was put into large-scale use. The questionnaire was developed and administered in Chinese considering that all participants had Chinese as their native language (See Appendix A). I translated the questionnaire into English only for this thesis (See Appendix B).

Section 1 Demographics. It collected information on participants’ school names; gender; age and their self-rated English proficiency (Excellent / Average / Poor) based on their previous test scores in English examinations.

Section 2 Student Learning. It included 30 multiple choice question items asking students about perceptions of English learning during senior 3 year (see Appendix B). Based on the 8 hypothesis discussed (see Chapter 2), these 30 question items were further classified into 4 subcategories: Hypothesis 5 (hereafter H5) what to learn (item 4 to 6; item 12 to 14, six in total); Hypothesis 6 (hereafter H6) how to learn (item 7 to 11, item 22 to 23; seven items in total); Hypothesis 8 (hereafter H8) rate and sequence of learning (item 16 to 18; item 24 to 25, item 27 to 28; seven items in total); Hypothesis 11(hereafter H11) attitudes towards the content, method, etc., of teaching and learning (item 1 to 3, item 15, item 19 to 21, item 28, item 29 to 30; ten items in total). A detailed discussion of these four subcategories were presented in Chapter 4.

Section 3 Test Anxiety. It included 20 question items, each comprising a statement on a five-point Likert scale ranging from 1= strongly agree to 5= strongly disagree. Getting a mean score lower than 3 was considered as being test anxious. The lower the score, the higher the test anxiety. Item 31 to item 34 and item 44 to item 45 (six
items in total) were designed by the research center based on the reforms in the NMET (ZJ) (Cronbach’s $\alpha=.263$). As the Cronbach of those 6 designed items was too low, then they were deleted from the final data analysis (see Chapter 4). The rest 14 items (item 35 to item 43 and item 46 to item 50) were adapted from the Test Anxiety Scale (Sarason, 1978). An excellent level of internal consistency (Cronbach’s $\alpha = .84$) was reported for Sarason’s Test Anxiety scale (Raju, P. M., Mesfin, M., & Alia, E., 2010). The internal consistency of those 14 adopted items in my study was also excellent (Cronbach’s $\alpha = .883$, see Chapter 4). Therefore, only the 14 adopted test anxiety items were used for further analysis in Chapter 4.

Data Collection

The data in the current study were collected on two stages. First, in 2016, interview data was collected and coded for the purpose of designing the student questionnaire. After the interview data was coded and analyzed, common themes such as “NMET reform”, “two tests a year” and “test anxiety” emerged from the interview results and were used to design some of the questionnaire items (see item 31 to 33 for example). As interview data was not used to answer the three research questions, I did not provide more information on that part in the study. And then, questionnaire data for the study was collected in Hangzhou in 2017. Overall, students from three schools participated in the study.

Data Analysis

The questionnaire responses were entered into the Statistical Package for Social Sciences (SPSS version 24). The data analysis procedures in the study were as follows.

First, missing data was carefully examined. One participant’s gender was missing coded which remained unknown whether it was female or male. One male and three female participants’ self-rated English were missing as well. Another three female
participants’ responses to item 5 were wrongly coded. One female participant’s response to item 32; and two male participants’ responses to item 38; item 44 respectively were wrongly coded as well. As their original questionnaires were hard to find, then all those participants’ responses were teased out from the data set. Then 395 valid questionnaires were used for the data analysis in the study. Descriptive statistics included the means, standard deviations, and Cronbach’s to examine psychometric properties of scales used in the study for Section 2 Test Anxiety. For Section 1 washback on student learning, frequency was counted to see how the student learning in various aspects was influenced by the NMET (ZJ) reform. This first set of analysis addressed research question 1.

Exploratory factor analysis (EFA) with principle axis factoring (PAF) was conducted to run the factor analysis of 20 test anxiety items as PAF allowed for random measurement. Direct oblimin rotation was used as it allowed for the factors to be correlated (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Two factors (which I labeled as test worry and test unconfident) were retained for further analysis (see Chapter 4). Then, mean scores of those 14 test anxiety items were counted to see how the test anxiety was influenced by the NMET (ZJ) reform. This set of analysis helped address research question 2.

For research question 3, a correlation analysis was conducted between dependent variables (test worry and test unconfident) and independent variables (school backgrounds and self-rated English proficiency). The school backgrounds showed no significant correlation with two dependent variables. Thus I did not conduct a further regression analysis among them. The English level showed a negative correlation with test unconfident but no correlation with test worry. As there was usually a correlation between English level and test anxiety as reviewed in Chapter 2, I continued to run a regression analysis with the dependent variable (self-rated English proficiency) and
independent variables (*test worry* and *test unconfident*) to see if there was any difference.

The results were presented in Chapter 4.
Chapter 4 - Results

This chapter presents the results of the questionnaire data analysis to answer the three research questions (see Chapter 1). Thirty student learning question items were divided into 4 different subheadings: H5 what to learn (item 4 to 6; item 12 to 14, six in total); H6 how to learn (item 7 to 11, item 22 to 23; seven items in total ); H8 rate and sequence of learning (item 16 to 18; item 24 to 25, item 27 to 28; seven items in total); H11 attitudes towards the content, method, etc., of teaching and learning (item 1 to 3, item 15, item 19 to 21, item 28, item 29 to 30; ten items in total) (see in Chapter 3). Frequency of those four subheadings was counted respectively to answer the first research question. Mean scores of 14 retained test anxiety question items (see Chapter 3) were then counted to address the second research question. For research question 3, a factor analysis of the 14 items was first conducted to extract two factors (discussed in details in this chapter below) for further analysis. Then, a correlation analysis was conducted between dependent variables (two test anxiety factors) and independent variables (school backgrounds and English level). After the correlation, only the English level was retained to conduct a regression analysis with those two test anxiety factors.

Descriptive Statistics

Demographics. After missing data was addressed (discussed in Chapter 3), the final data set included 395 secondary school students between 16 and 19, \( M=17.31, SD=.538 \). Among the 395 participants, there were 185 males (46.8%) and 216 females (53.2%). Participants’ self-rated English proficiency (\( M= 2.16, SD= .691 \)) was measured on three scales: excellent (1), average (2), and poor (3) (see more details in figure 4.1). Participants’ school backgrounds (\( M= 5.12, SD= .762 \)) were also measured on three
scales: top school (4), average school (5), low-level school (6) (see more details in figure 4.2 below).

**Frequency.** Thirty question items about student learning were categorized into four subcategories (as discussed at the beginning of the chapter). The Cronbach’ alpha was used to examine the internal consistency of those 30 questionnaire items. The result showed a .515. Thus, all items were used for the following analysis. An examination of the standard deviation, skewness, and kurtosis of all 30 items suggested a normal distribution. The frequency distribution of each category was run. For the thesis, I focused the results on the highest frequency responses to each question item. This was done for two reasons: 1) to achieve reader friendliness, and 2) to understand the most salient aspects of student learning influenced by the NMET (ZJ).

**H5 what to learn.** In my study, *what to learn* referred to the learning materials students used; the content students wanted to focus on and the content students reviewed. The highest frequency distribution of response to H5 showed that what students learned highly coincided with what NMET assessed (See Table 4.1 below). To be specific, first, mock NMET test papers were used mostly as learning materials (45.6%, item 4) and they were constantly practiced in class (55.9%, item 5). Second, 40.5% of participants expected teachers to teach more writing (item 6), which weighed the highest in the NMET (Table 1.1, Chapter 1). 52.4% of them further admitted they spent more time on writing compared to the other sections in a test (see item 14).
Table 4.1

*Frequency distribution for the highest response on “H5 what to learn” question items*

<table>
<thead>
<tr>
<th>Item</th>
<th>Response</th>
<th>Valid percent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. Learning Material</td>
<td>Mock NMET test papers</td>
<td>45.6</td>
</tr>
<tr>
<td>5. How to Take English Class</td>
<td>By practicing those Mock NMET test papers</td>
<td>55.9</td>
</tr>
<tr>
<td>6. Teachers’ Focus</td>
<td>Writing</td>
<td>40.5</td>
</tr>
<tr>
<td>12. Reviewed Content</td>
<td>Contents that I am not familiar with</td>
<td>56.7</td>
</tr>
<tr>
<td>13. Reviewed Test Task</td>
<td>The one that I have most difficulties with</td>
<td>65.8</td>
</tr>
<tr>
<td>14. Type of Task in a Test Spent Most Time on</td>
<td>Writing</td>
<td>52.4</td>
</tr>
</tbody>
</table>

**H6 how to learn.** In my study, *how to learn* referred to the way students improved their four linguistic skills (reading, listening, writing and speaking) in senior three. It also referred to their English homework and tutoring at home.

The results showed how students learned English in senior three much matched with those test-related activities (See Table 4.2 below). To be specific, first, improving four linguistic skills in senior 3 relied on test-related activities: 40.8% of test takers improved their reading ability by doing reading comprehension exercises (item 7); up to 63.5% practiced their listening by taking test-related activities (item 8); as for writing ability, 54.4% admitted they wrote mock NMET composition or others (item 9); and nearly 60% of test takers (59.5% to be exact) revealed they barely practiced speaking in senior 3 as it was not tested on the NMET (item 10). Also, 68.4% of student did test papers as daily English homework (item 22).
Table 4.2

*Frequency distribution for the highest response on “H6 how to learn” question items*

<table>
<thead>
<tr>
<th>Item</th>
<th>Response</th>
<th>Valid percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>7. How to Improve Reading</td>
<td>By doing reading comprehension practice</td>
<td>40.8</td>
</tr>
<tr>
<td>8. How to Improve Listening</td>
<td>By using listening textbooks or listening-tutorial materials</td>
<td>63.5</td>
</tr>
<tr>
<td>9. How to Improve Writing</td>
<td>By writing mock NMET topics or other types of topics</td>
<td>54.4</td>
</tr>
<tr>
<td>10. How to Improve Speaking</td>
<td>By scarcely focusing on speaking as it is not tested by the NMET</td>
<td>59.5</td>
</tr>
<tr>
<td>11. How to Finish English</td>
<td>I not only finish the practice but figure out the answers</td>
<td>53.9</td>
</tr>
<tr>
<td>Practice</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22. English Homework</td>
<td>Doing all types of test papers</td>
<td>68.4</td>
</tr>
<tr>
<td>23. English Tutor</td>
<td>I never plan to hire an English tutor</td>
<td>61.8</td>
</tr>
</tbody>
</table>

**Figure 4.1 English Tutor For Students with Different English Levels**

Second, 61.8% of students said they never wanted a tutor at home to help them (item 23). Figure 4.1 showed the distribution of students in three English levels (1 = excellent; 2 = average and 3 = poor) to all four responses of item 23 (1 = “consistently hiring one”; 2 = “used to hire one but now don’t”; 3 = “plan to have one”; and 4 = “never plan to hire one”). It showed students with excellent English level displayed the highest
percentage of “consistently having one” English tutor (14.93%) and students in the other two levels showed less than 10%.

**H8 rate and sequence of learning.** In my study, *rate and sequence of learning* referred to the frequencies and sequence that students learned English inside or outside the classroom. This was reflected in how often students took compulsory English classes or weekly English tests; how much time they spent on homework or extra study, and how they arranged their private English tutoring class on the weekend.

Results from the questionnaire responses to this subcategory showed that the rate of students’ taking English class and weekly or monthly English tests remained high (See Table 4.3 below). To be specific, 83.5% participants reported one English class per day (item 16) and 65.3% said they had to take at least one or two weekly or monthly English tests (item 25). Thus, 77.0% said they spent 30 minutes to one hour on their homework (item 17), and then up to 59.0% had to spend more time to review English to ensure a good score (item 18).

**Table 4.3**

*Frequency distribution for the highest response on “H8 rate and sequence of learning” question items*

<table>
<thead>
<tr>
<th>Item</th>
<th>Response</th>
<th>Valid percent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16. Amount of Daily English Class</td>
<td>One</td>
<td>83.5</td>
</tr>
<tr>
<td>17. Time spent on English Homework</td>
<td>Half to an hour</td>
<td>77.0</td>
</tr>
<tr>
<td>18. Time spent on English Except Homework</td>
<td>Less than half an hour</td>
<td>59.0</td>
</tr>
<tr>
<td>24. English Tutoring Class on Weekend</td>
<td>I don’t take any English tutoring class</td>
<td>75.2</td>
</tr>
<tr>
<td>25. Amount of Weekly English Test</td>
<td>One or two tests</td>
<td>65.3</td>
</tr>
<tr>
<td>27. Amount of Exercise or Mock NMET Test Paper</td>
<td>Four to six</td>
<td>52.2</td>
</tr>
</tbody>
</table>
Also, 75.2% reported never taking any English tutoring class (see item 24). Similar to item 23 (see figure 4.1), figure 4.2 below shows the distribution of students in three English levels (1 = excellent; 2 = average and 3 = poor) to all four responses of item 24 (1 = “no tutoring class”; 2 = “half a day”; 3 = “one day”; and 4 = “the whole weekend”). It showed that 23.88% of students “excellent” spent at least half a day on English tutoring (which was the highest of the three levels). It further added to what was presented in item 23 (see figure 4.1).

**Figure 4.2 Private Tutor Time Students with Different English Levels Spent**

**H11 attitudes towards content, method, etc. of learning.** In my study, **attitudes towards content, method, etc. of learning** referred to how students thought of and felt about the **content, method, etc. of English learning** in senior 3.

The results showed that students even reported high psychological pressure due to NMET (see Table 4.4). To be specific, 31.9% of students claimed NMET brought about huge psychological pressure (see item 1). 47.6% of students believed NMET was average difficult (see item 19), but a high score in the NMET was as important as I discussed much in Chapter 1. Thus, it explained why 49.1% of students felt stressed about weekly
or monthly tests at school (see item 26) as each test was a practice for the NMET.

Second, 43.5 of students further claimed peer pressure was the most significant contributing factor to their learning burden (see item 29). Even though 60.8% said they handled the burden well (see item 15), they did admit the existence of English learning burden in senior 3. Some students claimed huge psychological pressure from NMET (see item 2). But a higher percent (47.6%) of students said they felt the same as senior 2 (see item 1). Therefore, negative effects of NMET on student learning such as causing students learning burden or psychological pressure appeared as early as in senior 2. Students remained passive learners as 44.8% of students reported just following what teachers required (see item 3).

However, even though there were frequent English tests for students to take (see item 26), failure in those tests pushed them (41.3%) to study harder (see item 30) rather than take them down.

Table 4.4

*Frequency distribution for the highest response on “H11 attitudes towards content, method, etc. of learning” question items*

<table>
<thead>
<tr>
<th>Item</th>
<th>Response</th>
<th>Valid percent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. English Learning in Senior 3</td>
<td>The same as in senior 2</td>
<td>47.6</td>
</tr>
<tr>
<td>2. Influence of NMET on English Learning</td>
<td>It has brought huge psychological pressure</td>
<td>31.9</td>
</tr>
<tr>
<td>3. Plan for English Study</td>
<td>I just follow my English teacher</td>
<td>44.8</td>
</tr>
<tr>
<td>15. Learning Burden in Senior 3</td>
<td>Well handled</td>
<td>60.8</td>
</tr>
<tr>
<td>19. Difficulty of NMET</td>
<td>Average Difficult</td>
<td>47.6</td>
</tr>
<tr>
<td>20. Learning Burden Compared to Other Subjects</td>
<td>Average Burden</td>
<td>51.1</td>
</tr>
<tr>
<td>21. Daily English Class</td>
<td>No Special Feeling</td>
<td>54.2</td>
</tr>
</tbody>
</table>
**Test Anxiety.** In my study, the six designed test anxiety questionnaire items (item 31 to 34, item 44 to 45, see section 3.3, Chapter 3) showed a relatively low internal consistency (Cronbach’ alpha=.231, see Table 4.5 below) when fourteen adopted test anxiety items (see section 3.3, Chapter 3) showed an alpha of .883 (see Table 4.5 below). The overall internal consistency for all 20 test anxiety question items was lower (Cronbach’s alpha= .858) when the six designed items were included (see Table 4.5 below). Therefore, all the six designed items were teased out from the data set, and only the other 14 adopted items (discussed in details in Chapter 3) were retained in the following analysis. The test anxiety items were responded to on a five-point Likert scale ranging from 1= *strongly agree* to 5= *strongly disagree*. A mean score lower than three (out of five) was considered as being test anxious. The lower score they get, the more test anxious they are.

Table 4.5

*Reliability analysis for test anxiety question items*

<table>
<thead>
<tr>
<th>Description</th>
<th>Cronbach’s Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>All 20 items (item 31 to 50)</td>
<td>.858</td>
</tr>
<tr>
<td>14 adopted items (item 35 to 43, item 46 to 50)</td>
<td>.883</td>
</tr>
<tr>
<td>Six created items (item 31 to 34, item 44 to 45)</td>
<td>.231</td>
</tr>
</tbody>
</table>

Mean scores of all 14 adopted test anxiety items were counted; however, in this chapter, I mainly reported mean scores lower than 3 to see which aspect of test anxiety was mostly affected.
In Table 4.6 below, the results showed that of all 14 items, item 35 ("Getting a good score on one test doesn’t ease my worry towards a subsequent test") and item 39 ("Before an English exam, I always feel that my peers prepare better than me") displayed mean scores lower than 3. To be specific, item 39 displayed the lowest mean score (2.77). As Segool el al.’ model suggested (Segool et al., 2014), this social context aspect of test anxiety (see discussion on their model in Chapter 2), peer-related test anxiety, worked as the most significant contributing factor to overall test anxiety in my study. It further helped explain results in item 29 (see Table 4.4).

Additionally, item 35 displayed the second lowest mean score (2.94). As Segool et al.’ model suggested, this cognitive aspect of test anxiety, worry followed as the second contributing factor to overall test anxiety in my study. It also helped explain some of the results presented in the tables above. For example, students constantly practiced mock NMET test papers (see table 4.1) because they worried that their peers would surpass them.

Table 4.6

Descriptive statistics for test anxiety question item 35 and item 39

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean</th>
<th>SD</th>
<th>95%CI</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>35. Worry</td>
<td>2.94</td>
<td>1.44</td>
<td>[2.79, 3.10]</td>
<td>.06</td>
<td>-1.33</td>
</tr>
<tr>
<td>39. Peer-related Test Anxiety</td>
<td>2.77</td>
<td>1.41</td>
<td>[2.63, 2.91]</td>
<td>.21</td>
<td>-1.26</td>
</tr>
</tbody>
</table>

Furthermore, poor and average students suffered most from test anxiety in my study. To be specific, the frequency distribution of all three-level (1 = “Excellent”; 2 = “Average”; and 3 = “Poor”) students’ responses to item 39 (see Figure 4.3) showed that as English level increased, peer-related test anxiety decreased. Peer-related test anxiety negatively influenced poor students most. On the other hand, average students were most worried about tests among all the students (see Figure 4.4).
Figure 4.3 Peer-Related Test Anxiety of Students With Different English Levels

Figure 4.4 Worry of Students With Different English Levels

Factor Analysis

First, the 14 items (item 35 to 43, item 46 to 50) were tested for the appropriateness of factor analysis with a Kaiser-Mayer-Olkin (KMO) value of 0.905, which exceeded the recommended value of 0.6 (Kaiser, 1970; 1974). The Barlett’s Test of Sphericity also indicated statistical significance (p < .05). The results suggested that those 14 test anxiety items were good for factor analysis.

Through the procedures outlined in Chapter 3, 14 test anxiety items in the questionnaire were then factor analyzed. Two methods were adopted: the Kaiser Criteria method and the scree plot method. The use of the Kaiser Criteria method produced a two-factor model while the scree plot method resulted in a one-factor model (see Figure 4.5 below). After a comparison of these two models, considering the cumulative percentage of explained variance, two factors were retained, presented and discussed with results from the Kaiser Criteria as this model were easily comprehensible and met
the criteria of interpretability.

Table 4.7 showed a two-factor model with the factor loadings based on the rotated component matrix. Those two factors were extracted with the Eigenvalues above 1.0 and 50.44% of explained variance. Based on the overall descriptions of these question items, I labeled the first factor as test worry which contained ten variables: item 39 to 50. I labeled the second factor as test unconfident which contained four variables: item 35 to 38. Within the matrix, item 38 (“The closer I get to an English exam, the more anxious I feel because I don’t feel adequately prepared”), had double loadings slightly over than .300. Since this item asked about the readiness for the test and it showed a higher loading on factor 2 “test unconfident” (.404), it was clearer when it was included in factor 2.

![Figure 4.5 Scree Plot](image-url)
Table 4.7

*Factor Loadings of 14 test anxiety question items*

<table>
<thead>
<tr>
<th></th>
<th>Factor 1</th>
<th>Factor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>43. I start to feel tense once I get the test paper.</td>
<td>.825</td>
<td></td>
</tr>
<tr>
<td>40. Before an English exam, I sweat and tremble.</td>
<td>.819</td>
<td></td>
</tr>
<tr>
<td>41. Before an English exam, I always feel nervous and couldn’t sleep well at nights.</td>
<td>.786</td>
<td></td>
</tr>
<tr>
<td>47. During an English exam, I frequently get so nervous that I forget about facts I know.</td>
<td>.753</td>
<td></td>
</tr>
<tr>
<td>48. As soon as an English exam is over, I try to stop worrying about it but I just can’t.</td>
<td>.566</td>
<td></td>
</tr>
<tr>
<td>50. I get to feel very panicky when I have to take a surprise English exam.</td>
<td>.536</td>
<td></td>
</tr>
<tr>
<td>46. During an English exam, I find myself thinking of things unrelated to the test.</td>
<td>.535</td>
<td></td>
</tr>
<tr>
<td>42. Before an English exam, I stay up late to review the knowledge that may be tested.</td>
<td>.477</td>
<td></td>
</tr>
<tr>
<td>49. As soon as an English exam is over, I always worry about potential incidents with my answer sheet.</td>
<td>.414</td>
<td></td>
</tr>
<tr>
<td>39. Before an English exam, I always feel that my peers prepare better than me.</td>
<td>.338</td>
<td></td>
</tr>
<tr>
<td>36. I am afraid my English teacher will consider me as incompetent if I fail the test.</td>
<td>.658</td>
<td></td>
</tr>
<tr>
<td>37. I am afraid my peers might laugh at me when I fail the test.</td>
<td>.466</td>
<td></td>
</tr>
<tr>
<td>38. The closer I get to an English exam, the more anxious I feel because I don’t feel adequately prepared.</td>
<td>.404</td>
<td></td>
</tr>
<tr>
<td>35. Getting a good score on one test doesn’t ease my worry towards a subsequent test</td>
<td>.368</td>
<td></td>
</tr>
</tbody>
</table>
Reliability Estimates

Cronbach’s alpha was used to measure the internal consistency of the two factors extracted above. The instrument items should be related to other items if they measure a single construct (Crocker & Algina, 1986). The internal consistency of the two factors of test anxiety was .809. Thus, the two sub-constructs of test anxiety were retained as two dependent variables in the following analysis. Before conducting a regression analysis, correlation analysis between dependent variables (test worry and test unconfident) and independent variables (school backgrounds and English proficiency) was conducted first (see table 4.8 below).

Regression Analysis

Table 4.8

Correlation between dependent variables and independent variables

<table>
<thead>
<tr>
<th></th>
<th>test worry</th>
<th>test unconfident</th>
</tr>
</thead>
<tbody>
<tr>
<td>test worry</td>
<td>1</td>
<td>.626**</td>
</tr>
<tr>
<td>school backgrounds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>self-rated English proficiency</td>
<td></td>
<td>-.100*</td>
</tr>
</tbody>
</table>

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

First, the results presenting in Table 4.8 demonstrate that school background was not correlated with either test worry or test unconfident. Thus, there was no need for further regression analysis. It also revealed that two dependent variables (test worry and test unconfident) were significantly correlated (r= .626). It further displayed that test unconfident was negatively correlated with self-rated English proficiency (r= -.100). The results in Table 4.13 showed no correlation between test anxiety and self-rated English level, and the reason behind it remained unknown to me as these two were usually correlated as reviewed in Chapter 2. Therefore, I chose to continue to conduct a regression analysis between test worry and English level. Unlike the correlation analysis
(see Table 4.8), this time I used English level as a dependent variable and *test worry* as
the independent variable to see if there were any different results with those in the
correlation analysis (see Table 4.9).

Table 4.9

*Multiple regression analysis of self-rated English proficiency*

<table>
<thead>
<tr>
<th>$R^2$</th>
<th>Adjusted $R^2$</th>
<th>F change $F$ (df₁, df₂)</th>
<th>Variable</th>
<th>Final $\beta$</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.030</td>
<td>.025</td>
<td>11.309*(1,392)</td>
<td>test worry</td>
<td>.183</td>
<td>2.874</td>
<td>.004</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>test unconfident</td>
<td>-.214</td>
<td>-3.363</td>
<td>.001</td>
</tr>
</tbody>
</table>

Note: ***$p = .001$, *$p < .05$***

For the regression analysis, mean scores of *test worry* items and *test unconfident*
items were first counted and used as variables respectively in the regression analysis. The
correlation between two factors (*test worry* and *test unconfident*) was strong while
English level showed no correlation with *test worry*. Therefore, I conducted a stepwise
regression analysis which included both *test worry* and *test unconfident* as the
independent variables; the self-rated English proficiency as the dependent variable in
order to eliminate the multicollinearity and find a better model. The results showed that
both *test worry* and *test unconfident* were all significant predictors of participants’
self-rated English proficiency. As reported in Table 4.9, *test worry* was positively related
to self-rated English proficiency; to be specific, *test worry* positively predicted self-rated
English proficiency, $\beta = .183$, $p = .004$. However, it only explained 3% of the variance in
self-rated English proficiency. *Test unconfident* negatively predicted self-rated English
proficiency, $\beta = -.214$, $p = .001$. Thus, poor level students were most likely to feel
unconfident about the test.
Chapter 5 - Discussions and Conclusions

This chapter restates the purpose of this study and discusses some key findings. It also provides some implications and states the limitations of this study in the end.

The purpose of this study is to answer three research questions concerning the washback effects of NMET on student learning and test anxiety. It also aims to see how individual factors such as school backgrounds and English level influence test anxiety. Results of the NMET (ZJ) on student learning were discussed first, followed by a discussion on test anxiety. How English proficiency was related to test anxiety was presented next. Implications and limitations of the study were discussed in the final section.

The Washback Effects of the NMET (ZJ) on Student Learning

To answer the first research question “What is the washback effect of the NMET (ZJ) province on student learning?”; 30 question items concerning the washback effects of the NMET (ZJ) on student learning were divided into 4 subscales (H5 what to learn; H6 how to learn; H8 rate and sequence of learning; and H11 attitudes towards content, method, etc. of learning) based on Alderson and Wall’s 15 Washback Hypotheses (see discussions on the subscales in Chapter 3). Therefore, discussions on the first research question were presented in terms of those subscales sequentially.

Washback Effects on What to Learn

This study yielded a key finding in this subsection that strong negative washback effects on what students learned were observed and that students reported they learned English to the test (see Table 4.1, Chapter 4). To be specific, students’ choices regarding the learning areas to cover or ignore at school or home were mediated by NMET (ZJ) test
syllabi. And mock NMET test papers were constantly practiced inside and outside the classroom (see item 4 and item 5, Table 4.1). Those findings suggested that learning content in a high-stakes context is inevitably manipulated by the test just as other studies in and out of the NMET (ZJ) context suggested (Chen, 2011; Chen et al., 2018; Pan, 2016; Zhang, 2018).

Additionally, there was a negative washback on students’ learning of English writing. In more details, over half of the students (see item 14, Table 4.1) admitted they spent most time on the new writing task called “Continuation task” (discussed in Table 1.1, Chapter 1) and 40.5% of them said they really wanted teachers to teach them more writing skills for a better score on that task (see item 6, Table 4.1). And it turned out that those kinds of skills were just writing and memorizing NMET topics (see item 9, Table 4.2). This was contradictory to Zhang (2018)’s conclusions in Shanghai’s new NMET. She believed that the new writing task in NMET (Shanghai) brought about positive washback on students’ writing ability by finding students spent more time on writing. However, an increase in time on learning to write does not always meet students’ real needs as both Shohamy et al. (1996) and Qi (2005) agreed. In this study, there is always a lack of communicative context of writing and too much emphasis on meeting the rating standards. This writing task was intended for a positive change because the old one in the NMET (ZJ) was long criticized for being too testlike (Han, 2012). But the findings on writing suggested that this intention seemed not to be achieved as intended. Possible explanations for that could be that this is the second cohort of NMET (ZJ) test takers and it takes time for washback to operate. The high stakes of NMET (ZJ) is another explanation because tests with too much high stakes will always fail to exert the intended washback effect although they have the potential to do so (Qi, 2004).

To sum up, previous studies showed that tests affected student learning positively,
negatively or even in both directions (Andrew et al., 2002; Ferman, 2004; Gu, 2007; Tsagari, 2009; Watanabe, 2004). In this study, only negative effects were observed. It seems that when a major change was introduced into the existing system, students’ learning content immediately responded to that change by taking a learning-to-the-test strategy. Wall and Horak (2006) help us understand test effects on teachers’ reaction when a major change was introduced and this study added to that understanding from students’ perspectives. Shohamy et al. (1996) compared a low-stakes test (Arabic Second Language Test) and a high-stakes test (English Foreign Language Test) which affected entrance into higher education. They concluded that even slight changes in such a high-stakes test caused strong negative washback in terms of what students learned. Then, it would not be odd to see those findings in my study when there were big changes in terms of test frequency and test format.

**Washback Effects on How to Learn**

This study yielded a major finding in this subsection that strong negative washback effects on how students learn English inside and outside the class were observed (see Table 4.2, Chapter 4). In more details, the ways students improved their four linguistic skills (listening, speaking, reading and writing) were highly examination-oriented (see Table 4.2, Chapter 2). The difference in the way they treated speaking and writing is even more convincing in this subsection: over half (54.4%) of the participants practiced mock NMET writing topics because writing was the most valued section and 59.5% admitted they ignored speaking in senior 3 as it was not tested in the NMET (ZJ) (see item 9 and item 10, Table 4.2). This suggested that students learned English for the sake of testing instead of genuine learning and my conclusion on that is not alone (Cai, 2013; Chen, 2011).

Additionally, some washback studies were conducted by comparing test scores
while the promotion of four linguistic skills was less researched. Therefore, there was little evidence showing whether students have learned more as a result of a particular test (Tsagari, 2007). After all, the adding of test scores does not always imply the adding of English ability especially when some ability such as speaking or communicative ability which is key to language learning is not shown in a test. Qi (2007) lamented that there was a lack of the communicative context of learning but only an emphasis on the test situation and the assumed preferences of the markers in the Chinese NMET context. In my study, although the increase of value in some learning section such as writing brought about students’ more preparation into it (see item 6 and item 14, Table 4.1), that preparation remained to be NMET test-driven as usual (see item 9, Table 4.2). Thus, it would be just a dreadful process of learning to the test instead of learning to the learning itself. NMET studies in other provinces also suggested that test tasks brought about negative effects on how students learned English as there was just too much drilling on test training in the preparation process to get a higher score (Ma, 2010; Zuo, 2010). Washback studies outside the NMET context also concluded that there was a strong negative correlation between examination patterns and how students learned English and that secondary school students were highly relying on rote-learning, engaging in test-preparation activities and developing test-taking strategies (Ahmed, 2018; Rind & Mari, 2019). Thus, this study would conclude that a learning-to-the-test strategy may not be a Chinese-specific scenario but a universal one as long as high-stakes tests are valuable in that context.

At the end of this section, I would discuss some results on English private tutoring in NMET (ZJ). Yung (2015) concluded that English private tutoring (also called shadow education in Hong Kong context) helped students better prepare for all relevant examinations they faced but failed to prepare them with an authentic use of English.
Students admitted that English private tutoring was very helpful for them to learn examination skills, but they did not regard it as an effective means to increase their English proficiency for authentic communication or interaction (Yung, 2011). My study did not go that further and only found out that excellent students of all three English levels were most likely to hire a tutor (see Figure 4.1). Little is known about shadow education in either the NMET or the NMET (ZJ) context and its effects on student learning. This study did not explore much on that either. But it will be an interesting topic for future studies to explore further when considering shadow education is becoming an increasingly big business in China.

**Washback Effects on Rate and Sequence of Learning**

This study yielded a finding that negative washback effects on the rate and sequence of English learning were observed (see Table 4.3, Chapter 4). In more details, 83.5% of students took one English class per day (see item 16, Table 4.3) and including that, they (77.0%) spent half to an hour on daily homework (see item 17, Table, 4.3) and another half an hour doing extra study on English (see item 18, Table 4.3). In fact, this high rate of learning was supposed to be positive effects. However, considering what students did during that learning process was mainly test-oriented (see discussion in this Chapter above), it was shocking to see that students were rote-learning test-guided English at such a high rate. Another finding was about testing frequency. There was weekly or monthly English tests for most students (65.3%) in the class (see item 25, Table 4.3). English test frequency was high not only in my study but also in Shanghai’s NMET context (Zhan, 2018). This study would conclude that the new NMET (ZJ) did not only impact their rate of learning but also their rate of testing for that learning outcomes. Then, it would not be difficult to understand another finding about shadow education. Outside the classroom, this study found that excellent students of all three
levels were engaged in more intensive time and rate to take private tutoring class on the weekend (see Figure 4.2, Chapter 4). This finding was contradictory to Ferman’s (Ferman, 2014). This could be explained by the fact that just as discussed in Chapter 1, Chinese students from ancient time valued good results in a high-stakes test over many other things because those results led to huge consequences such as not being admitted to tertiary education (Yung, 2011; 2015). The higher the stakes, the higher the value.

This study did not go much further on the sequence of learning much and little is known about that in other studies either. For example, when students learn English, what would be their first priority? Is it grammar or anything else? It will be another interesting topic for follow-up studies to see the various impact of test on different section of English learning.

**Washback Effects on Attitudes towards Content, Method, etc. of Learning**

This study yielded a major finding towards students’ attitudes towards the learning content. And those attitudes were not intended. In more details, 31.9% students reported huge psychological pressure towards English learning due to NMET (ZJ) (See item 2, Table 4.4) and they (54.2%) said they had no special feelings towards their English daily class (see item 21, Table 4.4). It is well-stated that the MOE intended for a positive effect on students’ English learning attitudes towards English to make it less tense but more relaxing by changing the NMET (ZJ) test frequency and some test format. Well, that intention was not achieved in this study. Either is in Shanghai’s context (Zhang, 2018). Additionally, this study also found that the intensive preparation for NMET (ZJ) started as early as in Senior Two (see item 1, Table 4.4). When they entered Senior Three, some of them already felt ‘numb’ about what they learned in the class. It is even shocking to see that Dong (2014) concluded the NMET preparation started as soon as students just began their high school as freshmen and that test preparation got more intense as the
NMET approached. Similar conclusions were also made in the other high-stakes testing contexts outside China (Ahmed, 2018; Cho, 2010; Rind & Mari, 2019).

Second, another finding was on the attitudes towards the learning strategy. Nearly half of them (44.8%) did not have their study plan for English and just followed their teachers (see item 3, Table 4.4). Thus, most NMET (ZJ) test takers seemed to be just passive receivers of testing skills from their teachers in the learning process. This finding is not welcome because in an English class, students are supposed to be an active enquirer, not a passive receiver because these passive or even negative attitudes towards learning English may lead to bad consequences such as limited opportunities for creative expressions in the classroom (Aldabbus, 2008; O’ Dwyer, 2006). It’s even more pitiful to see this finding is not limited to the Chinese context (Ahmed, 2018).

The Washback Effects of the NMET (ZJ) on Test Anxiety

This study yielded its first finding on the type of test anxiety brought about by the test. As discussed intensively in Chapter, test anxiety is a complex construct. A convincing discussion of it could not stand up without a clear identification of its components in specific situation. In this study, item 35 ‘Getting a good score on one test doesn't ease my worry towards a subsequent one’ and item 39 ‘Before an English exam, I always feel that my peers prepared better than me’ displayed mean scores lower than 3 (see item 35 and item 39, Table 4.6, Chapter 4). Those results suggested that the cognitive aspect of test anxiety, worry (see item 35) and the social aspect of test anxiety, peer-related test anxiety (see item 39) contributed most to the overall test anxiety in my study. Thus, the test affects students’ cognition of it most and put them into a worrying situation. And it also affects their ideas about their peers.

The NMET in Shanghai found a similar result (Zhang, 2018). Students (44.57%) in that context complained that the second chance for the NMET (Shanghai) was not a
good choice because it made them a very tight study schedule. Thus, it made them more test anxious. That result is understandable. Because no matter what score students obtained for the first test, they still wanted to try second one hoping for a higher score (Hou, 2017), which in turn increased their test anxiety.

Another finding on test anxiety is about students’ different level of it. This study found that average students suffered most from worry and poor level students suffered most from their peers (see Figure 4.3 and 4.4, Chapter 4). That result is understandable. As high-stakes tests exist, test anxiety exist as well. Previous studies suggested that students in all levels were worried when asked to take a high-stakes test (Paris et al., 1991; Putwain, 2008). But this study provides a more detailed picture by finding the different amount of test anxiety on different students. On the other hand, students reported higher levels of test anxiety when they considered the social factors such as peers to be highly competitive, which was also called “little fish, big pond” effect (Marsh et al., 2001; 2014). In my study, poor level students showed the highest level of peer-related test anxiety, and that result was also supported by previous studies inside or outside the NMET context (Bossy, 1996; Goetz et al., 2008; Segool et al., 2014; Wang, 2007). An effort should be made to promote this cohort of students in dealing with negative emotions towards the test.

Test anxiety in washback studies is not well-researched previously. The role of negative emotions or so-called test anxiety in shaping washback should be valued (Murry, Raizi & Cross, 2012; Rao et al., 2003).

**Relationship between Test Anxiety and Self-rated English Proficiency**

In my study, I found two factors of test anxiety (I labeled them as test worry and test unconfident) based on the factor analysis results (see Table 4.7, Chapter 4). To be specific, test worry positively predicted English level but accounted for only 3% of the
variance. However, I believed this result called for further justification as there was no correlation between the two (see Table 4.8, Chapter 4) and the regression model seemed not to be perfectly fitted (see Table 4.9, Chapter 4). Second, test unconfident was negatively correlated with English level and further a negatively significant predictor of it (see Table 4.8, Chapter 4). It implied that poor level students felt most unconfident about the NMET (ZJ) which was echoed by study outside the NME (ZJ) context (Roderick & Engel, 2001). However, Chen et al. (2018) conducted their washback study of the 2017 NMET (ZJ) and reported a positive result. A possible explanation of the difference was maybe they used only two question items to measure test anxiety. As test anxiety was commonly known as a complex construct, I believe two items were unable to cover a holistic picture of it.

In summary, this study explored how washback operated in the NMET (ZJ) by using a student questionnaire. The NMET (ZJ) induced different levels of test anxiety among different students as discussed above. However, it was a pity that the regression model failed to see the relationships among school backgrounds, test anxiety and English level. Follow-up studies are highly recommended on this interesting topic. And a more well-designed test anxiety questionnaire is suggested as well. To ensure its good use, a pilot study might help.

Revisit the 15 Washback Hypotheses

The contribution of the 15 Washback Hypotheses to the whole field was well reviewed (see Chapter 2). In Alderson and Wall’s 1996 paper, they also revisited the Washback Hypotheses and suggested an expansion of the 15 Washback Hypotheses that “Tests will have different amounts and types of washback on some teachers and learners than on other teachers and learners.” It’s a pity that little expansion of the original hypotheses has been done since that. Washback does not always stay the same as testing
situations change. Thus, a static attitude towards it is not encouraged for better use of tests as effective levers for teaching and learning.

Based on the findings and discussions in this study, when I revisit the original 15 hypotheses, I do not find a hypothesis about test impact on test anxiety. Thus, this study suggests a 16th hypothesis that tests will have different amounts and types of washback effects on test anxiety of some learners’ than on other learners’. And that new one calls for more empirical evidence to justify and enrich it. Based on the results of the study, it will be fruitful for future studies to collect more large-scale data. Meanwhile, my study will also encourage teachers’ voices to be captured in this new hypothesis. How will teachers feel about test anxiety? Will they experience the same anxious situations as their students do, or maybe in a less serious level? If they do, how will they tackle those negative test emotions during the preparation process and will those emotions affect their teaching beliefs or even teaching behaviors in their classroom? Questions like those may merit follow-up studies.

To sum up, the consequences of the use of test such as test anxiety explored in my study within specific educational systems were less eye-catching. Not to mention the consequences of a reformed NMET (ZJ) as in the current study. Green (2013) points out that language testing is traditionally more concerned with the issues of test design as they lied more clearly within the control of policymakers. Bachman and Palmer (2010) lamented that no matter how hard test designers tried, the skills they intended could never be fully equal to those skills required for successful use in the target language situations. Perhaps it’s time for test makers to find out why.

Implications and Limitations

This study is robust in its findings: what and how students learn is negatively affected by the test. Specifically, the test negatively affects learning strategies through
pushing the students towards mastering the test-tackling strategies rather than learning English for the sake of genuine learning. It also negatively affects learning contents as test-oriented practice is students’ priority in the classroom. The majority of the students studied and practiced mock NMET test papers and materials as much as they can.

Besides, most of the students did self-study to master the strategies for the test at home. In the context of NMET (ZJ), this study has cast doubt on the changes of test frequency to decrease test anxiety. Test anxiety remains a confronting issue for poor and average students especially. Peer anxiety and students’ negatively cognitive comprehension of the test are the two biggest contributing factors to the overall test anxiety in this study. Based on those findings, several implications are shown as follows:

First, leveled learning is suggested and students in the same English level should take an English class in the same classroom. This kind of leveled learning should be taken into consideration after the first NMET (ZJ) test. In this way, students can rearrange their learning schedule in Senior Three if they do not take the second test. Then, they will no longer worry about the English test, and it can leave them more time and efforts for other subjects. For teachers, it also leaves them more energy to focus on those poor students who want to take a second test. Whether teachers can create environments that make low-achieving students feel supported in responding to new demands will form a significant impact on students’ test anxiety and passing rates.

Second, students’ differences in levels of test anxiety should also be focused. Both students and teachers should manage to find more appropriate ways to prepare students in different levels well for the test. Meanwhile, it is not realistic to eliminate test anxiety because as long as there is a high-stakes test, there will be test anxiety. What students and teachers can do is to maintain that level of test anxiety into a reasonable amount. Therefore, students themselves should learn to develop a proper attitude towards
it and come up with adequate ways to face it.

The findings of the study also have implications for future research. First, English private tutoring (also called shadow education) is receiving increasing attention as a global educational phenomenon and a big industry (Yung, 2011; 2015). The study does not provide a clear picture of the shadow education in the NMET (ZJ) context, but it is commonly well-known and used among students and parents in China. How this shadow education might affect student learning and their test anxiety during the NMET test preparation will be a fun topic to explore.

Despite its implications, it is also important to address the limitations of this study in terms of research methods and participants. First, only questionnaire data were analyzed to answer the research questions in the study. Without access to interviews with students to listen to their own experience and observation of their classroom behavior, it was not critical enough to conclude on some abstract concepts such as student learning and their test anxiety. For example, Chen et al. (2018) used a teacher questionnaire, a student questionnaire, and interviews with both teachers and students to explore the NMET (ZJ) reform. Also, the population of participants in this study was limited to Hangzhou without other cities in Zhejiang. Thus, a follow-up study which can cover a wider range of areas is encouraged to make the results more representative.

This study is a start to explore the washback effects of the latest NMET reform on student learning and test anxiety. In answering the third research question, the factors investigated were limited. First, the study included only two test anxiety factors and one of the two factors, test worry, was not a strong predictor of English level. Second, the study only included two traditional factors, and one of them (school backgrounds) could not be used for further analyses because it did not correlate with either of those two test anxiety factors. Therefore, other factors, especially those unique to NMET test takers
such as their prior NMET test experience or their first NMET test score), should be included in future studies.

Further, this study encourages research into strategies to reduce test anxiety among students. The findings of the study demonstrated that students were negatively affected by test anxiety. Further studies can try to find out the strategies to reduce test anxiety by interviewing students with a low level of test anxiety, teachers as well as test experts.

To sum up, the study is one of the few washback studies conducted on learning and learners in the NMET context. The findings of the study contradict the intention that the test was used as a lever for positive changes in the classroom. It is also one of the few washback studies conducted on test anxiety in the NMET context. The findings of the study were adequate to alert all stakeholders involved in the test to the test anxiety and its potential consequences on students. This study contributes to the understanding of how student learning and test anxiety were affected in the NMET reform context.
References


61


Cheng, L. & Qi, L. (2006). Description and examination of the National Matriculation


Methods, 4(3), 272-299.


LeBeau, R. T. et al. (2010). Specific phobia: A review of DSM-IV specific phobia and
preliminary recommendations for DSM-V. *Depression and Anxiety, 27.*


Marsh, H. W., Koller, O., & Baumert, J. (2001). Reunification of east and west German


Pan, J. (2016). *Backwash Effects of National Matriculation English Test on Senior High*


Putwain, DW, & Daniels, RA. (2010). Is the relationship between competence beliefs and


Segool, N. K. et al. (2013). Heightened test anxiety among young children: elementary school students’ anxious responses to high-stakes testing. *Psychology in the*
Schools, 50(5), 489-499.


Ye, J (1998, b). An investigation and reflection on the washback effects of the CET.
Foreign Language World, 71(3), 40-43.


University.

[A quantitative analysis of documents on examination anxiety in China from 2008 to 2018]. Teaching of Forestry Region, 05, 107-110.

Peking University Education Review, 08(2), 71-81.

Chen, F. (2015). 高中生语文、数学、英语考试焦虑的比较研究
[A comparative study of test anxiety among Chinese, Math and English in senior high school students]. Gaokao, (9), 208-209.

[Correlations between cognitive components of test anxiety and academic performance of students in junior middle school]. Journal of Clinical Psychosomatic Diseases, 12(1), 35-38.

[Explore the reform of College Entrance Examination System – the ‘3+1’ College Entrance Examination Subject Setting Scheme]. China Exam, 1.


Chen, Y. et al. (2018). 外语高考改革新模型的反拨效应研究

Dong, L. (2014). 高考英语测试的反拨效应探析—基于一所普通高校新生的问卷调查


Gu et al. (2001). 中学生考试焦虑及相关分析

[Examination anxiety of middle school students and related analysis]. *Career and Health*, 17(6), 58-59.


Gu, X. (2007). 大学英语四、六级考试对中国大学英语教学的反拨效应实证研究


[A longitudinal study on college English teachers’ perceptions of CET and its washback]. *Foreign Languages & Their Teaching*, (6), 37-41.

He, J. M. (2000). 中国高考报告


Hou, L. (2017). 例谈语篇分析与高考新题型读后感写的结合
[Discourse analysis and continuation task in NCEE]. New Curriculum, (6).

[An analysis of the advantages and disadvantages of College Entrance Examination in Zhejiang province: taking the Wenzhou area as the research object]. Education Modernization, (11), 67-69.

Huang, Q. (2003). 中学生外语焦虑研究--以全国英语高考成绩为例
[Study on foreign language anxiety among senior high school students: taking the NMET test scores as research objects]. Elementary EFL Teaching Journal (1), 28-31.


Huang, D., & Yang, B. (2002). 语言测试反拨效应研究概述
[A review of the washback effect research in language testing]. Foreign Language Teaching and Research, 34(4), 288-293.


Kong, Y. A., & Nie, J. (2002). CET 中复合式听写及其对教学的反拨作
[Compound dictation in CET and its backwash effect on language teaching].
*Foreign Language World, (2),* 51-57.


Liu, Q. S. (2008). 改革开放三十年来我国高考英语科的发展情况
[Development of the NMET over the past thirty years]. *Curriculum Teaching Material & Method, (4),* 22-27.

Lu, L. (2017). 近五年浙江省高考英语完形填空题型效度分析
[An analysis of the validity of Cloze task in College Entrance Examination in Zhejiang province in the past five years]. *School English, (9),* 124-125.

Ministry of Education. (2001). 教育部发布 2001 年高考“3+X”考试内容和形式改革方案[Reform scheme of content and form of "3+X" Examination in 2001 College Entrance Examination]. *China University Enrollment, (1)*


Qin, X. (2012). 充分利用四、八级考试的反拨效应,抑制教学质量滑坡
[Make full use of the washback effect of CET-4 and CET-8 to restrain the decline of teaching quality]. *Foreign Language World, (3),* 10-14.

[The State Council’s opinions on deepening the reform of examination enrollment system]. Retrieved from
Sun, Y., & Chen, G. (2012). 学生高考前考试焦虑状况及其影响因素研究
[Study on Status of Test Anxiety and Its Influencing Factors in Students before the Entrance Examination for College]. China Journal of Health Psychology, 20(6), 867-868.


[Study of NMET from the perspective of authentic evaluation]. China Examinations, 08.


Wen, X. M. (2011). 高考英语（四川卷）对高中英语学习的反拨作用研究
[Washback effects of NMET (Sichuan) English learning in senior high school levels ].
Folk Art and Literature, (2), 118-118.

Xiao, L. (2003). 高三学生社会支持与焦虑水平关系的测验研究
[On the correlations between social-support and anxiety of the students in grade
three of senior middle school]. *Journal of Binzhou Teachers College, 19*(4), 79-84.

Xu, Y. (2010). 从高考英语反拨效应看中国学生的雅思写作难题

[ Chinese students’ IELTS writing problems from the perspective of NMET washback effects]. *Overseas English, 10*, 133-134.

Ye, M. et al. (1999). 高三学生心理变化及对高考成绩的影响

[Psychological change of senior students and its effect on College Entrance Examination]. *Chinese Journal of Behavioral Medical Science, 8*(1), 18-20.

Zhan, L. L (2018). 上海高考英语“一年两考制”研究


Zhang, W. L. et al. (2013). 中学生应对倾向在成就目标与考试焦虑中的中介作用


[Study on the relationship between candidates’ test anxiety and social comparison]. *China Journal of Health Psychology, 20*(1), 70-72.


Appendix A  A Student Questionnaire (In Chinese)

高考英语反拨效应问卷调查（2017 届学生卷）

亲爱的同学：
为了进一步了解高考英语对高三学生的学习带来的影响，以及学生的学业负担及考试焦虑状况，我们组织了本次问卷调查。问卷调查采用匿名方式，有关结果只用于课题研究，与你的学业成绩评定无关。请结合你的真实情况和想法如实填写。感谢你对本研究的支持与合作！

华南师范大学外文学院课题组
2016 年 11 月 3 日

说明：本问卷共 4 页，分两部分，第一部分是关于答卷者信息的填空题和选择题，第二部分是关于高考英语反拨效应的单项选择题，分两小节。

第一部分答卷者信息
学校名称：性别：年龄：
个人英语水平在班中所处的位置（在方格内打√）： □ 优秀 □ 中等 □ 比较弱

第二部分高考英语反拨效应
第一节单选题（共 30 小题）
请根据个人情况选择最佳答案，并把与答案相应的编号字母写在题前括号内。
（）1. 你如何描述你个人高三的英语学习感受？
    A. 比较愉快  B. 与高二时差不多，没有什么特别感觉
    C. 很紧张  D. 有厌倦英语学习的感觉
（）2. 你觉得高考对你的英语学习有什么影响？
    A. 有很大的促进作用  B. 没有任何影响
    C. 影响了正常的学习  D. 带来很大精神压力
（）3. 你对高三英语学习有个人规划吗？
    A. 没有规划，只是跟随老师的复习计划  B. 只规划某个方面（如语法等）
    C. 对各方面复习有大致规划  D. 针对每阶段都有详细的计划
（）4. 高三期间，你最常用哪一类的英语学习材料？
    A. 教材  B. 高考模拟试卷与真题  C. 教辅  D. 课外读物（如报刊）
（）5. 高三英语课堂中最常见的上课内容是什么？
    A. 模拟题、真题、套题等的讲练  B. 教材中内容的学习
    C. 语法（词汇）的讲解  D. 高考题型专项训练
（）6. 你最希望老师在课堂上加强哪块内容的讲练？
    A. 词汇和语法  B. 写作  C. 阅读理解  D. 完形填空
（）7. 在高三，你主要通过什么方法提高阅读能力？
    A. 多做模拟卷中的阅读理解题  B. 多背单词
    C. 经常看英文杂志和短篇小说  D. 经常关注文章的文体和结构
（）8. 在高三，你主要通过什么方法提高听力能力？
    A. 观看英文影片、电视等  B. 听国外的英文广播
    C. 做高考历年真题和模拟题  D. 使用听力教材或资料进行训练
（）9. 在高三，你主要通过什么方法提高写作能力？
A. 以历年高考写作题型为主，练习各种题型或写作主题
B. 背诵往年高考经典范文
C. 背诵含开头、结尾的各类作文模板
D. 经常用英文写电子邮件、日记或周记

（）10. 在高三，你主要通过什么方法提高口语能力？
A. 与别人（如同学、朋友、家长）练习对话
B. 坚持朗读英语课文
C. 在课堂上积极用英语完成老师给的任务
D. 因高考不考口语，不大重视，只在课堂上偶尔说几句英语

（）11. 在高三阶段做题时，你的最常见做法是？
A. 只求正确完成，其他不关注
B. 不但要完成，还会弄清正误原因
C. 做出答案后，找出关联知识进行复习
D. 将错题汇总记录

（）12. 公然进行复习时，你认为哪些是最重要的复习内容？
A. 高考常考的知识点
B. 自己掌握得不好的知识
C. 考纲上列出的知识范围
D. 教材课文里的重点难点

（）13. 公然进行复习时，你认为最需要关注什么题型的专项技能训练？
A. 自己最感兴趣的题型
B. 高考中最容易得分的题型
C. 自己感到最困难的题型
D. 高考中占分值最多的题型

（）14. 在实际的复习中，你在以下哪类题型上花的时间最多？
A. 语法填空
B. 阅读理解
C. 写作
D. 完形填空

（）15. 就你自身而言，你认为高三的英语课业负担如何？
A. 没有课业负担
B. 负担一般，可以承受
C. 负担比较重
D. 负担很重

（）16. 你每天平均上多少节英语课？
A. 1 节
B. 2 节
C. 3 节
D. 4 节及以上

（）17. 通常情况下，你每天需要多长时间完成英语家庭作业？
A. 0.5 小时-1 小时
B. 1 小时以上-2 小时
C. 2 小时以上-3 小时
D. 3 小时以上

（）18. 除老师和家长布置的作业外，你每天自主学习英语的时间有多长？
A. 不到 0.5 小时
B. 0.5-1 小时
C. 1 小时以上-2 小时
D. 2 小时以上

（）19. 通过做真题、模拟题等，你觉得当前高考英语试题的难度如何？
A. 比较简单
B. 难度一般
C. 比较难
D. 非常难

（）20. 与其他科目相比，你感觉英语学科给你造成的学习压力如何？
A. 没有什么压力
B. 压力一般，可以承受
C. 压力比较大
D. 压力很大

（）21. 一般情况下，上完英语课后你感觉如何？
A. 还比较轻松
B. 没有特别感觉
C. 感到疲劳
D. 非常累

（）22. 在你每天的英语家庭作业中，什么类型的作业最多？
A. 背诵（课文单词知识点等）
B. 抄写（单词句子课文等）
C. 练习册
D. 各类试卷

（）23. 你请英语家教进行辅导的情况是怎样的？
A. 一直都在请
B. 以前请过但现在没有请
C. 现在没有请，但以后打算请
D. 从来没有打算请家教

（）24. 每周双休日，你大约要用多长时间参加各类英语补习？
A. 不参加
B. 参加半天
C. 参加 1 天
D. 参加 2 天

（）25. 包括课堂测验在内，你每周英语测试有几次？
26. 对于学校的周测、月考等英语考试，你感觉压力如何？
A. 能轻松应对   B. 没什么压力   C. 有一定压力   D. 压力很大
27. 请你估算一下你手上的英语练习册、模拟题等教辅材料一共有几本？
A. 3 本及以下   B. 4-6 本   C. 7-9 本   D. 9 本以上
28. 你的教辅材料中有几本是由你自己决定购买的？
A. 3 本及以下   B. 4-6 本   C. 7-9 本   D. 9 本以上
29. 你认为你的英语学习压力最主要来源于哪方面？（如果你感到没有压力，请考虑你的同学的压力来源）
A. 父母的要求   B. 老师的期望   C. 同学间的竞争   D. 自我要求
30. 一次英语考试的失利会对你带来什么影响？
A. 没什么影响，无所谓   B. 觉得自己很没用，对英语学习失去信心
C. 情绪波动很大，需要时间调整   D. 促使自己更努力地学习
第二节单选题（共 20 小题）
请阅读以下说法，然后根据你的真实情况或想法选出代表你对相关说法的认同程度的答案，并把与答案相应的编号字母写在题前括号内。
以下各题的选择项是：
A. 非常符合我的情况（想法）   B. 比较符合我的情况（想法）
C. 基本符合我的情况（想法）   D. 不怎么符合我的情况（想法）
E. 根本不符合我的情况（想法）
31. 我赞成英语高考只考一次，考两次让我感到很大压力。
32. 因为我有多于一次参加高考的机会，所以我的焦虑情绪会减轻。
33. 虽然我在第一次英语高考中成绩不错，但我还是要参加第二次，争取更好成绩。
34. 进入高三以来，我常感到有大脑蹦紧的感觉。
35. 尽管目前这次英语测试成绩还不错，但我还是会担心下一次可能会考砸。
36. 我担心如果自己的考试成绩不好，英语老师会觉得我能力不强。
37. 我害怕自己英语考砸了的话同学们会笑话我。
38. 越临近英语考试，我越觉得自己没有复习好，紧张不安。
39. 参加英语考试前，我会觉得别人的复习工作比我做得更好。
40. 参加英语考试前，我会出现手心出汗、身体发抖等不适现象。
41. 参加英语考试前，我常常感到紧张，晚上睡不好觉。
42. 参加英语考试前，我会熬夜复习可能考的知识点。
43. 英语考试时，一拿到试卷，我就觉得很不自在。
44. 英语考试时，一拿到试卷，我有时会忘了填写个人信息而急于答题。
45. 英语听力考试时，我总担心自己使用的听力设备有问题。
46. 英语考试时，我发现自己无法集中精力，容易在考场想一些与考试无关的事情。
47. 英语考试时，我经常感到很紧张，以至于本来记得的内容也会忘记。
48. 英语考试后，我试图停止对考试的担忧，但做不到。
49. 英语考试后，我总担心答题卡等问题出现问题。
50. 面对英语老师的“突袭式”考试，我感到十分紧张。
--------- 调查结束感谢合作 ---------
Appendix B

A Questionnaire Survey on the Washback Effects of the National Matriculation English Test (NMET) (for 2017 test takers)

Dear Students:

We conduct this questionnaire survey for a better understanding of the effects of the NMET on your learning; your study pressure as well as your English test anxiety. The survey will be anonymous, and its results will only be used for academic purposes. It has nothing to do with the evaluation of your academic performance so please feel free to share your ideas. We feel very thankful for your support and your cooperation as well for this study.

Research Group in the School of Foreign Studies,
South China Normal University
November, 3rd, 2016

Note: This is a four-page questionnaire with two sections. This first section is a demographic information part and the second section consists of multiple choice questions asking about the washback effects of the NMET dividing into two sub sections.

The First section: Demographic Information
School Name: Gender: Age:
Your self-rated English level (Put √ in the square) □Excellent □Average □Poor
The Second Section: the Washback Effects of the NMET
The first part: multiple choice questions (30 question items)
Please choose only one answer based on your situation.

1. How do you think of your English learning in senior 3?
   A. Quite pleasant.         B. The same as in senior 2.

2. To what extent does the NMET influence your English learning?
   A. It has a very positive influence.
   B. It does not influence at all.
   C. It has some influence on my regular study.
   D. It has brought huge psychological pressure.

3. Do you have any plan for your English study in senior 3?
   A. No, I don’t. I follow my English teacher.
   B. I have plans for some aspects of English learning (e.g., grammar).
   C. I have general study plans for all aspects of English learning.
   D. I have detailed plans for English study.

4. Which type of English learning materials do you usually adopt?
   A. Textbooks.
   B. Mock NMET test papers.
   C. Teaching-tutorial materials.
   D. Extracurricular readings (e.g., magazines).

5. What are the contents that are usually learned in your English class?
   A. Practicing those Mock NMET test papers.
B. Learning textbooks.
C. Learning grammar.
D. Training for each tested section in the NMET.

( ) 6. What do you want your teachers to emphasize in your English class?
A. Vocabulary and grammar.       B. Writing.
C. Reading comprehension.        D. Cloze.

( ) 7. How do you usually improve your English reading in senior 3?
A. By doing Reading comprehension practice.
B. By memorizing English words.
C. By reading English magazines and novels.
D. By paying attention to the genre and structure of English passages.

( ) 8. How do you usually improve your English listening in senior 3?
A. By watching English movies and televisions.
B. By listening to English radio.
C. By doing Mock NMET test papers.
D. By using listening textbooks or listening-tutorial materials.

( ) 9. How do you usually improve your English writing in senior 3?
A. By writing Mock NMET topics or other types of topics.
B. By memorizing model essays in the Mock NMET papers.
C. By memorizing all kinds of composition templates including their beginning and end.
D. By writing e-mails; daily or weekly diaries in English.

( ) 10. How do you usually improve your English speaking in senior 3?
A. By practicing conversation with others (e.g., peers, friends or parents).
B. By persistently reading aloud English texts.
C. By actively completing classroom tasks assigned by teachers in English.
D. By scarcely focusing on speaking as it is not tested by the NMET.

( ) 11. How do you usually finish your English practice in senior 3?
A. I only focus on accuracy.
B. I not only finish the practice but figure out the answers.
C. After I finish the practice, I usually review knowledge related to it.
D. I usually categorize all my errors.

( ) 12. What is essential for you when you review for your studies?
A. Contents that were mostly tested by the NMET.
B. Contents that I am not familiar with.
C. Contents listed on the testing syllabi.
D. Those problematic and relevant contents listed in the textbooks.

( ) 13. What test task do you think is the most important when you review?
A. The one that I am most interested in.
B. The one that I score most easily in the NMET.
C. The one that I have the most difficulty with.
D. The one that weighs the highest in the NMET.

( ) 14. What type of task in a test do you spend most of your time on?
A. Gap filling.       B. Reading comprehension. C. Writing.       D. Cloze.

( ) 15. How much burden are you experiencing from your study in senior 3?
A. No burden at all.       B. Well to handle the burden.
C. Average burden level.   D. High burden level.

( ) 16. How many English classes do you take every day?
A. One class       B. Two classes       C. Three classes D. Four and above
17. How much time do you spend on your daily English homework?
   A. Half to an hour.   B. One to two hours.
   C. Two to three hours.   D. More than three hours.

18. How much time do you spend on extra English learning except for homework?
   A. Less than half an hour.   B. Half to an hour.
   C. One to two hours.   D. More than two hours.

19. What is the difficulty of Mock NMET tests you did?
   A. Quite simple.   B. Average difficult.
   C. Quite difficult.   D. Very difficult.

20. Compared to other subjects, how much burden are you experiencing from English?
   A. Not much burden.   B. Average burden.
   C. Quite a lot of burden.   D. Much burden.

21. How would you describe your daily English class?
   A. Quite relaxing.   B. No special feeling.

22. Which type of English homework do you spend most of your time on?
   A. Reciting (e.g., words in textbooks).
   B. Copying (e.g., words and sentences in textbooks).
   C. Doing an exercise book.
   D. Doing all types of test papers.

23. Have you hired any English tutor at home?
   A. I have been consistently hiring an English tutor at home.
   B. I used to have an English tutor, but now I don’t.
   C. I don’t have an English tutor now, but I plan to have one.
   D. I never plan to hire an English tutor.

24. Do you take any English tutoring class on the weekend?
   A. I don’t take any English tutoring class.
   B. I spend half a day on it.
   C. I spend a whole day on it.
   D. I spend the whole weekend on it.

25. Including classroom tests, on average, how many weekly English tests do you have?
   A. Barely any.   B. One to two tests.
   C. Three tests.   D. Four or above.

26. Do you feel stressed about the weekly or monthly English tests at school?
   A. I can handle them well.   B. I don’t feel much stress.
   C. I feel some stress.   D. I feel a lot of stress.

27. How many exercise books or Mock NMET test papers do you have?
   A. Three or below.   B. Four to six.
   C. Seven to nine.   D. More than nine.

28. How many of those exercise books are chosen by yourself?
   A. Three or below.   B. Four to six.
   C. Seven to nine.   D. More than nine.

29. What is the most significant contributing factor to the English learning burden? (If you do not feel any burden, please consider your peers).

30. How does failure in one English test influence you?
A. No influence and I don’t care.
B. I feel myself so useless and lose interest in English.
C. I suffer from mood swings and need time to adjust myself.
D. It pushes me to study harder.

The second part: multiple choice questions (20 question items)

Please read carefully and choose only one answer based on your real situation
A. Strongly agree
B. Agree
C. Slightly Agree
D. Disagree
E. Strongly Disagree

31). I agree that NMET should be held only once. Twice makes me very nervous.
32). My anxiety will be reduced as there is more than one chance to participate in NMET.
33). I would continue to study for the second NMET to get a better score even though I perform well in the first.
34). My brain always feels so tight ever since I enter Grade 3.
35). Getting a good score on one test doesn’t ease my worry towards a subsequent test.
36). I am afraid my English teacher will consider me incompetent if I fail the test.
37). I am afraid my peers might laugh at me when I fail the test.
38). The closer I get to an English exam, the more anxious I feel because I don’t feel adequately prepared.
39). Before an English exam, I always feel that my peers prepare better than me.
40). Before an English exam, I sweat and tremble.
41). Before an English exam, I always feel nervous and couldn’t sleep well at nights.
42). Before an English exam, I stay up late to review the knowledge that may be tested.
43). I start to be tense once I get the test paper.
44). During an English exam, I’m sometimes so tense that I forget to complete the demographic information at first.
45). During an English exam, I always worry about my listening equipment not working for the listening section.
46). During an English exam, I find myself thinking of things unrelated to the test.
47). During an English exam, I frequently get so nervous that I forget about facts I know.
48). As soon as an English exam is over, I try to stop worrying about it but I just can’t.
49). As soon as an English exam is over, I always worry about potential incidents with my answer sheet may.
50). I always panic when I have to take a mock English exam.

-------- The survey is over and thank you for your cooperation --------
Appendix C Ethics Approval Level

August 24, 2018

Xing Zheng

Master’s Student
Faculty of Education
Queen's University
Duncan McArthur Hall 511 Union Street West
Kingston, ON, K7M 5R7
GREB Ref #: GEDUC-911-18; TRAQ # 6024425

Title: "GEDUC-911-18 Exploring the washback effects of the Chinese National Matriculation English Test (NMET) on student learning and test anxiety”

Dear Xing Zheng:

The General Research Ethics Board (GREB), by means of a delegated board review, has cleared your proposal entitled "GEDUC-911-18 Exploring the washback effects of the Chinese National Matriculation English Test (NMET) on student learning and test anxiety” for ethical compliance with the Tri-Council Guidelines (TCPS 2 (2014)) and Queen's ethics policies. In accordance with the Tri-Council Guidelines (Article 6.14) and Standard Operating Procedures (405.001), 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, Ms. Gail Irving, at University Research Services for further review and clearance by the GREB or Chair, GREB.

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

Dean Tripp, Ph.D. Chair
General Research Ethics Board

c: Dr. Liying Cheng, Supervisor
Dr. Benjamin Bolden, Chair, Unit REB
Mrs. Erin Rennie, Dept. Admin.
### Appendix D Outline for Interview (In Chinese)

<table>
<thead>
<tr>
<th>采访目的</th>
<th>为了进一步了解一年两考改革对高三英语教学和学生学习负担，考试焦虑的影响</th>
</tr>
</thead>
<tbody>
<tr>
<td>采访时间</td>
<td>2016 年 10 月</td>
</tr>
<tr>
<td>采访地点</td>
<td>浙江杭州</td>
</tr>
<tr>
<td>采访对象</td>
<td>两位高考专家和三位英语教师</td>
</tr>
<tr>
<td>采访工具</td>
<td>录音笔</td>
</tr>
<tr>
<td>采访问题</td>
<td>1. 你认为一年两考这个改革政策对你工作计划有什么影响呢？</td>
</tr>
<tr>
<td></td>
<td>2. 你认为一年两考这个改革政策对你教学内容有什么影响呢？</td>
</tr>
<tr>
<td></td>
<td>3. 你认为一年两考这个改革政策对你的技能有什么影响呢？</td>
</tr>
<tr>
<td></td>
<td>4. 你认为一年两考这个改革政策对你的教学方法有什么影响呢？</td>
</tr>
<tr>
<td></td>
<td>5. 你认为一年两考这个改革政策对你的教学负担有什么影响呢？</td>
</tr>
<tr>
<td></td>
<td>6. 你认为一年两考对学生的学习方面有什么影响呢？</td>
</tr>
<tr>
<td></td>
<td>7. 你认为一年两考是增负还是减负呢？</td>
</tr>
<tr>
<td></td>
<td>8. 你觉得造成这些影响的原因是什么呢？</td>
</tr>
</tbody>
</table>
# Appendix E Outline for Interview (In English)

<table>
<thead>
<tr>
<th>Purpose</th>
<th>For a better understanding of NMET (ZJ) reform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>In October, 2016</td>
</tr>
<tr>
<td>Location</td>
<td>Hangzhou, Zhejiang province</td>
</tr>
<tr>
<td>Participants</td>
<td>Two NMET experts and three English teachers from Hangzhou</td>
</tr>
<tr>
<td>Instrument</td>
<td>A recorder</td>
</tr>
<tr>
<td>Open-ended questions</td>
<td>1. Has the reformed influenced your teaching plan?</td>
</tr>
<tr>
<td></td>
<td>2. Has the reform influenced your teaching content?</td>
</tr>
<tr>
<td></td>
<td>3. Has the reform influenced your teaching skills?</td>
</tr>
<tr>
<td></td>
<td>4. Has the reform influenced your teaching methods?</td>
</tr>
<tr>
<td></td>
<td>5. Has the reform brought you any teaching burden?</td>
</tr>
<tr>
<td></td>
<td>6. Has the reform influenced student learning?</td>
</tr>
<tr>
<td></td>
<td>7. Has the reform increased learning burden or decreased it?</td>
</tr>
<tr>
<td></td>
<td>8. What are the reasons behind the influence?</td>
</tr>
</tbody>
</table>