

8-6-2014

Washback Effects of the Reformed CET-4 on College English Teaching and Learning in China: Students' Perspectives

Zhiling Wu

Indiana University of Pennsylvania

Follow this and additional works at: <http://knowledge.library.iup.edu/etd>

Recommended Citation

Wu, Zhiling, "Washback Effects of the Reformed CET-4 on College English Teaching and Learning in China: Students' Perspectives" (2014). *Theses and Dissertations (All)*. 679.
<http://knowledge.library.iup.edu/etd/679>

This Dissertation is brought to you for free and open access by Knowledge Repository @ IUP. It has been accepted for inclusion in Theses and Dissertations (All) by an authorized administrator of Knowledge Repository @ IUP. For more information, please contact cclouser@iup.edu, sara.parme@iup.edu.

WASHBACK EFFECTS OF THE REFORMED CET-4 ON COLLEGE ENGLISH
TEACHING AND LEARNING IN CHINA: STUDENTS' PERSPECTIVES

A Dissertation

Submitted to the School of Graduate Studies and Research

in Partial Fulfillment of the

Requirements for the Degree

Doctor of Philosophy

Zhiling Wu

Indiana University of Pennsylvania

May 2014

© 2014 by Zhiling Wu

All Rights Reserved

Indiana University of Pennsylvania
School of Graduate Studies and Research
Department of English

We hereby approve the dissertation of

Zhiling Wu

Candidate for the degree of Doctor of Philosophy

Michael M. Williamson, Ph.D.
Professor of English, Advisor

Lilia P. Savova, Ph.D.
Professor of English

Virginia A. Perdue, D.A.
Associate Professor of English

Jeannine M. Fontaine, Ph.D.
Associate Professor Emeritus of English

ACCEPTED

Timothy P. Mack, Ph.D.
Dean
School of Graduate Studies and Research

Title: Washback Effects of the Reformed CET-4 on College English Teaching and Learning in China: Students' Perspectives

Author: Zhiling Wu

Dissertation Chair: Dr. Michael M. Williamson

Dissertation Committee Members: Dr. Lilia P. Savova
Dr. Virginia A. Purdue
Dr. Jeannine M. Fontaine

The influence exerted by language tests on language teaching and learning has been called washback (Pearson, 1988). Although an increasing number of empirical washback studies have appeared after the call by Alderson and Wall (1993), the perspectives of the most immediate stakeholders, the test takers, are largely under-researched. Equally under-researched are the washback effects of language tests at the college level, since most reported research has been conducted on tests given before the tertiary level (e.g. Cheng, 2005; Manjarrés, 2005; Qi, 2004; Shohamy, 2001; Wall & Alderson, 1993; Watanabe, 2004). Hence the present study, from students' perspectives, examines how the reformed national College English Test Band 4 (CET-4) affects English teaching and learning at the college level in China. A four-phase mixed method is employed with 414 student participants from three Chinese universities. Suggestions are proposed to further promote a more effective English language assessment in China.

Keywords: washback effects, the reformed CET-4

ACKNOWLEDGMENTS

I am very grateful to Dr. Michael M. Williamson, my dissertation advisor, for his magnificent intellectual guidance, humanity, hospitality, humor and constant encouragement and support. Sometimes he believes in me more than I believe in myself. I am also grateful to Dr. Lilia Savova, Dr. Jeannine Fontaine, and Dr. Virginia Purdue, my committee members, for their patience, understanding and constructive advice along the way.

Special thanks go to the International Research Foundation (TIRF) for English Language Education 2010 Doctorate Dissertation Grant (DDG), which greatly relieved my financial burden by funding my field research. Of course, I was to express heartfelt thanks to my family and all my student participants and teachers, administrators and colleagues in the three universities in China who made my field research happen. In addition, I am indebted to Ms. Meiyen Chen and Ms. Yu Yi from the IUP Applied Research Lab, my mentors/friends - Dr. Zongmin Kang, Dr. Carolyn Princes, Mr. Yongming Luo, Mr. Ronald Hamilton, Mr. Adam Wolf, my sister - Fanglin Wu, and Ms. Julie Englemann and Ms. Vera Bonnet - my editors, etc. who either helped proofread the paper or offered me invaluable advice and support. I will definitely pay all your kindness forward!

TABLE OF CONTENTS

Chapter		Page
1	INTRODUCTION	1
	Statement of the Problem.....	2
	Questions to be Researched	4
	Methods of the Study	5
	Significance of the Study	7
2	LITERATURE REVIEW	9
	Washback or Backwash	9
	Washback Domains	12
	Washback Direction.....	14
	Washback Direction and Test Quality	17
	Inherent Limitations vs. Desirable Features of Language Testing	19
	Empirical Studies on Washback	21
	Washback and Centralized Educational Systems	23
	Washback and High-Stakes Testing	23
	Washback and Norm-Referenced Tests.....	25
	Washback, English Language Testing and EFL Contexts.....	26
	Washback and Commutative Language Teaching (CLT) and Communicative Language Competence	27
	CLT-Related Testing and Negative Washback.....	29
	Negative Washback and Other Variables	32
	Washback and the Teacher Variable.....	34
	Washback Studies and Often-Ignored Students' Perspectives	36
	Necessity of Washback Studies on Students' Perspectives	40
	Washback Mechanism/Models/Hypotheses	44
3	MACRO RESEARCH CONTEXT AND EVOLUTION OF THE CET	49
	Macro Research Context: English Learning and Testing in China.....	49
	The College English Test (CET).....	52
	Old Versions of the CET-4 and Washback.....	55
	Reformed CET-4.....	63
4	METHODS	71
	Research Methods Used in Prior Washback Studies	71
	Methodological Framework.....	74
	Sites for the Study.....	80
	Questionnaire Construction	86

Chapter		Page
	Overall Procedures.....	92
	Phase I.....	95
	Phase II.....	97
	Phase III.....	100
	Phase IV and Brief Summary.....	102
5	DATA ANALYSIS.....	106
	Basic Characters About the Participants.....	106
	Research Question One.....	113
	A Short Summary of the Responses to Research Question One.....	127
	Research Question Two.....	127
	A Short Summary of the Responses to Research Question Two.....	138
	Research Question Three.....	139
	A Short Summary of the Responses to Research Question Three.....	169
6	CONCLUSIONS, DISCUSSION AND SUGGESTIONS.....	170
	Conclusions.....	170
	Discussion and Suggestions.....	174
	Further Discussion and Suggestions of the Emerged CET-4 Washback Effects.....	187
	REFERENCES.....	208
	APPENDICES.....	222
	Appendix A – Recruitment Notice for the Questionnaire.....	222
	Appendix B – Letter of Information for the Questionnaire.....	224
	Appendix C – Informed Consent Form.....	226
	Appendix D – Survey Regarding the Influence of the Reformed National College English Test 4 (CET-4) on Your college English Teaching and Learning (English Version Distributed in Phase I).....	228
	Appendix E – Survey Regarding the Influence of the Reformed National College English Test 4 (CET-4) on Your College English Teaching and Learning (Chinese Version Distributed in Phase I).....	233
	Appendix F – Major Interview Questions in Phase II.....	237

Appendix G – Questionnaire Regarding the Influence of the Reformed National College English Test 4 (CET-4) on Your College English Teaching and Learning (English Version Distributed in Phase III).....	239
Appendix H – Questionnaire Regarding the Influence of the Reformed National College English Test 4 (CET-4) on Your College English Teaching and Learning (Chinese Version Distributed in Phase III).....	244
Appendix I – Major Interview Questions for Further Follow-up in Phase IV	249
Appendix J – An Overall Timeline of the Study	251

LIST OF TABLES

Table	Page
1 CET-4 Before January 2007 (150 Minutes).....	68
2 CET-4 After January 2007 (125 Minutes).....	69
3 Participants in Phase I and Phase II	99
4 Participants in Phase III and IV in University A	104
5 Participants in Phase III and IV in University A, B and C: A Further Follow-Up of Participants in Phase II.....	105
6 Participants from Phase I to Phase IV.....	105
7 Age.....	106
8 Gender.....	107
9 Birthplace.....	107
10 English Study Lengths	109
11 Matriculation Statistics of English Exam Grades in the Three Universities Respectively	110
12 Perception of Closeness Between one's Major and English.....	115
13 Test Statistics ^{a,b} Regarding Time Spent Daily on Studying English After Class by Students at the Three Universities.....	118
14 Collective Student Opinion of the Influence and Relevance of the Reformed CET-4 on Students' English Learning.....	140
15 What is the Respective Student Opinion in Regards to the Relevance of the Reformed CET-4 Influencing Students' English Learning?.....	141
16 What is the Collective Student Opinion in Regards to the Relevance of the Reformed CET-4 Influencing Teachers' English Teaching in the Three Universities?	141
17 What is the Respective Student Opinion in Regards to the Relevance of the Reformed CET-4 Influencing Teachers' English Teaching?.....	142

Table	Page
18 What is the Collective Student Opinion in Regards to the Influence of the Reformed CET-4 on Students' Learning Content?.....	142
19 What is the Collective Student Opinion in Regards to the Influence of the Reformed CET-4 on Teachers' Teaching Content?.....	142
20 What is the Collective Student Opinion in Regards to the Influence of the Reformed CET-4 on Teachers' Teaching Methods?	143
21 Frequency of Communicative Activities that were Conducted in English During Students' English Classes at College.....	143
22 Frequency of Using Multi-Media Assisted Teaching in English Classes.....	145
23 Frequency of Using Multi-Media in English Classes	145
24 What is the Collective Student Opinion in Regards to the Influence of the Reformed CET-4 on Students' Learning Methods/Strategies?	147
25 What is the Perspective Student Opinion in Regards to the Influence of the Reformed CET-4 on Students' Learning Methods/Strategies?	147
26 Percentage of Listening that was Influenced in Students' Learning.....	148
27 Percentage of Vocabulary that was Influenced in Students' Learning.....	149
28 Percentage of Writing that was Influenced in Students' Learning	149
29 Percentage of Listening that was Influenced in Teachers' Teaching	149
30 Percentage of Writing that was Influenced in Teachers' Teaching	149
31 Percentage of Reading that was Influenced in Teachers' Teaching	150
32 Frequency of Practicing English Skills out of Class at University A	152
33 Frequency of Practicing English Skills out of Class at University B	152

Table	Page
34	Frequency of Practicing English Skills out of Class at University C152
35	Phase I and Phase II251
36	Phase III and IV in University A252
37	Phase III and IV: A Further Follow-up of Participants in Phase II252

LIST OF FIGURES

Figure		Page
1	Primary Purpose of English Learning at College	111
2	College Graduation Plans	112
3	CET-4 and CET-6, Which Test is More Important?.....	113
4	Hours Spent Daily on Studying English After Class at College.....	118
5	Frequency of Practicing Different English Skills at University A	154
6	Frequency of Practicing Different English Skills at University B.....	154
7	Frequency of Practicing Different English Skills at University C.....	155

CHAPTER 1

INTRODUCTION

Examinations have been a crucial means for upward mobility throughout Chinese history (Cheng, 2009; Hu, 1984). In the 21st century Information Age, English has been recognized worldwide as an international language of commerce, politics, and culture (Crystal, 2003; Ha, 2008; McKay, 2002). Accordingly, English as a subject along with English examinations have received more and more attention, not only in the context of English learned as second language (ESL), but in English learned as foreign language (EFL). In China, English testing is now required at different educational levels, beginning with primary school (Cheng, 2008). According to the Second International Conference on English Language Teaching in Shanghai in 2006, more than 300 million Chinese were estimated to be learning English nationwide and taking English tests for purposes ranging from education to employment, promotion to entertainment (“China has,” 2006). More than 100 million of these students are registered in classes from the elementary to the college level. Of these English learners in China, 27 million are university students (Cheng & Curtis, 2009). The National College English Test Band-4 (CET-4) is the only national test for non-English majors at the college level, drawing the largest number of test takers in the world (Jin & Yang, 2006). In 2006, for example, more than 13 million students took the CET-4 (Zheng & Cheng, 2008).

Since its inception in 1987, the CET-4 has made an undeniable contribution to promoting English teaching and learning at the college level in China, despite some negative influences and misuse of test scores. In 2005, as part of the national Higher Education Undergraduate Level Teaching Quality and Teaching Reform Project initiated

by the Chinese Ministry of Education (MoE) and Ministry of Finance (MoF), MoE started a reform of college English teaching. The ministry cited “the pressing social need for college and university graduates with a stronger communicative competence in English” (Jin & Yang, p. 21). Reform of the CET-4 was considered essential because the CET-4 is the only English assessment for all non-English majors at the college level. It was expected that the reformed CET-4 would “maximize its positive backwash effect on teaching and beneficial impact on society” (Jin & Yang, p. 34). In other words, the reformed CET-4 is to help equip college students with more English communicative competence through promoting more of its positive washback effects on college English teaching and learning in China.

Statement of the Problem

Though the influence exerted by language tests on language teaching and learning has been called *washback* in the field of language testing (Pearson, 1988), it is also referred to as *backwash* by scholars such as Spolsky (1994) and Biggs (1995). In general, washback is the term currently preferred in professional publications, although the meaning of the term, washback is still being debated (Bachman & Palmer, 1996; Bailey, 1996; Messick, 1996; Spolsky, 1994). In this dissertation, I adopt the definition most often cited in the literature, in which washback is defined as “the effect of testing on teaching and learning” (Hughes, 2003, p. 1).

Given that there are not many empirical studies on washback, Alderson and Wall (1993) made an appeal and developed 15 important hypotheses to direct future washback research. Of the limited number of empirical studies that have appeared on washback so far, the majority focus primarily on teacher’s perspectives and how their English teaching

has been influenced by high-stakes tests, both in terms of what they teach and how they teach, (e.g. Manjarrés, 2005; Wall & Alderson, 1993; Watanabe, 1996). In contrast, the perspectives of the test takers, who are the most immediate stakeholders, have been largely ignored. Shohamy (2001) has pointed out that test writers are not particularly interested in the test takers' voice, for "in the testing literature test takers are often kept silent; their personal experiences are not heard or shared" (p. 97). Fullan (2007) also lamented that in educational change, the students were rarely considered "participants" but were rather considered "potential beneficiaries of change" (p. 15). This kind of bias leaves seven of the 15 washback hypotheses developed by Alderson and Wall (1993) under-researched. The hypotheses specifically related to language learning are:

Hypothesis 5: A test will influence what learners learn.

Hypothesis 6: A test will influence how learners learn.

Hypothesis 8: A test will influence the rate and sequence of learning.

Hypothesis 10: A test will influence the degree and depth of learning.

Hypothesis 11: A test will influence attitudes to the content, method, etc., of teaching and learning.

Hypothesis 14: Tests will have washback on all learners and teachers.

Hypothesis 15: Tests will have washback effects for some learners and some teachers, but not for others. (pp. 120 –121)

Since most of the published studies focus on tests administered at the middle school or high school level (e.g. Andrews, Fullilove & Wong, 2002; Cheng, 2005; Luxia, 2005; Shohamy, 2001; Wall & Alderson, 1993), the washback of high-stakes, national English tests at the college level is similarly under-researched.

Jin (2006) emphasized that “the real challenge facing the National College English Testing Committee (NCETC) is to improve the test’s washback and impact on English language teaching and learning at the tertiary level in China” (p. 68). Since its official implementation nationwide in June 2007, the reformed version of CET-4 has been administered six times by 2009. How college students perceive the reformed CET-4, the current state of college English teaching and learning, and the extent to which the reformed CET-4 has exerted a positive influence on English teaching and learning in China are the subjects of this study.

Questions to be Researched

As stated above, washback research has not focused on what test takers think and do. Existing washback literature often tries to establish the connection between examinations and the perceptions and/or behaviors of the various stakeholders, especially teachers (e.g. Alderson & Wall, 1993; Cheng, 2005; Hughes, 1993; Pearson, 1988; Spratt, 2005; Watanabe, 1996). This study, therefore, investigates the washback effects of the reformed CET-4 from the perspectives of the student. Specifically, the major questions researched in this study addressed students’ perspectives as follows:

1. What are college students’ beliefs about English learning; what are their expectations and experiences of studying English at college, and what are their perceptions of the reformed CET-4?
2. What plans did the students make to prepare for taking the CET-4, and how did they implement their plans?
3. To what extent has the reformed CET-4 influenced students’ perceptions of the practices of English teaching and learning at the college level?

In order to examine these questions, the following related questions were initially considered:

1. What are the learning philosophies of contemporary Chinese college students and what are their beliefs about learning English?
2. What prior English learning experiences have they had, and how did they compare to their current college-level experiences?
3. What are their current goals and future plans in terms of English learning?

Because it was not possible to include every college or university in China, study was conducted at three universities, all located in Shanghai, the biggest city in China. All three universities specialize on science and engineering, but they represent three different rankings in the university system: highly reputed, intermediate, and ordinary. With the reform of the CET-4 in 2007, each university executed its own practice plan for college English teaching, CET-4 test requirements, and testing preparation. The CET-4 is issued twice annually, in June and December. Because the MoE allows each university to decide whether their students should take the CET-4, in their first or second year and/or in June and December, the universities have different policies about when the majority of students should take the test. This study includes students who planned to take the CET-4 in June or December 2010. Data collection took place from May 8th to December 20th, 2010. A four-phase mixed method was employed.

Methods of the Study

The study was conducted four phases. Phase I and III made use of surveys. Phase II and IV were follow-up qualitative studies, which involved primarily interviews, and emails, essay writing, online chat, phone calls, and self-recordings. Phase I took place

from May 8, 2010 to May 18, 2010 and Phase II ran from May 19, 2010 to June 30, 2010. Phase III from November 1, 2010 to November 18, 2010. Finally, Phase IV lasted from November 19, 2010 to December 20, 2010. Descriptive statistics were used principally to analyze the questionnaire data, and Grounded Theory (Strauss & Corbin, 1998) was used to analyze the qualitative data.

In Phase I, surveys were conducted in three pre-scheduled classes in each of the three universities. In Phase II, ten students (eight planned to take the CET-4 in June and two in December) from each university were selected from among all of the students who had indicated their interest in participating in the follow-up interview at the end of the survey. The selection was based on gender, major, English classes registered, and English exam grades at matriculation. Since the majority of students expected to take the CET-4 in December in Phase III, a slightly revised survey was conducted in two classes in University A again in November and December. In Phase IV, four students in each class in Phase III were selected among the volunteers who expressed their willingness to participate further in the follow-up interview. What is more, the six students who were interviewed in Phase II, yet were going to take the CET-4 in December 2010 at this time were contacted to complete the survey and to be interviewed again. All 14 students in Phase IV were asked to keep journals or self-recordings about their CET-4 test-taking experience. Of course, not all participants stayed throughout the study. Altogether, 414 participants participated in the survey portion (5 participants were surveyed twice), 34 were interviewed (6 participants were interviewed twice), and 7 submitted their self-recordings at a later date.

The survey data were analyzed using descriptive statistics through SPSS 18. The

follow-up qualitative study data - individual interviews, students' self recordings, and/or emails regarding their English learning and preparation for the CET-4 - were transcribed and coded based on a Grounded Theory approach. When necessary, ambiguous information was further clarified by emails or phone calls to the participants concerned. The final results of the study were shared with the participants and the universities in which the four-phase studies were conducted.

Significance of the Study

Since there are only a limited number of empirical studies on the washback effects of high-stakes language tests, this empirical study contributes to a further understanding of how these tests impact English teaching and learning. Particularly, it contributes to the understanding of the 7 hypotheses related to washback effects on learning and learners developed by Wall and Alderson (1993).

In addition, this research is one of the few empirical studies that focuses on washback effects of high-stakes English tests at the college level from the perspective of the student. This focus is especially important because students are always the most immediate stakeholders in testing; hence, their voices should be heard and respected. What is more, as mentioned earlier, most of the previous washback studies focus on language tests given before the college level and primarily from teachers' perspectives (e.g. Manjarrés, 2005; Wall & Alderson, 1993; Watanabe, 1996). Compared with students at the secondary level, college students usually serve as the backbone of the future development of a society, and they have a clearer idea of their educational goals. Given the 27 million university students who are learning English and taking the CET-4 in China, and given the invaluable human resources, money and time that have been

invested by the MoE, it is certainly worth examining the washback effects of the reformed CET-4 to see whether it helps maintain students' enthusiasm for learning English, and more importantly, it meets the standards for communicative competence standard set for the college-level students in China.

Theoretically, evidence of washback is related to the consequential aspect of construct validation enquiry, according to the unified validity argument made by Messick (1996). In this sense, the findings in the current study will also contribute to the ongoing validation process of the reformed CET-4. In the increasingly globalized society of the 21st century, studies on maximizing positive washback effects of high-stakes English tests at the college level will also benefit students in broader EFL learning contexts.

CHAPTER 2

LITERATURE REVIEW

The influence exerted by language tests on language teaching and learning has been labeled as *washback* in the field of language testing (Pearson, 1988). However, even though the influence caused by tests was recognized soon after tests were administered, the term *washback* did not attract wide attention until the 1980's. Alderson (1993) once even cast his doubts on the existence of washback. Davies (1990) argued that “testing always has a ‘washback’ influence and it is foolish to pretend that it does not happen” (p. 24). Today, it has been generally acknowledged that “washback effect is an inherent quality of any kind of assessment” (Cheng, 2005, p. 26), but different definitions have been proposed as well as a “considerable variety in opinions as to how it functions” (Bailey, 1996, p. 257). Specifically, there is little agreement on the domains impacted by washback effects, or on how washback operates, or in what direction and to what extent it operates. (Alderson & Hamp-Lyons, 1996; Alderson & Wall, 1993; Bailey, 1996; Cheng, 2005; Spratt, 2005). The controversial opinions proposed by various linguists and educators, as well as salient theories generated from empirical studies based on aspects of washback literature, will be discussed in the following section.

Washback or Backwash

The initial usage of the terms *backwash* and *washback* has not been clearly documented, nor is it clear why the use of one term has been preferred to the other at different times. Yet the reason for figuratively combining *wash* and *back* to this reverse phenomenon is well illustrated by Pearson (1988). He stated that “public examinations influence the attitudes, behaviors and motivations of teachers, learners and parents, and

because examinations often come at the end of a course, this influence is seen working in a backward direction, hence the term ‘washback’” (p. 98). According to Alderson and Wall (1993), *washback* is the more accepted term among British applied linguists while *backwash* is more commonly used by educators. This, in fact, is not quite the case, because *backwash* is preferred, for instance, by Hughes (1989, 2003) and Biggs (1995, 1996).

One of the leading British scholars in language testing circles, Hughes uses *backwash* in both editions of his best seller – Testing for Language Teachers. In a footnote, he even points out that he is not familiar with the origin of *washback*, nor could he find a dictionary entry for the term. His claim that only *backwash* could be found in comprehensive dictionaries is also supported by Cheng (2005). In the field of general education, *measurement-driven instruction* (Frederickson, 1984; Popham, 1983, 1987; Bracey, 1987) appears to be used more frequently than *washback*. Shohamy (1992) asserted that washback effects and measurement-driven instruction refer to the same phenomenon. Spolsky (1994) even contended that “backwash is better applied only to accidental side effects of examinations, not to those effects intended when the first purpose of the examination is control of the curriculum” (p. 55).

At one time *backwash* was the more commonly used term. However, *washback* is the term that has currently gained prominence in language testing and applied linguistics circles and it is now generally agreed that *washback* and *backwash* are synonyms (Cheng, 2005; Davies, 1990; Green, 2007). As to the definition of *washback*, multiple interpretations have been proposed from various angles: Alderson and Hamp-Lyons (1996) claimed that “‘washback’ (sometimes also known as ‘backwash’) is the

influence that writers on language testing, syllabus design and language teaching believe a test will have on the teaching that precedes it” (p. 280). Bachman (1990) described it as “the effect of testing on instruction” (p. 283) and Wall and Alderson (1993) defined *washback* as “the impact of a test on teaching” (p. 41). This emphasis on the effects of washback on teaching is similarly shared by Richards, Platt, and Platt (1992). As simple as these definitions appear, the emphasis on teaching in so many seminal studies, has greatly influenced the course of washback research. As illustrated later in this literature review, research on the influence of washback on other stakeholders – notably learners – has largely been ignored.

On the other hand, *washback* has been described as “the connections between testing and learning” by Shohamy, Donitsa-Schmidt, and Ferman (1996, p. 298). Although they do not explain exactly what these connections are, their understanding differs from the previous researchers’ in that they do not ignore the influence of testing on learning. Messick (1996) vividly depicted *washback* as “the extent to which the introduction and use of a test influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning” (p. 241). While Messick implicitly discusses both intended and non-intended effects of testing, Cheng (2005) used *washback* only to “indicate an intended direction and function of curriculum change on aspects of teaching and learning by means of a change of public examinations” (p. 28). She uses the term *backwash* for unintended effects.

Without considering the nature of the different effects that can be produced by testing, Hughes (1989; 2003) defined *washback* as “the effect of testing on teaching and learning” (p. 1). Likewise, Gates (1995) stated that *washback* is “the influence of testing

on teaching and learning” (p. 101). Davies (1990) concurred with Hughes and Gates’ definition, and further pointed out that language testing “provides a methodology for experiment and investigation in both language teaching and language learning/acquisition” (p. 1). After a brief review of washback studies, Bailey (1996) also concluded that “washback is generally defined as the influence of testing on teaching and learning” (p. 259). This dissertation uses the term *washback* as it refers to testing effects on both teaching and learning, in the belief that teaching and learning are interrelated yet separate issues.

Washback Domains

The literature exhibits even greater differences when referring to the domains to which washback effects may extend. Buck (1988) limited it to “the influence of the test on the classroom” (p. 17). Alderson and Wall (1993) hypothesized about the effects of testing on content, methodology, rate, degree, attitude, etc. in language teaching and learning. Operatively, Hughes (1993) suggested that washback involve three core components: participants, process and products. He explained that participants included administrators, material developers, publishers, students and teachers. Further, washback effects could be found in the participants’ perceptions and behaviors, as well as in their learning outcomes. Bailey (1996), however, has countered that not all the participants’ processes lead directly to learning as Hughes suggests. She categorizes students separately from the other participants, and proposes two types of washback: washback to the learners and washback to the program.

Besides examining testing as it relates to educational activities, some researchers take a much broader view of the domains influenced by washback. For instance,

Rea-Dickins (1997) extended it to government, the workplace and the marketplace.

Taylor (2000), the Performance Testing Co-coordinator of the University of Cambridge Local Examinations Syndicate (UCLES), compiled an even more comprehensive list of professions that may be influenced by testing: learners, teachers, school administrators, test writers, consultants, examiners/raters, test-center administrators, material writers, publishers, receiving institutions, government agencies, professionals, employers and academic researchers. Taylor argued that because they were more or less influenced by testing, these professionals always showed an interest in different segments of the testing procedure. This claim has certainly been acknowledged in the assessment field.

In addition to research on the washback effects on teaching and learning, there have been sporadic studies of effects on some of the above – mentioned domains, for example: studies have been done on teaching materials (Saville & Hawkey, 2004), textbook writers (Hughes, 1993), textbook publishers and testing agencies (Hamp-Lyons, 1997), curriculum innovation (Andrews, 1994), administrators and test developers (Luxia, 2005), parents (Shih, 2007; Pearson, 1988), broad social dimensions (McNamara, 2006), and ethical use of tests by different stake holders involved (Shohamy, 1993; Spolsky, 1994), to name a few.

Acknowledging the broader aspects of washback found in the literature, Bachman and Palmer (1996) made a further distinction, differentiating between *washback* and *impact*. According to them, the effect of *washback*, defined as the influence of tests on teaching and learning, occurs at a micro level. In this framework, *impact* refers to the effects on individuals, policies and practices beyond classrooms and on the educational system and society at large. Therefore, *washback* is merely “an aspect of impact” (p. 30).

In line with this distinction, Shohamy (2001) summarized the use and abuse of the power of tests, and proposes the “need to develop critical strategies to examine the use and consequences of tests, to monitor their power, minimize their detrimental force, reveal the misuse, and empower the test takers” (p. 131). In like manner, McNamara and Roever (2006) advocated looking into the social dimension of language testing.

The author agrees, on the one hand, that the distinction between *washback* and *impact* has made a significant contribution to a more holistic understanding of language testing and its effects. On the other hand, the interwoven influences of testing on the macro and micro levels make it difficult to maintain clarity between these distinctions in practice. The broad scope of the social dimensions impacted by testing further complicates the implementation of an empirical study. As a result, this research will make use of the terms *washback* and *impact* interchangeably, as suggested by Andrews, Fullilove and Wong (2002) and Manjarrés (2005). Since it is hard to refer to one level without inferring the other, this paper will focus on the influence of testing on teaching and learning in school contexts, but will unavoidably touch upon the influence of testing on society as well.

Washback Direction

The direction that washback effects work is another issue debated heatedly. *Backwash* as a term seems to be interpreted with a more negative tone than *washback*. This was true even before *backwash* became a term circulated in language testing or in the discourse of educational measurement. The New Webster Comprehensive Dictionary of the English Language defined *backwash* in 2002 as “the unwelcomed repercussion of some social action”. The term meant “unpleasant after-effects of an event or situation”

according to Collins Cobuild Dictionary of English Language in 1995 (Cheng, 2005). Davies (1990) mentioned that the term was “usually felt to be wholly negative” (p. 1) when describing the testing effects on teaching and learning. As stated earlier, Spolsky (1994) believed that backwash represented the negative side effects caused by testing. Likewise, the synonym *measurement-driven instruction* is constantly criticized for its negative connotation, which will be illustrated later.

In fact, the cry to prevent negative washback once became so strong that Morrow (1986) suggested the concept washback validity. Washback validity is the validity of a test biased on the extent to which it meets the demands of intended beneficial washback. The case for drawing more positive attention to washback is understandable, but its extreme nature has never been widely endorsed in professional measurement circles. This point will be further illustrated when the complications of the test validation process are discussed.

As regards to washback and the direction of the effect, it is currently strongly acknowledged that so long as there is a test, there will be washback effects, both intended and unintended. Studies in different contexts cited below also verify this point empirically. What then constitutes intended positive washback, and what constitutes unintended negative washback? Pearson (1988) believed that good tests would produce good washback effects. Wall (2000) contested that it was too simple an assumption to believe that a *good* test would produce positive washback. Messick (1996) argued that “a poor test may be associated with positive effects and a good test with negative effects because of other things that are done or not done in the educational system. Technically speaking, such effects should not be viewed as test washback but rather as owing to good

or bad educational practices apart from the quality of the test” (p. 242). Let’s consider the example of a writing test. One might assume that a writing test that uses only multiple-choice questions results in negative washback, whereas a writing test with direct writing tasks leads to positive washback. Linn, Baker and Dunbar (1991), however, found that washback from testing on direct writing is not necessarily positive. Teachers employed a formulaic approach to writing proved to be a key reason for score improvement, yet the researchers asserted that actual writing ability had not improved. As a result, Alderson and Wall (1993) suspected that “the quality of washback might be independent of the quality of the test” (p. 18), because any test can foster participants’ motivation and aggravate anxiety at the same time. Shohamy (2000) seemed to agree, and advanced the hypothesis that “good or bad, [a test] will have beneficial washback, if it increases learning activities and intentions, making teachers and learners work harder” (p. 48).

Ideally, a genuinely perfect test should produce exclusively intended washback while unintended washback should be produced only by an imperfect test. The problem is, does a genuinely perfect test exist? Needless to say, every test is designed from a specific perspective by subject test writers, and it is used by subjective test-takers, employers, teachers, parents, etc. So without doubt, the answer to this question is that there is no perfect test. The reasons will be further explained when inherent limitations of language tests are addressed.

In short, I believe that as long as there is washback, there will be both intended and unintended effects, since nobody lives in a social vacuum. As Bailey (1996) noted, “washback can either be positive or negative to the extent it either promotes or impedes the accomplishment of educational goals held by learners and/or program personnel” (p.

269). With the empirical studies cited later, I would further argue that it is not possible for a test to yield only positive washback and no negative washback at all, or vice versa. Like medicine designed to treat patients, every test has unintended side effects. Unintended negative washback cannot be eliminated completely. Of course, efforts can be made to maximize the positive effects and minimize negative ones.

On the other hand, although a ‘good’ test will not produce intended positive washback exclusively, I do not agree that the direction of washback is independent of test quality. On the contrary, I believe that the desirability of washback effects is closely connected to the quality of a test, even though there is no perfect test. Rather, a high quality test is expected to generate more positive washback and less negative washback, other variables being equal. A poor quality test, however, is expected to generate more negative washback than positive washback. A review of empirical studies will lend support to this assertion later.

Washback Direction and Test Quality

An obvious question then arises: what makes a high quality test? To answer this question, a review of some of the well-established theories in the field of educational measurement is in order. In the field of educational assessment it has been universally accepted that reliability and validity are two essential psychometric principles.

Reliability is mainly concerned with consistency of test performances: a valid test must first and foremost be reliable. Validity, which in its traditional sense verifies whether a test measures what it intends to measure, now indicates “the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores. Test validation has, thus, turned into the process of accumulating evidence to support such

inferences” (American Psychological Association, 1985, p. 9). In this sense, validity should not be considered an inherent quality of a test, but a value judgment. In other words, a test itself has no validity, but the inferences made through the test scores do.

Moreover, Messick (1988) argued for a unified validity, in which construct validity subsumes other types of validity, such as content validity, criterion validity, and predictive validity. According to Messick (1988), construct validity serves as the evidential basis for test interpretation and test use. Evidence of washback then is related to the consequential aspect of construct validity. This view provides a theoretical basis for the importance of washback studies in the field of testing. Operationally, Messick (1996) emphasized that construct under-representation and construct-irrelevant variance were two major threats to construct validity.

According to Messick, construct under-representation occurs when a test is so narrow that it fails to incorporate the representative components or aspects of the underlying abstract theories of the target language. Construct-irrelevant variance comes about when a test is so broad that it includes components or aspects that are not fundamentally related to the underlying theories of the target language. He suggested that the former jeopardize authenticity and the latter jeopardize directness. However, in either case, the test score obtained will fail to be a good indicator of test takers’ ability in the non-test domain of that language.

What is more, Messick has reiterated that washback is just one aspect of consequential validity. Consequential validity, like washback validity – the term proposed by Morrow (1986) above – is not sufficient to evaluate the validity of a test. Validation must be based only on unified validity as a whole, which consists of six

essential aspects of construct validity. These aspects are content, substantive, structural, generalizability, external and consequential. Moreover, according to Messick, only effects that can evidentially relate to the use of the test be counted as washback.

While agreeing with Messick on the above point, I would also argue that a test that produces positive washback is not assuredly valid, yet washback evidence is a necessary part of any test validation. Besides, other things being equal, the more a test satisfies criterion of unified validity, the more likely is it to generate significant positive washback. Below is a review of empirical studies which demonstrate how washback direction is influenced by test quality, particularly how unintended negative washback is closely connected to tests that violate construct-underrepresentation and/or construct irrelevant variance.

Inherent Limitations vs. Desirable Features of Language Testing

Of course, it is first necessary to review the inherent limitations pertaining to language testing, so that tests are not subject to unrealistic expectations, especially in the design of educational reforms. As to the limitations of testing, Bachman (1990) has pointed out that, like all measures of mental ability, language tests are invariably characterized as indirect, incomplete, imprecise, subjective and relative. In order to present a clear picture of testing, each of these limitations will be briefly explained.

First, language tests are all indirect in the sense that observable behaviors are used to evaluate unobservable, abstract language competence in tests. No technology so far can decipher precisely how the nerves in the brain work to enable human beings to perform various language functions. Test users can only make inferences about test takers' level of language ability based on the specific performance of test tasks. Even the

direct testing advocated by Messick (1996) is indirect by nature. However, theoretically, direct testing should produce more valid influences than indirect testing since the skills and criteria being tested bear greater resemblance to the characteristics of target language use. So, in a sense, direct testing can improve the construct validity of a test, but it cannot overcome the indirect nature of language testing in general.

Second, the small sample of tasks selected for evaluation in any language test can never be completely representative. Obviously, it is impossible to test all the bits of language knowledge possessed by any language user, given that human beings are capable of generating countless new sentences in different contexts. Neither is it practical to lengthen a test for the sake of completeness. To better infer the level of the test takers' language ability, the sample tasks are limited to only those clearly represent the core underlying construct of knowledge about language and language use.

Furthermore, sample test tasks included in any language tests can only be relatively accurate when assessing language competence. This does not mean that the tasks are inaccurate, but that there is no perfect match between the level of test task difficulty and the level of true language competence. It is hard to maintain internal consistency so that all test tasks are at the same level of difficulty. The pass/fail cutoff score on any test is not precise, either, since it is based on the subjective and finite selection of test task samples. Those who fail the test are not necessarily incompetent. Further, there is no guarantee that no other ability is involved, because a single test task often engages more than one language skill, as well as other more general skills. On the whole, test-taking is a subjective process leading to subjective conclusions.

Despite the inherent limitations described above, tests have many high-valued merits. As summarized by Shohamy (2001), a) tests are administered by powerful institutions; b) tests use the language of science; c) tests use the language of numbers; d) tests use written forms of communication; e) tests rely on documentation; f) tests use objective formats, etc. The public understands the power of these features and the extent to which they outweigh the limitations of testing. In reality, ever since examinations became a form of assessment and selection, they have trumped selection based on inheritance (Williamson, 1994) and “nepotism, favoritism and even outright corruption in the allocation of scarce opportunities” (Bray & Steward, 1998; Eckstein & Noah, 1992, as cited in Cheng & Curtis, 2004, p. 5). In fact, more often than not, test scores are considered to be “objective proofs” (Shohamy, p. 40) of learning outcomes, not only by the general public, but by administrators who want to prove their effectiveness as leaders. High test scores are always interpreted as being strong indicators of the skills of language users, teachers and administrators. Also, because of this widely-perceived symbolic power of the test as an indicator of success (Bourdieu & Thompson, 1991), language testing is now often used in top-down educational innovations, with high hopes of generating positive washback.

Empirical Studies on Washback

Even though washback effects have become a widely studied topic, a review of the language testing literature shows that few studies of washback have been conducted using empirical methods (Alderson & Wall, 1993; Bailey, 1996; Cheng, 2005). Moreover, most existing accounts are based on teachers’ self-reports, not demonstrated activities of what teachers or learners do. As noted above, Alderson and Wall (1993) had cast doubts

on whether washback exists. Based on a review of educational measurement and applied linguistics, as well as interviews with teachers about their experiences with testing, the authors conceived 15 hypotheses and called for more empirical studies on washback research. This frequently-cited article, serves as a cornerstone for more empirical washback studies in language testing.

Using explicit research directions incorporated into the hypotheses, a considerable number of researchers have conducted empirical studies on washback in response to this call (e.g. Alderson & Hamp-Lyons, 1996; Andrew, Fullilove & Wong, 2002; Cheng, 2005; Huhta, Kalaja, & Pitkanen-Huhta, 2006; Manjarrés, 2005; Qi, 2004; Shohamy, 2001; Wall & Alderson, 1993; Watanabe, 2004). As a result, According to Alderson (2004), enough empirical evidence has now been collected to demonstrate that washback does exist. Further, Hughes (2003) has made an appeal to use tests as more than simply a final step in the teaching process; instead he states, “we should demand of it that it is supportive of good teaching, and, where necessary, exerts a corrective influence in bad teaching” (p. 2).

More and more people are meeting the call to create the desired positive washback. After all, the aim of assessment should not be to simply measure. Rather it should aim to improve test takers’ target abilities (Wiggins, 1998). Of course, many aspects of the hypotheses introduced by Alderson and Wall (1993) still need further research. What follows is a review of the washback studies that have been referenced above, with emphasis on the common characteristics shared by these studies, and their contributions to the field, as well as some emerging issues that require greater attention.

Washback and Centralized Educational Systems

As indicated above, language testing has been widely used in educational practice worldwide. Foster (1992) believed that this was particularly true in developing countries, because “there are far fewer places available in the upper levels of education” where testing serves a gate-keeping function (as cited in Wall, 2000, p. 501). However, Shohamy (2001), has found that the practice of testing to be more salient in countries with centralized educational systems. The latter seems to better describe educational practice in the world, since developed countries such as Japan (Watanabe, 1996), the United States of America (Oneill, Murphy, Huot, & Williamson, 2006), Australia (Burrows, 1998) and Hungary (Huhta, Kalaja, & Pitkanen-Huhta, 2006) have all been reported to use high-stakes language tests to screen candidates for various purposes. With the exception of the United States, almost all countries have centralized educational systems, e.g. China (Qi, 2005), Colombia (Manjarrés, 2005), Sri-Lanka (Wall, 1996), Iran (Salehi, Mustapha & Yunus, n.d.), Israel (Shohamy, 1993) and so forth. Therefore, language testing is widely utilized in centralized educational systems worldwide. Washback, hence, is a common effect that exists in both developing countries and developed countries.

Washback and High-Stakes Testing

High-stakes testing is an important variable influencing washback effects. High-stakes testing usually refers to tests that have real life-changing effects on test takers (Cheng, 2005), because the tests determine the future educational and employment opportunities. Examples include the Instituto Colombiano para la Evaluación de la Educación (ICFES) exams for students who finish upper secondary school; the O-level

exams at the end of the 11th year for desirable higher education or employment in Sri-Lanka; the National Matriculation English Test (NMET) in China; the entrance examination to college in Japan; the Entrance Examination of Universities (EEOU) in Iran; an oral test for English as a foreign language at the 12th grade level, a component of the national matriculation examination in Israel; and the entrance foreign language entrance examination for college in Hungary.

Certification tests are also high-stakes. These include the certification exam in Spoken and Written English in Australia, and the Hong Kong Certificate of Education Examination (HKCEE). In addition, studies on Tests of English as a Foreign Language (TOEFL) (Alderson & Hamp-Lyons, 1996) and the International English Language Testing System (IELTS) (Green, 2007) are also on the rise. Unlike other tests, these two proficiency exams are language prerequisites mainly for non-native English speakers who apply for study in the United States, the United Kingdom, or other English-speaking countries.

Though administered in different countries, these tests are used to determine admission, graduation, employment, immigration, certification, personal advancement or reputation. Whether passing or failing, the results can be the turning point of a test taker's life. Sometimes "a formal trigger of consequences need not [even] be built into the testing program for the stakes to be high" (Corbet & Wilson, 1991, as cited in Champmen & Snyder, 2000, p. 458). In other words, a test may involve high-stakes if the perceived consequences are high, even if the test developers did not intend it. Madaus (1988) even believed that stakeholders' perceptions of the potential stakes were more important than the stakes intentionally built into a test. Furthermore, both built-in and perceived stakes

seem to play a part in generating washback. Like Shohamy, Donitsa-Schmidt, and Ferman (1996), the author of this paper also believes that washback effects can change over time as the different variables involved change. A change in either built-in stakes or perceived stakes will result in changes to the washback effects.

Washback and Norm-Referenced Tests

As to the frame of reference, Madaus (1988) stated that high-stakes tests could be norm-referenced or criterion-referenced, internally or externally administered achievement tests, or proficiency tests. Nonetheless, almost all the tests considered in washback publications are externally-given, norm-referenced achievement or proficiency tests. Norm-referenced tests are tests designed to interpret a test taker's score with reference to test scores of a group whose characteristics are similar to those of the test takers. The scores obtained in criterion-referenced tests, by contrast, are interpreted with reference to different stages of a pre-determined content domain (Bachman, 1990). In this case, test takers will know their own progress without comparing themselves to others, as in driving test. Given that the major purposes of testing are to infer the test takers' ability, and to compare individuals and select a small number from a large pool based on their test performance, it is not surprising that high-stakes tests are rarely criterion-based but norm-referenced. Davies (1990) even claimed that the criterion-referenced test was "more properly regarded as suitable as an exercise... it is one use of norm-referenced test" (p. 19).

Norm-referenced high-stakes tests do not enjoy a good reputation in the field of educational measurement. Due to the serious potential consequences perceived by teachers, students, parents, and other stakeholders, norm-referenced high-stakes testing

has been constantly criticized for distorting curricula, forcing students to cram, inducing students to train on test-taking strategies, and restricting teachers' autonomy to alter their instructional or evaluation methods. What is more, these tests have been widely reported to incur enormous pressure and anxiety while failing to improve learning or critical thinking (Herman & Golan, 1991; Madaus, 1988; Nobel & Smith, 1994; Shepard, 1990; Smith, 1991; Vernon, 1956).

On the other hand, despite all the accusations, high-stakes testing still remains the most popular way to introduce important educational innovations. Low-stakes testing, though useful in many respects and non-anxiety-producing, generally does not attract sufficient attention to create educational policy change (Shohamy, 2001). The desire to be successful and to have their success acknowledged by others (Popham, 1993) is probably another reason that stakeholders are motivated to continue participating in norm-referenced high-stakes tests of various kinds each year all over the world.

With the growing recognition of the potential negative influence of high stakes testing, there has been greater emphasis on redesigning tests to increase positive washback. In order to take advantage of the power of washback effects and generate more desired positive washback, a high correspondence between what is taught and what is to be tested should be achieved. Messick (1996) also stressed that “ideally, the move from learning exercises to test exercises should be seamless” (p. 241).

Washback, English Language Testing and EFL Contexts

Another common point shared among the above-mentioned language tests is that almost all are tests of English. What is more, in most cases English is learned as a foreign language in these contexts. Although the majority of these tests are just one component of

a larger battery of entrance examinations, a low score on English examinations can be so detrimental that test takers will have to give up their efforts to move forward (e.g. Cheng, 2005; Qi, 2004; Watanabe, 2004). In the meantime, few question the credibility of these tests. As Ross (2008) put it after examining the language assessment practices in Japan, Korea, China, Hong Kong, Thailand, and India, “a commonly accepted assumption is that a foreign language learned in the context of formal schooling yields suitable subject matter for making high-stakes inferences about qualifications for admissions or employment” (p. 8). Ross shows concern about the negative washback effects that arise with the assumption that a foreign language test can predict aptitude or achievement in these contexts. Although it is beyond the scope of this dissertation to discuss whether high-stakes English testing can provide “an equitable basis for the estimation of scholastic merit” (p. 8), it is still worth pausing to think about the legitimacy of this practice.

Washback, Communicative Language Teaching (CLT) and Communicative Language Competence

Bailey (1996) has pointed out that developing students’ communicative language competence is a common goal for nations or cities undertaking broad educational reform. A closer look at the above-mentioned tests reveals that almost all test writers claim to base their tests on the construct of communicative language competence. At the same time, failing to implement Communicative Language Teaching (CLT) among teachers and/or failing to cultivate communicative language competence among students is an important cause of the negative washback effects occurred in these studies.

Since the early 1980s, CLT has been embraced as a paradigm in language teaching. It can actually be traced back to the late 1970s, when language learners were frequently criticized for mastering linguistic analysis in the classroom, but failing to use language appropriately for actual communication outside the classroom setting. CLT, hence, came into being as a teaching approach that “acknowledges the interdependence of language and communication” (Larsen-Freeman, 2000, p. 21). Language was then taught with a communicative intent that paid “systematic attention to functional as well as structural aspects of language” (Littlewood, 1981, p. 1).

What is more, specific communicative language competence was then defined and modified. It is now well established as a construct that consists of four components: grammatical competence, discourse competence, sociocultural competence and strategic competence (Savignon, 1997; Canale & Swain, 1980; Canale, 1983). Specifically, grammatical competence refers to mastery of linguistic knowledge and the ability to apply grammatical rules. Discourse competence is concerned with the ability to connect discrete words or phrases into coherent and meaningful discourses. Sociocultural competence requires learners to know when, where, why, how and what to say in specific socio-cultural contexts. Lastly, strategic competence is the ability to cope with unfamiliar situations or even situations where communication has broken down.

Although there is no one recipe for developing communicative language competence, generally all four language skills – listening, speaking, reading and writing (Celce-Murcia, 2001) – are important and there is a strong tendency to focus on meaning over linguistic structures. Accordingly, CLT shifts the roles of teachers and students: teaching is expected to be student-centered, and students are encouraged to be treated as

active learners. Teachers, rather than serving as knowledge transmitters, are transformed mainly into facilitators whose job it is to engage students in the negotiation of meaning. (Lee & Vanpatten, 1995; Yalden, 1981).

In the published empirical studies about language tests, the intended positive washback of almost all tests was to enhance communicative language competence and CLT. Some test writers asserted explicitly that they were working toward achieving positive washback by redesigning the test. For instance, the revised ICFES exam in Colombia, “seeks to evaluate the communicative competence of the students...[T]he document issued by the ICFES as framework for the examination makes explicit mention of the CLT approach as the official orientation of the teaching of languages in the country” (Manjarrés, p. 2). This is also true of the National Matriculation English Test (NMET) in China. Luxia (2005) has noted that a mild change towards the CLT approach; since Chinese test writers expect more emphasis to be put on teaching the use of language in middle schools. Similarly, writers of the O-level English exam in Sri-Lanka (Alderson & Wall, 1993) try to address the issue, claiming that “even those who did pass [the old O-level English exam] were usually not adequately prepared for the situations in which they needed English. The teaching program they had followed was structurally based” (p. 42).

CLT-Related Testing and Negative Washback

Nobody would question the goals of language testing and the good intentions of CLT for achieving communicative language competence. After all, everyone wants language learners to be successful in managing the use of English in and out of class. Then why are most of the English tests that are designed for communicative language

competence troubled by deleterious, negative washback effects? The author of this dissertation believes that one important reason is that some of these tests fail to meet the requirement of construct representation and/or construct-relevant variance. That is, they do not provide an adequate representation of language in use. This oversight threatens the construct validity of the test. And as pointed out earlier, the less a test satisfies the unified validity criteria discussed by Messick, the less likely it is to generate intended positive washback. The following empirical studies show results in accordance with this statement.

Studying the impact of the O-level English exam in Sri-Lanka, Wall and Alderson (1993) did not observe any change in teachers' methodology towards CLT in their classroom observation, even when the teachers claimed to have made the change. However, it should be noted that the original plan to have the O-level exam measure all four language skills was aborted. The oral skill measurement had to be dropped for practical and political reasons. So the new O-level English test proved to be a test of old-fashioned construction. Like its previous version, it focused only on reading and writing. In this case, as a test that failed to fulfill construct-representation, it can hardly be expected to generate the intended positive washback that the reformers had in mind.

The Entrance Examination of Universities (EEOU) English test used in Iranian high schools includes 25 multiple-choice and a cloze test. What is more, according to the observations of Salehi, Mustapha, and Yunus (n.d.), 11 of the multiple-choice questions are on grammar and vocabulary. So, this test not only threatens construct-representation, but may implicitly encourage more effective test-taking strategies at the cost of learning actual communication in English. It is not surprising, then, that test takers complain of poor skills in English listening, speaking and writing as reported in the study.

A similar situation exists in Colombia where the Instituto Colombiano para la Evaluación de la Educación (ICFES) administers exams for students who finish upper secondary school. Even though, as pointed out earlier, communicative competence was supposedly being assessed, according to Manjarrés (2005), “only linguistic competence with its three subcompetencies – grammatical competence, textual competence and textual coherence – is included in the obligatory test all the students have to take” (p. 7). So in his conclusion, even Manjarrés has to admit that “obvious underrepresentation of the concept of communicative competence in the test” (p. 14) is an important factor that results in negative washback.

For the NMET in China, Luxia (2005) mainly attributed negative washback to conflicting goals that were imposed on test writers, distracting them from focusing on linguistic knowledge. Nonetheless, she failed to address the issue that the NMET was not a valid communicative language test in the first place. From her analysis we know the NMET does not include assessment of listening and speaking ability, and it makes use of a multiple-choice format. As Hughes (2003) stated, “if the test content and testing techniques are at variance with the objectives of the course, there is likely to be harmful backwash” (p. 1). If the test had a better criterion representation, the goals of selection and of promoting educational changes may have complemented each other.

For tests that have a better criterion representation, intended positive washback effects are generally found. For instance, in Israel’s EFL oral test, Shohamy (1996) has reported that students are required to perform role plays, extended interviews, and an extensive reading task. Results of teacher interviews and student surveys indicate that the oral skills evaluated by the test have been taught in class. Positive washback effects are

found in teaching content and methodology. The English inspectorate also has a high regard for the intended positive washback achieved by the test. The Use of English (UE) Oral examination in Hong Kong achieves some positive washback effects as well. The UE Oral exam is a prerequisite test for students of secondary 7 who want to attend college. According to Andrews, Fullilove and Wong (2002), some of the washback effects seem superficial. However, they found traceable – although delayed – intended positive washback when comparing a cohort that did not prepare for the UE Oral test with two cohorts that prepared by using an instrument originally designed for the UE Oral test.

Negative Washback and Other Variables

Of course, there are cases in which tests fail to achieve intended washback even though they employ content and testing techniques that largely match the communicative language competence construct. A case in point is the well-known longitudinal washback study of the Hong Kong Certificate of Education Examination in English language (HKCEE). According to Cheng (2005), HKCEE is a performance-based assessment with a Target Oriented Curriculum. It was introduced to secondary school students for positive washback. In this new test, listening, speaking, reading and writing skills are all covered. What is more, the test format includes not only multiple-choice, but tasks that bear a high resemblance to real target-language use, such as integrated listening, reading, and writing, role play, and group interaction. Theoretically, a new test like this one should result in the intended positive washback.

However, Cheng (2005) found that “the washback effect is clearly limited to the surface level” (p. 246). Even senior HKEA officials concluded that the new HKCEE only

changed the *what* in teaching and learning in Hong Kong, not the *how*. Teachers' negative attitudes and actions toward a more communicative approach remained unchanged, although they incorporated new test-task activities in their teaching. Teachers still played a dominant role in the classroom and used basically the same teaching methods they used before the test was implemented. Cheng (2005) attributed these unintended washback effects to the fact that the new HKCEE protocol did not provide sufficient support to the teachers to help them implement the required changes.

The washback effects of the International English Language Testing System (IELTS) -Academic Writing Module on academic study is the subject of Green's (2007) research. He compares the outcome of IELTS Test Preparation courses with English for Academic Purposes (EAP) courses among students who plan to take the IELTS. In both types of classes he finds positive washback, including writing tasks other than test-like writing practices. However, the data also show that improved test scores do not necessarily mean improved target writing abilities. Nor does everyone strongly benefit from the preparation courses as has been claimed.

These two cases demonstrate that it is a prerequisite to design valid tests to create intended positive washback, but a valid test cannot be guaranteed to produce intended positive washback. Rather, test design is just one element and it is closely connected to and influenced by a larger set of variables in the complicated educational reform process. As Wall and Alderson (1993) have pointed out, washback is mediated by a number of variables throughout the test use process. This view also explains why educational reform will not proceed smoothly if tests are used as the major lever for educational innovation without taking other accompanying variables into account. In fact, Shohamy (1993)

claimed that “testers can no longer be viewed as technicians whose work is done when they reach satisfactory reliability coefficients; rather, they must consider the social, psychological, ethical, curricular, and educational consequences of the test they produce” (p. 1). To this, Davies (1997) added that “an ethical perspective for a language tester is ... necessary, it is not possible for a tester as a member of a profession to take account of all possible social consequences” (p. 335). The author of this paper agrees that testers are not solely responsible for the use of the test, yet they should be aware of ways to lessen and guard against negative consequences and misuses of the test. Other variables definitely play an important role in mediating the direction and intensity of washback effects. But what are these variables?

Washback and the Teacher Variable

In the course of test use, as has been discussed earlier, test content, test format, stake levels, and testing contexts all play a part in overall washback performance. And washback can change over time with the fluctuating dynamics of these variables (Shohamy, Danitsa-Schmidt & Ferman, 1996). What is more, in the washback literature, one of the identified and often-reported variables is the teacher. Insufficient teacher training is a problem commonly cited in many studies. Wall and Alderson (1993) believed that in their Sri-Lanka washback study, the lack of teacher training on the use of the textbook was a main hindrance. They believe that the lack of suggested methodology in the Teacher’s Guide for Years 10-11, as well as the teachers’ limited access to the Teacher’s Guide, were responsible for negative washback on teaching methodology. Likewise, Watanabe (1996) concluded that teachers’ educational backgrounds, experiences and beliefs are major reasons for the continuing use of the grammar-

translation method in language classrooms in Japan. Actually, it seems that in certain educational contexts, other strategies have been implemented beyond mere teacher support. These include developing and implementing new curricula and new textbooks (e.g. Cheng, 1998; 2005; Manjarrés, 2005; Wall & Alderson, 1993). However, the specific reason for not retraining teachers remains unclear. It sounds quite reasonable when Shohamy (2001) sharply argued that “using tests as disciplinary tools is a cost-effective strategy for policy making. In comparison to reforms through teacher training and development of new curricula or new textbooks, changing the test is a substantially cheaper venture” (p. 40).

In the meantime, some researchers have gone so far as to claim that teachers are the sole agent of washback effects and educational innovation. For instance, in Spratt’s (2005) review on washback studies of English language tests on teaching and learning, she claims that

the teacher [is] in the driving seat in some important ways as far as washback is concerned. When and where the teacher is in control of the factors determining washback, washback itself is largely under the teacher’s control. It is the teacher who can then determine to a greater or lesser extent whether to allow washback to operate, what areas it should operate in and how. (pp. 23–24)

Cheng (2005) came to a similar conclusion in her comprehensive longitudinal washback study that “In the end, the change is in the teachers’ hands” (p. 254).

These studies certainly demonstrate that teachers play an extraordinary part in eliciting intended desirable washback. However, a closer examination of these studies finds that the majority of them focus heavily on teachers’ perspectives and how the test

influences their classroom teaching. The other important party involved in the language classroom – the students – is scarcely investigated. What these test takers think of the test and how their learning has been influenced by tests remains largely under-researched. Even though classroom observations are sometimes employed in washback studies, rarely is attention focused on what the students are doing during the observation period, and what they think of their learning experience in preparation for, and following the test. Next, I will elaborate on why I believe the above-mentioned studies exaggerate the teacher's role in the washback mechanism without taking the students' voice into account.

Washback Studies and Often-Ignored Students' Perspectives

In a washback study of the new O-level English exam in Sri-Lanka, Wall and Alderson (1993) trained local teachers to conduct teacher interviews and classroom observations at different times during the study. Trained observers were required to provide a detailed description of each lesson in the classroom, – especially the types of texts and types of activities employed by the teacher, – and to judge the impact made by the test on the content and methodology of each lesson. In addition, the teachers being observed were interviewed immediately after the observation if they used a different textbook or methodology than the one suggested by the Teacher's Guide. What the students were doing and what they thought of their learning process were hardly mentioned. The teachers' selection of textbooks and activities were assumed to accurately reflect the students' needs in their preparation for the next exam or for their English learning in general.

Similarly, Watanabe (1996) investigated whether the grammar-translation method was part of the washback of implementing the university entrance English exam in Japan.

Watanabe interviewed two teachers both before and after classroom observations. Student activities were included in his observation agenda, as well as the materials used and the time teacher on his/her selected classroom activities. However, in his discussion, Watanabe writes solely about teacher factors in the choice of the grammar translation method. Though he deemed it important to ask for student opinions, he was more concerned that the time spent on questions might affect student performance on the upcoming exam. Furthermore, he believed it would be difficult for the students to answer questions about whether their English learning strategies outside the classroom were influenced by the types of questions raised by teachers in class. I would argue that if this was a cause of concern, then no washback research should be done before exams, because any research method, including classroom observation, might affect students' exam performance in some way. In addition, if it is difficult for students to describe the impact of different types of questions on their classroom learning methods outside of class, who else would be more qualified to describe it? In other words, how can we rule out that the grammar translation method is not what students preferred?

In investigating the washback of the ICFES exam in Colombia, Manjarrés (2005) went one step further by including interviews of students and teachers, in addition to making classroom observations. Out of five ethnographic classroom observations of the 10th grade at a public school, only three students from the classes being observed were chosen for interviews. What is more, in his 16-page paper the results of the student interviews were reported in fewer than 150 words if references are not counted.

Therefore, it appears that classroom observations largely serve to cross-check what the teacher is doing in class and to determine whether this corresponds to claims

made by the teacher either in a survey or interview. Even though they outnumber everyone else in a classroom, students are usually only cursorily examined in the course of washback studies. Rarely are their voices taken into account in researchers' classroom observation analyses. From the literature review, we see that most washback research is conducted from the teacher's perspective. Ironically, the implementation of student-centered CLT and learning by reform-oriented professionals appears to exert little or no impact on the dominant role of teachers in the classroom.

Occasionally, student interviews or student surveys are used as the major research methodology in washback studies. For instance, when investigating washback effects of the NMET in China, Luxia (2005) conducted in-depth semi-structured interviews with different stakeholders, including ten students. However, how these ten students were selected, and to what extent her findings could be replicated or utilized in other contexts is not explained. Also, the purpose of conducting the interviews is mainly to elicit questions for follow-up using relatively large-scale surveys for teachers and students. Thus, the extent to which this study represents students' perspectives needs further investigation.

In comparison, Cheng's (2005) four-year longitudinal massive-scale study of the English subtest of HKCEE is carefully-designed, and includes key informant interviews, student surveys, teacher surveys, classroom observation, and follow-up interviews with teachers. Since the research was done at the time when the last cohort of students who took the old exam and to a second group who took the next one. The survey results were then compared. Cheng found that the washback on student learning was superficial. Students' perceptions toward public exams remained largely unchanged. However, in

Cheng's follow-up classroom observation phase, her attention shifted completely towards teachers' behaviors. She even stated clearly that her focus on the classroom observation was to record "teachers' behavioral changes in the actual teaching and teaching contexts" and to have "detailed follow-up interviews of the three teachers who participated" (p.74).

Under such circumstances, therefore, it is not surprising that one of the conclusions most commonly drawn from washback studies on high-stakes English tests from 1993 to today is that high-stakes tests will have washback on what teachers teach but not on how they teach; i.e., they will not have much impact on the methodology employed by teachers (Cheng, 2005; Luxia, 2005; Manjarrés, 2005; Shohamy, 2001; Wall & Alderson, 1993; Watanabe, 1996).

These conclusions, reasonable as they may sound, have certainly understated the fact that students – the test takers – are actually the most immediate stakeholders in testing. How testing influences student learning deserves more attention. After all, testing and teachers are both mediators in students' learning. As we all know, a student can have the best teacher in the world, but whether he/she can learn successfully depends ultimately on the student. It is said in China that the sage Confucius only successfully cultivated 72 students out of a total of 3,000. An English saying expresses a similar idea: you can lead a horse to water, but you cannot make it drink.

Underestimating the role of test takers or exaggerating the role of teachers or other variables can never provide a correct or complete picture of washback. Even if students are assumed to be empty receptacles waiting for teachers to fill them with knowledge, the transformation still needs to be digested and internalized by the students themselves. Moreover, the goal of most of these tests is to promote CLT and to develop

communicative language competence, for which the expectation is that students be treated as active participants. As a result, only limited number of studies has been conducted on students' experience of educational change. Fullan (2007) argued that "educational change, above all, is a people-related phenomenon for each and every individual. Students, even little ones, are people, too. Unless they have some meaningful (to them) role in the enterprise, most educational change, indeed most education, will fail" (p. 170). Shohamy (2001) also suspected that test writers were not interested in the voice of test takers, for "in the testing literature test takers are often kept silent; their personal experiences are not heard or shared" (p. 97). She has stated that listening to the voice of test takers is a significant way to identify different uses of tests.

Necessity of Washback Studies on Students' Perspectives

The need for more washback studies from students' perspectives is enjoying increased awareness. Admitting that "we know very little about students' perceptions of tests (as opposed to their teachers' impressions of their perceptions) and even less about how new tests influence what students know and can do" (p. 506), Wall (2000) has suggested that more studies are needed in this area. The tendency to ignore student perceptions may partially explain why, out of the 15 hypotheses put forth by Alderson and Wall (1993), each hypothesis concerning teaching is paired with a counterpart regarding learning.

Likewise, Hamp-Lyons (1997) has stressed that tests bring different meanings to different stakeholders. She made an appeal for more washback research on students' views to enhance test writers' professional responsibility in language testing. Green (2007) has concluded that even now student perspectives remain "under-investigated in

the literature” (p. 314). He asserted that “variability at the individual level is central to an understanding of the complex process of washback and that the nature and extent of washback to learners does not bear a transparent relationship to washback to the teacher” (p. 314).

In peer-reviewed journals and books published to date, one of the few washback studies focusing primarily on student perspectives was done by Huhta, Kalaja and Pitkanen-Huhta (2006). They studies what test preparation, the test-taking process and getting score reports meant to test-takers using discursive approach, in which “[r]ather than seeing discourse as the product of psychological process, it [discursive psychology] considers the way in which psychology is produced *in talk* as parts *of practices*” (p. 330, as cited in Potter, 2003, emphasis in original). The English exam which is the subject of this study is the one compulsory component in a battery comprising the highly-competitive Finnish matriculation exam. The exam has two parts, the first being listening. The second part is administered several weeks after the first, and includes reading, writing, structure and vocabulary. Nine students in three different English levels participated in the study. They were asked to record their thoughts and experiences in an oral diary during preparation for the different parts of the English exam and through receipt of the final English grade. The students could decide when and how much to record, but six recordings at specified times were required in total. Group and pair discussions were held throughout the study. Topics were suggested both for student recordings and discussions.

This four-month longitudinal study shows that the test-taking process is far from linear, but is instead recursive. The researchers identified four pairs of roles relating to

different degrees of hard work, skillfulness, luck, and anxiety among the test takers by the researchers (p. 344). On the whole, the study reveals the complexity of the test-taking process from the test-takers' perspectives. With increased awareness of the different roles assumed by students, and students' own understanding of those roles, a more balanced test preparation process can possibly be developed.

The IELTS washback study by Green (2007) is another study that has considered washback to students as well as washback to teachers. It examines the effects of various preparation courses for academic writing in higher education. A total of 476 international students from 50 nationalities at 15 institutions in the UK at different levels participated in the study as well as a number of teachers who took part in a survey or in focus group interviews. Each student was enrolled in IELTS preparation or pre-session English for Academic Purposes (EAP), or in the two courses simultaneously. All of these students participated in student surveys and many of them also participated in student focus groups. In addition, Green complemented the study with classroom observations. His classroom observations are distinguished from others in that he pays equal attention to teacher actions and student actions in his observation agenda. Thus, he comes to acknowledge how complicated and diverse washback effects can be to learners. More importantly, he concludes that it is essential for the students' perspective to be considered to understand overall washback effects.

One more study that has paid due attention to the students' perspective is the washback study of the General English Proficiency Test (GEPT) on English learning in Taiwan by Shih (2007). The GEPT is one of the degrees required by many colleges and universities in Taiwan. It includes five levels: elementary, intermediate, high-

intermediate, advanced, and superior, according to the Language Training and Testing Center (LTTC) in Taiwan. At all levels except the first, the exam is divided into two phases which must be taken consecutively: the first tests listening and reading skills, and the second tests speaking and writing. Only with passes in both phases can the certificate be granted. Noting the lack of research on washback on learning, Shin focuses on this subject.

The study consists of interviews conducted in two private institutions of higher education in Taiwan. The interviews are administered to 14 to 15 students from the day, night or weekend division of each institution, as well as to each department chair, two to three English teachers, and even three students' family members. Besides classroom observations, department self-study center observations and document reviews are among the research methods used in this study. Of these two private institutions, the first does not have a GEPT degree requirement, while the second requires its day-division students to pass either the first stage of the GEPT's intermediate level or the school-administered make-up exam to fulfill the degree requirement.

Shih finds that the intensity of washback is stronger in the second institution. Students in general spend more time preparing for the first phase (listening and reading), while they vary in their attitudes towards speaking and writing preparation. Given the intricacy of washback effects, Shih proposes a tentative washback model of student learning, which will be discussed next in the context of different washback mechanisms/models/hypotheses that have emerged in the literature.

Washback Mechanisms/Models/Hypotheses

As mentioned earlier, the most well-known washback hypotheses were introduced by Alderson and Wall (1993). These hypotheses help to clarify some widely existing assumptions that need to be tested empirically. Originally there were 15 of them and they are:

1. A test will influence teaching.
2. A test will influence learning.
3. A test will influence what teachers teach.
4. A test will influence how teachers teach.
5. A test will influence what learners learn
6. A test will influence how learners learn.
7. A test will influence the rate and sequence of teaching.
8. A test will influence the rate and sequence of learning.
9. A test will influence the degree and depth of teaching.
10. A test will influence the degree and depth of learning.
11. A test will influence attitudes to the content, method, etc., of teaching and learning.
12. Tests that have important consequences will have washback.
13. Tests 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. (pp. 120 –121)

Later, after examining the washback effects of TOEFL preparation classes and non-TOEFL preparation classes, Alderson and Hamp-Lyons (1996) revised Hypothesis 15 to state: “Tests will have different amounts and types of washback on some teachers and learners than on other teachers and learners” (p. 296).

These 15 hypotheses distinguish between washback to teaching and washback to learning. What is more, further distinctions are made between content, methodology, strategy, rate and sequence, and degree and depth of teaching as well as learning. These hypotheses have set the standard and direction for studies of washback. They have also legitimized the study of washback effects to learning as opposed to washback effects to teaching.

Hughes (1993) then proposed a washback model. Although the Hughes paper was never published, an illustration by Bailey (1996) has given the model wide recognition. In Hughes’ washback mechanisms, washback is composed of participants, process and product. Hughes notes that participants refer to all stakeholders, including teachers, students and administrators. Process generally means the broad learning process undertaken by participants, such as choosing textbooks, teaching methodology and learning strategies. Product refers to learning outcomes. Hughes asserts that the influence of a test can have a chain effect from perceptions of the participants to product.

Along with the introduction of Hughes’ mechanism, Bailey (1996) also challenged the mechanism, claiming that “not all the participants’ processes lead directly to learning” (p.262). Partly influenced by the distinction between washback to learning and washback to teaching, Bailey proposes a dichotomic washback model: washback to the learner and washback to the program. She related Alderson and Wall’s hypotheses 2,

5, and 6 above to her washback to the learners, and hypotheses 1, 3,4,7,9 and 11 to her washback to the program. She has elaborated that due to high-stakes testing, learners may involve themselves in a variety of learning activities. In this sense, Bailey's model places more stress on the importance of washback to the learners.

In his comprehensive review of washback research, Green (2007) offered a model of washback that incorporates washback direction, variability and intensity. Basically, this model demonstrates that the greater the overlap between test characteristics (e.g. test format, content, etc.) and focal construct, the more likely it is that the test will generate positive washback. At the same time, test stakes and participants' characteristics and values play a part in producing variability in washback. In other words, if participants know what needs to be changed, they will embrace the changes and get properly trained and equipped to make the changes; therefore, more intended positive washback can be anticipated. If participants are aware of the extreme importance of a test, and the test is reasonably difficult, the test is more likely to generate the expected degree of washback intensity.

Unlike previous models, the Green model depicts the relationship between testing and changes on language and learning as being far from linear, in addition to the close relationship between validity and washback effects in the first place. As discussed earlier, in most empirical studies the more valid the test, the more likely it is to produce intended positive washback effects, as conditioned by test stakes, test difficulty, material resources, teacher resources, participants' perceptions and actions. However, in Green's model the difference between variability and intensity is not quite clear, nor are the dynamics between washback variability and washback intensity. Another debatable issue is whether

there exists a state of no washback. As pointed out earlier, washback is an inherent quality of a test; whenever any external test is employed, a degree of washback is anticipated. So the author of this paper would suggest that the counterpart of intense washback be changed from no washback to less intense washback.

Given the important role played by students and the scarcity of literature on washback in learning, Shih (2007) has proposed a washback model of student learning based on his washback study on the General English Proficiency Test (GEPT). This model highlights extrinsic factors, intrinsic factors and test factors. Shih deems that these three factors are interrelated and act simultaneously to make a difference on the washback of a test on students' learning and psychology. Extrinsic factors include subcategories such as socioeconomic factors, school and educational factors, family, friends and colleagues, and personal factors. Intrinsic factors focus more on individual differences, characteristics and perceptual differences. Test factors refer to the test stakes, content, difficulty, etc. Influenced by all these factors, the washback of a test on the psychology of student learning is revealed in their learning content, time, strategies, motivation, anxiety and so on. This washback continues to affect results of the test and subsequent learning. What is more, the results of the test can directly and indirectly counter-influence the three primary factors, particularly those which are intrinsic.

Compared with the other hypotheses/models above, this is the only model that attempts to illustrate washback to learning. It is also significant because it incorporates socioeconomic factors besides school, educational and personal factors. In contrast, Green's Model of washback, incorporating intensity and discussion (2007) seems to imply that test quality plays the most salient role in generating washback. Likewise,

Alderson and Wall (1993), Hughes (1993) and Bailey (1996) disregarded extrinsic factors in their models, even though every test is mediated by socioeconomic conditions and these conditions are always interacting with each other. Since this dissertation examines washback effects from the perspective of student, Shih's tentative washback model of student learning provides a good reference for my study. Moreover, Shih's study is conducted in two private vocational institutions where all his participants are English majors in applied foreign language departments, suggesting that studying different tests in different research contexts will provide valuable insights into his model. A comparison between extrinsic factors and the results of the test, and against the factors and the results described in his model should also be fruitful.

Last but not least, in Shih's Test Factors, he has not clearly demonstrated the idea that the greater the overlap between test characteristics and focal construct, the greater is the potential positive washback. In addition, based on the hypotheses of Alderson and Wall (1993), the factors of learning and psychology, rate and sequence, degree and depth can be added to the category designated as Washback of a test on students. Nevertheless, since the focus of this paper is on washback on student learning, Shih's tentative washback model of student learning will be the major reference point for this study.

CHAPTER 3

MACRO RESEARCH CONTEXT AND EVOLUTION OF THE CET

Macro Research Context: English Learning and Testing in China

Compared to tests such as the Test of English as a Foreign Language (TOEFL) and International English Language Testing System (IELTS), the Chinese CET-4 is not widely known to the international community. This study investigates the latest reform of the CET-4, and the washback effects created by these reforms on the teaching and learning of English in Chinese colleges and universities. Before dissecting the details, a brief account of the CET-4 history is in order, including the reasons why the reforms were instituted, and who was responsible for instituting them. Like the extrinsic factors included in Shih (2007)'s washback model of student learning, Watanabe (2004) emphasized that in washback studies, both macro and micro levels of context should be described in detail. By macro level, he has referred to the broader educational and social environment in which the test in question is implemented. Micro-level context, then, is the specific classroom and school setting where the study takes place and where the test is used. According to Watanabe, it is important to give very explicit descriptions of both contexts, because only then will readers have a clear idea of the role played by the test and be able to apply reasonable "transferability" (p. 25) to their teaching, learning and testing contexts when useful.

Bachman (1990) pointed out that "tests are not developed and used in a value-free psychometric test-tube; they are virtually always intended to serve the needs of an educational system or of society at large" (p.279). The practice of English teaching, learning and testing in China nowadays is no exception to this notion. In fact, this should

be obvious to even the casual observer considering the rapid development and changing demands of the Chinese political, educational, and social system after the foundation of the People's Republic of China in 1949.

Since that time, a series of major historical events, such as the deleterious relationship between China and the former Soviet Union in the 1960s, President Nixon's visit to China in 1972, and the ending of the Cultural Revolution in 1976, paved the way for English to replace Russian as the most popular foreign language in China. Ever since the implementation of the Reform and The Open - Door Policy in the 1980s, China has witnessed the steady development of English education nationwide (Lam, 2002). In the 1990s, foreign languages were officially established as a compulsory testing subject in the National Matriculation Test (NMT), whether the student chose the Art Stream or the Science Stream in senior high school. Given the fact that the majority of elementary schools teach English as the foreign language, the English test has been widely accepted as the default foreign language test. Other language options like Russian, French, German, Japanese, Spanish and Arabic are officially considered minor foreign languages in China and only a very small number of students take these tests each year.

What is more, students begin to receive formal English education in school at an even earlier age. In the 1980s, most students started studying English in middle school. Gradually, primary schools in the big cities or coastal areas began teaching English in the 5th or 6th grade, resources permitting. However, "since the mid -1990s, English began to be taught for Grade 3 in primary education" (Cheng, 2008, p. 16). In 2001, with China's entry into the World Trade Organization (WTO), the Ministry of Education (MoE) mandated that English education should start in the 3rd grade in order to cater to future

demands. It is not surprising to see that many kindergartens offer basic English classes upon the request of parents. English, accordingly, became a compulsory testing subject at entrance exams for middle schools and high schools nationwide. Even for students who do not plan to study beyond the 9-year compulsory education requirement, the English test is a component of the battery of exit tests.

In the meantime, China's active engagement in world affairs and the spread of globalization has made effective English communication a pressing social demand. The early 2000s, particularly, witnessed a significant fervor for English learning in China due to the achievement of significant diplomatic milestones: In July 2001, Beijing was designated as the host city for the 29th Summer Olympic Games to be held in 2008. Then in December 2001, as the biggest developing country in the world, China was accepted as the 143rd member of the World Trade Organization (WTO) after years of painstaking negotiations. What is more, in December 2002, Shanghai was selected to host the 2010 World Exposition. All of China was excited about these international events, but the people were aware that these were not just opportunities for the world to get to know China; they were opportunities for the Chinese people to introduce China to the world. As part of this initiative, the people of China were challenged to improve their overall English proficiency level. Now, with the successful conclusion of the 29th Summer Olympics games in Beijing in 2008 and the World Exposition in Shanghai in 2010, the Chinese increasingly believe that in the globalized village, a good command of the English language can "provide its learners with the necessary qualification they need to be members of these expert communities [in science, technology, business, commerce, and so on]" (Widdowson, 2003, p.56). English language teaching and learning is no

longer simply an educational issue, but an issue related to long-term comprehensive national strength and China's international competitiveness. The mismatch between rapid economic development and a rather stagnant English teaching and learning environment caused more and more concern among upper-level officials in China, since low English proficiency had become a palpable barrier to introducing high technology, and cooperating in the global marketplace.

In fact, an English credential has now become a prerequisite, not just for joint-venture employment, but for promotion in state-owned work places in China. Chinese testing experts under the supervision of the Ministry of Education (MoE) have designed a variety of tests catering to the burgeoning demand for accuracy in measuring English ability at various levels and for different purposes. Besides the National College English Test (CET), other influential tests include the National Matriculation English Test (NMET), Business English Certificates (BEC), the Public English Testing System (PETS), China Accreditation Test for Translators and Interpreters (CATTI), the National English Test for Professional Title Promotion (NETPTP), the Waiyu Shuiping Kaoshi (WSK), the Graduate School Entrance English Examination (GSEEE), and the Test for English Majors (TEM).

The College English Test (CET)

As mentioned above, ever since the steady implementation of the Reform and The Open-Door Policy in the 1980s, a great demand has emerged for intellectuals who are proficient in English. In particular, because college graduates usually serve as the crucial work force driving the substantial economic and technological development of the country, the MoE began placing great emphasis on English education at the higher levels.

In order to standardize the quality of college education in English, the first National College English Teaching Syllabus (which will be referred to as the “Syllabus”) was published in 1986 after a large-scale longitudinal investigation (Jin, 2010) by a team of senior English scholars.

In the first Syllabus, college English courses consisted of six bands. Each band was equal to one semester’s formal college English study. The courses from band 1 to band 4 were required general English courses for all non-English majors. Courses from band 5 to band 6 were designed to teach English for specific purposes, and elective classes were tailored to specific majors. The objective of learning English was mainly to develop good reading skills, some listening skills, and basic writing and speaking skills, so that students could, with English as a tool, draw on information related to their major and build a solid foundation for future improvement in their English language competence.

Tests have long been used as a means of enforcing national policies in Chinese history, and they inevitably reflect the social realities of the period. Thus, the implementation of these relatively low standards is appropriate to the social realities of that time, given that for ten years Russian language learning had been emphasized, followed by another ten years in which English education came to a standstill during the Cultural Revolution.

Today, the development of the standardized college English tests is the task of the national College English Test Design Group appointed by the MoE. In accordance with the Syllabus, the Design Group developed two testing levels based on different bands: CET Band 4 and CET Band 6. CET-4 is given at the end of the fourth semester of college,

and CET-6 at the end of the sixth semester. However, only those who pass the CET-4 are eligible to take the CET-6. If the student does not pass the test the first time, he/she can keep taking it before they graduate.

In 1987 the first CET-4 was administered, and around 100,000 students sat for the test. In 1989 the first CET-6 was administered and 60,000 students took the test. With the test gaining greater acceptance, in 2003 more than 9.15 million college students took the CET (Jin & Yang, 2006). In November 1999, with increasing demand for improved oral English communication, the CET-SET (Spoken English Test) was introduced. The CET became a test battery composed of CET-4, CET-6 and CET-SET, the latter being administered independently of the other two tests. Moreover, the CET is the only national standardized test at the college level in China. Each of the three tests is offered twice a year. Both the CET-4 and the CET-6 are administered on the first Saturday in December and the third Saturday in June. Most examinees take the test in June, when they are in the fourth semester or sixth semesters at college. The CET-SET, on the other hand, is offered in May and November.

Before June 2005, students who passed the CET-4 or CET-6 were issued a certificate by the National College English Testing Committee (NCETC) on behalf of the Higher Education Department of the MoE. Specifically, two kinds of certificates were awarded: examinees who received a score between 60 and 85 earned a pass certificate, and those who received a score of 85 or higher earned a distinction certificate.

CET-SET, however, is reported on a letter grade scale: A, B, C and D, with A as the highest grade level and D the lowest. Certificates are awarded to those who achieve A, B and C levels only. Before June 2005, examinees whose CET-4 score was 80 or above

or whose CET-6 score was 75 or above, were eligible to register for the CET-SET. The CET-SET is conducted by well-trained examiners authorized by the NCETC. The whole CET-SET takes 20 minutes and consists of a face-to-face interview with the examiner, an individual presentation, and a discussion with two or three other examinees on a given topic. As to the registration fee, the charge for the CET is much less than for the TOEFL or IELTS tests. The fee is approximately around 25 Ren Min Bi (RMB - the Chinese unit of currency) for the CET-4 or the CET-6 and 50 RMB for the CET-SET. In other words, registration amounts to no more than \$10 per test.

Since a far larger number of students take the CET-4 as compared to the CET-6 or the CET-SET, the CET-4 was selected as the subject of this dissertation. In the rest of this chapter, the old version of the test and the corresponding washback effects will be described, followed by description of the reformed version and its intended washback effects.

Old Versions of the CET-4 and Washback

The first version of CET-4 was based on the Syllabus. It was designed in 1986 and administered in 1987. In 1999, the Syllabus was revised, as was the CET-4. The revised Syllabus specified that the objective of college English was to develop fairly strong reading ability and to foster considerable listening, speaking, writing and translating skills, so that students would be able to exchange information in English. College English courses were expected to give students a solid foundation in the English language, enable them to master useful language study skills, and enhance their cultural knowledge in order to meet the needs of social and economic development.

To assess these various skills, both the CET-4 and CET-6 were designed according to the “skills and elements” model (Bachman, 2007, p. 46, as cited in Carroll, 1968; Davies, 1977; Lado, 1961) of language testing. In this model, examinees are tested in relatively specific skills such as listening, speaking, reading and writing. Both before and after the reform of the CET-4 in 2005, the CET-4 and CET-6 were composed of distinct parts for testing listening, reading, writing and translating skills.

Based on the Syllabus, the CET-4 is composed of five parts. Part I, Listening Comprehension, includes two sections: Section A consists of ten short conversations which to be only read once, and Section B includes either three short-passage comprehension items that are each read only once, or a compound dictation (i.e., a dictation which requires listening to a passage and filling in missing words or sentences) that is read three times at different speeds. According to the Syllabus, CET-4 examinees should be able to understand listening materials on familiar topics delivered at a speed of 130-150 word per minute (wpm). Except for the compound dictation, the questions in Part I are all multiple choice. The compound dictation did not appear in the CET-4 until 1997. Both basic and higher meta-cognitive listening skills are tested in the two sections. In the 90s, partly as a washback effect, most colleges and universities offered one 90-minute English listening course in addition to courses in comprehensive English. In the first 15 years of CET-4 administration, with the practice and training students received through English courses, the mean score on the listening component rose from 10 to 12.5 (Jin & Yang, 2006).

Part II of the CET-4 is Reading Comprehension, in which students are given 35 minutes to read four passages of approximately 1,500 words. The selected passages are

usually argumentative or expository piece. According to the Syllabus, CET-4 examinees should be able to read them at a speed of 70 wpm. These passages are followed by 20 multiple choice questions ranging from fact identification, to tone, and inference. Since reading is greatly emphasized in the Syllabus (accounting for 40% of the total score) emphasis in college English classes has been on reading. In some colleges and universities even the Comprehensive English courses are called Intensive Reading courses, and every text is analyzed word by word and sentence by sentence. With this intensive training, the mean test scores in the Reading section of the CET-4 have also risen from 25 to 27.5, after 15 years of administration of the CET-4 (Jin & Yang, 2006).

Part III is called Vocabulary and Structure. This part includes 30 incomplete sentences. Students have 20 minutes to choose the best answer from four choices. The questions cover a mixture of grammar and vocabulary. Although this part only comprises 15% of the total score, it has always been the most controversial part of the CET-4. In some colleges or universities, in order to make sure students score well on this part, English classes are turned into mini grammar and vocabulary classes, deconstructing texts to learn new vocabulary and practice grammatical structure. Because the general vocabulary requirement for CET-4 examinees is 4,500 words, as established in the Syllabus, it is not surprising to see students every day memorizing thick CET-4 vocabulary dictionaries on Chinese college campuses.

Part IV is either Cloze, Short-Answer Questions, or Translation. If cloze, the examinee is expected to choose the best answers (multiple-choice questions) for 20 blanks in a given passage. The short-answer questions, which appeared after 1997, are usually composed of a short passage with five questions or incomplete statements. The

examinee is then asked to answer the questions or complete the statement without exceeding ten words. The time limit is 15 minutes. This part comprises 10% of the total score. Finally, the translation component, which is from Chinese to English, includes questions based on the Reading Comprehension section.

Part V is Writing, most of which requires composing argumentative or expository essays, because these are genres that a student will most likely encounter in the context of his/her future work environment. Usually a brief outline in Chinese will be given as a prompt. The examinees are then asked to write at least 100 words within 30 minutes. Since writing contributes greatly to the overall validity of the CET-4, starting in 1990, the test was administered on a separate sheet, 30 minutes before the end of the test so that the examinee could not sacrifice writing time for other parts of the test, or vice versa. Moreover, in 1998 a pass/fail cut-off score was established: those whose writing scores are lower than 6 out of 15 are penalized on their overall score. A zero in writing automatically results in a failing grade on the overall test. These changes greatly boosted the teaching of writing in college English classes, as seen by the rise of the average writing score from 4 points before the changes were introduced, to more than 8 points afterwards (Song, 2005, para. 11).

Usually, for students who are not English majors, Chinese universities offer six 45-minute English classes every week for the first two years of college. The classes are divided into four Comprehensive English courses and two English Listening courses. At the end of the second year of study, students are allowed to take the CET-4 in June. The Comprehensive English course is the main course for preparation for the CET-4. The MoE usually recommends a range of textbooks for these courses, all compiled in

accordance with the CET-4 requirements. Every college/university is free to choose one textbook from among these options for use in teaching comprehensive English. The teacher's manual and students' workbooks are usually recommended as well. The manual – for instance, College English, one of the most popular textbooks edited by Dong, Yafen – focuses primarily on vocabulary and reading strategies instead of developing writing, listening and speaking skills. Before the major reform of the CET-4, students in the Comprehensive English courses spent a lot of time on grammar, new vocabulary and reading strategies. This was to be expected since Reading comprised 40% of the total score in the old CET-4, and Vocabulary and Structure comprised 15%. As a result, teaching to the test in College English classes was strongly condemned as causing negative washback (Gao, 2003; Li & Wang, 2003; Zhang, 2004).

Moreover, among all the skills tested, speaking still remains the most underdeveloped component, even though the CET-SET was issued in response to widespread criticism of 'dumb' English learners after 10 years' study at school. Records show that by 2004 only 200,000 students were qualified to take the CET-SET. Yet in 2004 alone, 110,000,000 students took the CET-4 and CET-6 (Wang, 2004). The ever-expanding English classes at college posed an even more serious problem to English teaching: how it is possible to offer every student an adequate and fair opportunity to speak English or practice writing?

Before June 2005, both CET-4 and CET-6 test scores were based on a 100-point scale. Unlike other national, standardized foreign language tests, the CET is a criterion-related, norm-referenced test. It is criterion-related because it is designed to correspond to the criteria prescribed in the NCET Syllabus. It is norm-referenced because the test score

indicates the percentile position of the examinee against that of a norm of approximately 10,000 undergraduates from six top Chinese universities: Beijing University, Tsinghua University, Shanghai Jiaotong University, Fudan University, University of Science and Technology of China, and Xi'an Jiaotong University (Yang & Jin, 2001). Whether or not it is reasonable to set performance standards based on students from top universities remains a question. Also, many people assume that the CET-4 is a proficiency test. According to Bachman (1990), “syllabus-based tests are generally referred to as achievement tests” (p. 71). This is exactly the case with the CET-4, since it is a test battery developed to measure to what extent examinees have met the requirements of courses designed on the basis of the NCET Syllabus (Zhang, 2003). Cai (2006) has reiterated that the CET-4 and CET-6 are “standardized achievement testing” (p. 234).

As a criterion-related, norm-referenced test, the main purpose of the CET is to assess whether a student has met the prescribed requirements of the syllabus, not a student's overall English proficiency. Nonetheless, since the inception of the CET-4 in 1987, it has drawn enormous attention as a proficiency test, and has had a major impact on society. Especially since the mid-1990s, with the abolition of the nationally guaranteed, unified job assignment between employers and college graduates, the adoption of two-way choice between employers and college graduates, and the existence of freedom of choice in employment, the CET-4 has become one of the essential standards for filtering job applicants, whether the job requires knowledge of English or not.

Then in 1999, as a national policy, colleges and universities in China started expanding their enrollment to accommodate more students. Because of this expansion, in

2001 there were 1.15 million college graduates, there were 1.45 million in 2002, and this grew to 3.8 million in 2005. The rapid expansion unavoidably resulted in fiercer job competition. In 2001 the number of unemployed college graduates was 340,000, which became 370,000 in 2002. By 2005 the number had risen to 790,000 (Xue & Wang, 2006). Under such circumstances, big cities like Beijing, Shanghai, and Guangzhou have explicitly announced that without a CET-4 certificate, newcomers would not be accepted as official residents of these cities. Accordingly, more and more college students took the CET-4 to enhance their marketability.

Meanwhile, the CET-4 test gained importance in ranking universities. According to the NCET Syllabus in 1999, CET Band 4 is the officially recognized basic English language requirement for Chinese colleges and universities nationwide. The pass rate of the CET-4 is the only indicator that is comparable throughout China, so it is used to rank the overall level of educational success for each college or university. Moreover, the MoE makes public the employment rate of graduates in each university. Obviously a university's prestige is associated with its graduates' employment rate. Under such pressures, official university policy of many universities states that students who do not pass the CET-4 cannot be granted a diploma, which is required for employment. These policies exist even though the NCETCC (National College English Testing Committee Commission) makes it clear that the CET-4 should be taken on a voluntary basis.

This policy has had a tremendous influence on college English teachers and students. In order to boost the pass rate of the CET-4, some colleges or universities interrupt their normal English classes and replace them with exclusive CET-4 preparation classes, which may be held one or even two or three months ahead of the actual test date.

Students are then given previous CET-4 tests or simulated CET-4 tests instead of regular English courses, which led to a notable decrease in some students' motivation for learning English (Cai, 2006, p. 268). In a survey by the China Foreign Language Education Research Center in October 2004, involving 4,000 non-English majors, 19% of students claimed that almost all of their spare time was spent on English, and 56% of students spent the majority of their spare time on English ("Survey shows," 2005, para. 1). Since every college student usually takes more than five courses every semester besides English, the allocation of time is clearly skewed, driven by the high stakes of the CET. What is worse, many employers complained that CET-4 certificate holders could not communicate effectively in English. As a result, more and more people appealed to officials to remedy the situation. Cheng (2008) even concluded that "the CET has exerted a huge amount of influence, reportedly negative, on English language teaching and learning at the tertiary level in China since its first administration" (p. 19). A famous senior English professor, Runqing Liu from Peking University, has openly called for its complete abolition because of the negative impact of the CET-4.

It is sad to learn that failing to pass the CET-4 is one of the top three reasons listed for suicide, and thoughts of suicide among college students. These findings were made by the first College Students Suicidal Psychology Survey conducted in 2007 by Chongqing Medical University among 10,000 students at more than 10 universities. In 2006, one male college senior in the city of Chengdu and one 25-year-old female student in the city of Nanjing committed suicide, mainly because of repeated failures to pass the CET-4.

As to the administration of the CET-4, severe penalties are applied for releasing the CET-4 test content ahead of time, cheating, or facilitating cheating. Nevertheless, every year a number of teachers, administrators, students and proctors choose to risk their academic careers by doing so. For instance, in December 2004 the former vice director of the English department at Xi'an Jiaotong vocational college and another employee were sentenced to jail for 18 months for releasing the CET-4 test questions to students before the test. In June 2004, it was reported that eight graduates were dismissed from their schools in the city of Wuhan because they were found taking the CET-4 for others.

In short, with increasing unintended social consequences incurred by the CET-4, a strong case could be made for substantial reform to the CET, as well as for college English teaching and learning in general.

Reformed CET-4

While the unintended negative washback effects on college English teaching and learning were being criticized by more and more people, after more than 18 years of use, it was clearly time to substantially reform the CET-4 in order to meet the ever-demanding social need for college graduates with practical communicative English skills.

As stated earlier, College English teaching reform was an essential part of the Higher Education Undergraduate Level Teaching Quality and Teaching Reform Project of 2005. This project primarily consisted of three major reforms: First was reform of the previous college English teaching curriculum - changing the focus of college English teaching and learning from reading to the development of comprehensive ability, particularly listening and speaking. Second was reform of the previous teaching mode: shifting from lecturing and chalkboard to student-centered instruction, use of computer

technology, and internet resources. Third, the reform of the CET-4 and CET-6, the English ability assessment tools at the college level. In response to this national project, the CET committee published a more up-to-date manual called College English Curriculum Requirements (CECR) in 2004. To some extent, this served as the first move towards the reform of the old version of the CET-4. According to the new CECR manual, the College English course is an integral part of higher education and a required basic course for undergraduate students. English linguistic knowledge, practical applications, learning strategies, and cross-cultural communicative skills are the main components of College English courses. The objective of College English has also shifted away from mainly developing the students' reading ability, to developing the ability to use English in a well-rounded way, particularly through listening and speaking skills, so that they can effectively exchange information in both spoken and written form in their future work and social interactions. At the same time, students are expected to enhance their ability to study autonomously and raise their overall cultural awareness.

Four newly designed textbooks published by four famous Chinese publishing houses were adopted by the MoE for the reform of College English and the CET-4: New Era Interactive English by Tsinghua University Press, New Horizon College English by Foreign Language Teaching and Research Press, New College English by Shanghai Foreign Language Education Press, and College English New Experience by Higher Education Press. With the new CECR manual and new textbooks, Professor Yan Jin, chairperson of the NCETC (National College English Testing Committee Commission), and Professor Yang, Huizhong, former chairman of the Commission, remarked that “as a result, the task of the CET designer today is to provide a comprehensive assessment of

the testees' communicative language ability so as to promote the teaching and learning of English to meet the challenges posed by social and economic development" (Jin & Yang, 2006, p.35).

In February 2005, the MoE held a news conference to announce the official reform of the CET-4 to the public. One of the biggest changes recommended was the abandonment of the former certificate, and substitution of a new grading scale ranging from 220 to 710. Given an estimated normal distribution, the average score would be 500. This new scoring system does not define a specific pass/fail score. Rather, employers decide on a desirable score depending on their own needs. What is more, the MoE and the CET-4 committee reaffirmed that the CET-4 is an achievement test, not a proficiency test; therefore, as of December 2007, non-college students are not allowed to take it. Also, a minimum score of 425 is required in order to take the CET-6 and the cut-off scores for taking the CET-SET are 550 and 520 for the CET-4 and CET-6, respectively.

In response to the new teaching objectives, many changes were made to the original CET-4. Compound dictation (i.e., a dictation which requires listening to a passage and filling in missing words or sentences) has become a fixed feature of the Listening Comprehension section instead of an optional one. The listening materials are now adapted from real conversations, presentations and broadcasts from English-speaking countries and are therefore more authentic. Besides multiple-choice questions, students will be asked to fill in missing words, phrases and two sentences, resulting in a more accurate assessment of students' listening ability. Finally, the weight carried by the Listening section rose from 20% to 35% of the total score.

The Reading section has been further divided into Reading in Depth and Fast Reading. There are three reading passages in Reading in Depth. The first two require answering multiple-choice questions. The third one can be a Banked Cloze or a 200-word passage followed by Short Answer Questions. For Banked Cloze, students are required to reconstruct the original piece of 200 words by choosing extracted words from a word bank and placing them in the correct blank spaces. This test format differs from other blank-filling cloze in that it provides fifteen words for ten blanks. Students have to figure out which five words are redundant. For Short Answer Questions, students are asked to answer specific questions about the passage using fewer than ten words. In the new version of the CET-4, Banked Cloze has replaced Grammar and Vocabulary and the latter section has been cancelled completely on the assumption that grammar and vocabulary will be tested in a more contextualized manner through Banked Cloze and in other sections of the test. Again, validity of the test is improved with the inclusion of questions which are not multiple-choice. As for the Fast Reading section, it requires students to take fifteen minutes to skim or scan one passage of approximately of 1,000 words and answer questions. The multiple-choice questions have been replaced by but seven true-or-false statements and three fill-in-the-blank questions. This change is more application-oriented and makes reading more meaningful.

Cloze, follows the reading, with the goal of assessing students' reading comprehension using a top-down approach. According to the new CECR manual, the test format in this section can at times be Error Correction, in which students are required to identify and correct ten mistakes in ten numbered lines of a short passage of approximately 280 words. So far Error Correction has never appeared in the CET-4, but it

is a fixed test format in the CET-6. The Writing section has been changed to Writing and Translating. However, the writing task itself remains unchanged: an impromptu argumentative or expository essay introduced by a brief outline in Chinese. Five short translation exercises – from Chinese to English only – follow the essay writing component. Part of the sentence is given in English and students are then asked to translate the remaining part of the sentence according to English usages and standards.

In addition to the changes to test content, the sequence of the different sections of the CET test has also been adjusted. Instead of having the Listening section first and the Writing section last as before, a 30-minute essay writing is followed by a 15-minute fast reading. These tasks, which are performed on Test Paper I, are then collected. The proctors then proceed to distribute Test Paper II, which starts with Listening and ends with Translation, and lasts for 80 minutes. According to the instructions, Listening should be finished within 35 minutes, In-depth Reading 25 minutes, Cloze 15 minutes, and Translation 5 minutes.

As to scoring, four subscores are reported: Listening, Reading, Writing and Integrated Part. According to the NCETC (National College English Testing Committee) in 2005, the integrated Part counts for 15% and includes the scores of two parts: Cloze or Error Correction, and Translation or Short Answer Questions. However, from the score reports I gathered, the integrated part appears to only account for 10% of the total, which is 70 points, because the scores of the other three categories are all higher than 100. The scores of Listening and Reading are usually higher than 200. So, as the lowest score among the four subscores, the score of the Integrated Part may be the score of the Cloze, which is worth 10% of the total. Therefore, I think that in the score report, the total score

for Listening is 249 (35%), Reading 249 (35%), and Writing and Translating 142 (20%).

The following chart is a comparison of the CET-4 before and after the January 2007 reform:

Table 1

CET-4 Before January 2007 (150 minutes)

Sections	Content	Sub-Contents	Format	Percentage
I - Listening	Conversation (Con.)	Short Con.	Multiple choice	20%
		Long Con.	Multiple choice	
		Passage or	Multiple choice	
		Compound Dictation	Fill-in-the-blank	
II – Reading	In-Depth Reading	Passage	Multiple choice	40%
III – Vocabulary and Structure	Vocabulary and Grammar		Multiple choice	15%
IV – Cloze or			Multiple choice	10%
Short Answer Question			Fill-in-the-blank	
Or Translation			Fill-in-the-blank	
V - Writing			Essay writing	15%

Table 2

CET-4 After January 2007 (125 minutes)

Sections	Content	Sub-Contents	Format	Percentage
I - Listening	Conversation	Short Con.	Multiple choice	35%
	(Con.)	Long Con.		
	Passage	Passage	Multiple choice	
	Compound Dictation	Compound Dictation	Fill-in-the-blank	
II - Reading	In-Depth Reading	Passage	Multiple choice	35%
		Vocabulary	Banked cloze	
		Or Short Answer Questions	Fill-in-the-blank	
	Fast Reading	Passage	True or false Fill-in-the-blank	
III – Cloze	Cloze or		Multiple choice	10%
	Error Correction		Error correction	
IV – Writing and	Writing		Essay writing	15%
Translating	Translation		Fill-in-the-blank	5%

Prior to releasing the reformed CET-4, the MoE called for a pilot study in 2005 which included 180 universities, chosen on voluntary basis. Each of the universities was free to select one of the four textbooks and software cited above. The MoE appropriated 20,000 RMB (around \$2,929) for each university to purchase the textbooks and software

needed. In the pilot study, all participating universities were to reform their college English teaching syllabus, and adjust their teaching mode, collect data, and evaluate advances in student English proficiency through surveys. They were to poll teachers' and students' responses to the reform, upgrade software and hardware facilities in their language classrooms, and adjust teaching loads as well. The reformed CET-4 was also administered in the students of these 180 universities in June 2006. Upon determining that the reformed CET-4 as an achievement test was eliciting intended washback effects, the reformed CET-4 was then adopted nationwide as of January 2007.

CHAPTER 4

METHODS

Research Methods Used in Prior Washback Studies

The review of washback studies in the Chapter Two shows that a number of research methods have been employed in various contexts. The most widely adopted methods are surveys, interviews, document reviews, and classroom observations. However, results vary widely even when the same methodology is used. Interviews, for instance, are conducted in almost all of the empirical washback studies published so far (Cheng, 2005; Green, 2007; Huhta, Kalaja, & Pitkanen-Huhta, 2006; Qi, 2005; Shih, 2007; Shohamy, 1996; Wall & Alderson, 1993; Watanabe, 1996). However, the interview participants differ: Qi (2005) interviewed ten teachers, ten students, eight test constructors, and six English inspectors. Wall and Alderson (1993) and Watanabe (1996) interviewed primarily two teachers. Shih (2007) interviewed the department chair, fourteen to fifteen students, two to three teachers and three students' family members in the two departments in his study. The interviews were either semi-structured or structured, and almost all of them were audio-recorded. Besides these one-on-one interviews, Green (2007) administered focus group interviews to 21 course directors and teachers, as well as 41 students, and Qi (2005) conducted group interviews.

As to surveys, Qi (2005) conducted 976 student surveys and 378 teacher surveys after conducting the interviews “to see how the interview results could be applied to a larger group of participants” (p. 147). Green (2007) surveyed 476 students in eight institutions and teachers in four universities. Like many researchers, Shohamy (1996) administered one student survey once, while Cheng (2005) collected student and teacher

surveys twice to compare attitudinal differences before and after the Hong Kong Certificate of Education Examination (HKCEE). Shohamy (1996), Shih (2007) and Manjarrés (2005) complemented their studies with a review of relevant documents. Most studies are cross-sectional, but the Sri-Lanka O-level study is a two-year examination impact study (Wall & Alderson, 1993). Cheng (2005)'s HKCEE washback study is a three-year longitudinal study. Both included baseline study lasting approximately one year.

To further explore washback effects, Andrews, Fullilove, and Wong (2002) used a parallel UE test, and Green (2007) constructed Tests of Grammar and Vocabulary to measure participants' academic progress. Huhta, Kalaja, and Pitkanen-Huhta (2006) used oral diaries to track the experiences of participants. Another increasingly popular method in washback studies is classroom observation. For the O-level impact study, Wall and Alderson (1993) hired seven local teachers to observe a total of 105 classes. Watanabe (1996) observed two intensive language courses taught by two teachers in a private cram (test preparation) school in Tokyo. Menjarrés (2005) conducted five ethnographic classroom observations and Cheng (2005) observed four lessons by three teachers. Shih (2007) twice observed classroom teaching with two teachers; he also observed the self-study center at various times, as well as eight weeks of teachers' meetings.

Observations are useful according to Wall and Alderson (1993), what participants claim in surveys or interviews is not always consistent with what is observed in the actual classroom setting. However, they do not suggest that classroom observations are methodologically better than other techniques. Bailey (1996) encouraged washback researchers to include classroom observation, but she did not consider classroom

observation a 'must' methodology for washback studies. Instead, as Wall and Alderson (1993) have stressed, since classroom observations take place in specific contexts, the absence of observations does not invalidate a washback study. Moreover, each methodology has its advantages and disadvantages. Whether one methodology needs to be adopted or not depends mostly on the nature of the study and the research questions in that particular context. The availability of funding, energy and resources is another consideration, because not every researcher can afford to hire seven observers. Nor does every researcher have the time to conduct a study for more than two years. What is more, due to the sometimes intrusive nature of classroom observations, most studies schedule only two or three classroom observations on average. Compared to the time spent by students in class every semester, two or three classroom observations do not amount to a very impressive record.

Furthermore, almost all of these classroom observations are video-taped and when a video camera is filming, both students and teachers tend to behave differently. Even when researchers are sitting among the students quietly taking notes, they may also affect and distort the real classroom picture. Usually classroom observations are prescheduled so that the teachers have time to prepare, but when it comes times actually to teach, they may be too nervous to teach effectively, or they may teach more attentively than usual. Having been observed in class, the author of this paper can testify that, more often than not, the observation had an effect on her teaching style. Thus, the extent to which the results of classroom observation accurately represent the actual characteristics of classroom teaching remains in question.

Since the inclusion of classroom observations does not guarantee more accurate results, it is not a requirement for washback studies. As previously stated, the methodology must fit the chosen research questions within a given context. There is no best method, although some methods are more suitable than others for specific research projects. Each method provides a different perspective on the research question. As illustrated by many of the washback studies cited above (e.g. Cheng, 2005; Manjarrés, 2005, Qi, 2004; Shohamy, 2001; Wall & Alderson, 1993; Watanabe, 2004), mixed-method approach, is the one most commonly used.

Methodological Framework

The use of mixed methods is increasingly attractive for multi-faceted applied research nowadays. In the words of Creswell, Plano Clark et al (2007), “Using both numbers and words [researchers] combine inductive and deductive thinking” (p.10). For instance, in-depth, one-on-one, contextualized interviews provide significant complementary data to the relatively decontextualized close-ended data of multiple-choice surveys. Together, the two approaches can bring a fuller understanding of the question than either method employed alone.

Based on the prior methodology review and the goals set for this research, the author of this paper determined that the research questions posed in this study would be best answered by using both avenues of inquiry, given the strengths and weaknesses of quantitative and qualitative methods. Thus, the mixed methods approach was adopted for this washback study.

Specifically, this study employs an explanatory design: participant selection model (Creswell & Plano Clark, 2007). The explanatory design is a type of sequential,

two-phased, mixed-method design in which both quantitative and qualitative approaches are used. It is sequential in that a quantitative survey is followed by qualitative research methods. In fact, the qualitative study is conducted on the basis of the earlier quantitative study (Onwuegbuzie & Teddlie, 2003). “The overall purpose of this design is that qualitative data helps explain or build upon initial quantitative results” (Creswell et al, 2003 as cited in Creswell & Plano Clark, 2007, p. 71). For the explanatory design: participant selection model - unlike the other variant of the explanatory design, the follow-up explanations model - the main purpose is not to explain the significant differences identified in the initial quantitative data analysis, but to understand purposefully selected participants through a follow-up qualitative study, as well the macro study context. Compared to the explanatory design: follow-up explanations model, the explanatory design: participant selection model places more emphasis on the follow-up qualitative study than on the initial quantitative study. Of course, both quantitative data and qualitative data are utilized collectively to offer insight into the research questions.

In the first phase of the explanatory design: participant selection model, surveys are used to efficiently gather large amounts of data in a short time and select participants for the next phase. The second phase makes use of follow-up interviews to clear up ambiguities from the survey results and complement the survey data with more interview data. The current study investigates aspects of teaching and learning that have been influenced by the CET-4, as well as the scale of the influence – all from the student’s perspective.

It is well-known that students may have quite different feelings about a class, even when they are taught by the same teacher using identical activities. This is especially true when students are adults who have been taking courses for 18 years or longer, and who come from provinces/cities in different parts of China. In the follow-up interviews for this study, lengthy and candid responses were required in order to determine whether positive washback occurred after the national reform of the CET-4. If such positive washback did occur, in what ways did it happen? If not, what specifically went wrong, and what actions should be taken to improve English learning for future college students in China?

The oral diaries utilized by Huhta, Kalaja, and Pitkanen-Huhta (2006) in their washback study of Hungarian students served as an inspiration for the author: students participating in Phase IV were asked to make oral self-recordings. These were done at their own convenience, and were used to describe study method, particularly regarding preparation for the CET-4, and to record their reflections about feelings and activities associated with their preparatory activities.

Diary-keeping has long been considered one of the best and most appropriate means to capture learners' personal thoughts and actions in TESOL (Bailey, 1990). However, the Chinese college students in this study were too busy to keep daily diaries. They usually were not willing to write regularly over a period of time in addition to completing their daily homework assignments. An oral diary seemed to be the best alternative, because students in general were interested in technology, and the oral diary required them to simply turn on tape recorder when they felt like expressing themselves. It imposed no extra burden on them, as there was little time required and little to do to

make the recording. It could be recorded anytime and anywhere so long as it did not disturb others. It could even be done spontaneously. The technology allowed students to play back recordings, erase them easily if the recording was not to their liking, and start a new one immediately. What is more, portable digital voice recorders were easy to carry and allowed the data to be sent via email. Unlike traditional diaries, there is no need to examine the physical notebooks or to try to decipher students' handwriting at the time of transcription. Rather, the digital data is simply stored on laptop, ready to be replayed whenever time is available. Furthermore, not only did the diary convey content, it also communicated the emotional tone of the speaker.

As mentioned in Chapter 1, 343 students in nine classes from the three participating universities took part in the survey during Phase I. In Phase II, ten students from each university (eight planned to take the CET-4 in June, and two who planned to take it in December) were selected from among the Phase I volunteers who expressed their willingness to participate in the follow-up interview. The selections were based upon gender, major, English classes registered, and English exam grades at matriculation. In Phase III, a second, slightly changed survey (see p. 240) was administered to 71 students in two classes at University A, and again in November and December, since these students normally took the CET-4 in December. In Phase IV, four of the Phase III volunteers from each class were selected to participate in the follow-up interview, based upon the same criteria used in Phase II. What is more, the six students who were interviewed in Phase II, who were going to take the CET-4 in December 2010 were asked at this time to complete the survey and to be interviewed again. All 14 students (eight new and 6 old participants) in Phase IV were asked to keep journals or self-recordings

about their CET-4 test-taking experience. In the end, 414 surveys (5 participants were surveyed twice) and 34 interviews (6 participants were interviewed twice) were conducted, and 7 self-recordings were collected. Please see Appendix J for an overall timeline of the study.

The different types of data in this study triangulate with each other methodologically. On average, two students from each of the three classes in the three participating universities were selected for follow-up interviews. Thus, for example, the influence exerted by a certain teacher's classroom style, could be mitigated by observing differences among the students in terms of washback. Like Cheng's (2004) study, both within-method and between-method methodological triangulation (Brannen, 1992) were utilized in this study – the former because the methods in Phase I and Phase II were repeated in Phases III and IV, and the latter, because both quantitative and qualitative methods were used to elicit data from the participants. Moreover, researcher triangulation (Cheng, 2004, p. 72) was employed as well, because besides the author's participation, the Zhong Jian Ren (a teacher who serves as a liaison for other teachers) at University A helped collect data in Phase III and IV.

Although the purpose of this study is not to generalize findings, the use of triangulation and mixed methods allow the strengths from each methodological approach to be employed to explore the research questions from different perspectives. The criteria for the selection of participants for Phase II and part of Phase IV of this study are demographic characteristics such as gender, major, birthplace, overall length of English study, grade in the matriculation English test, English learning goals, and the extent to which the CET-4 has influenced the participants' study.

Originally, when designing the blueprint for this research project, the plan had been to conduct a two-phase study in just one university, using mixed methods. In Phase I, to address some general issues regarding college English teaching and learning and CET-4 preparation, a relatively large scale quantitative survey was to be conducted among students who were about to take the CET-4. As part of an explanatory design: participant selection model, this stage was to provide a cohort for the recruitment of participants in Phase II, using interview data, essay writing, emails and self-recordings to explore the extent of the washback effects of the reformed CET-4 on the students' perspectives. However, when it became evident that, in response to the College English Teaching and Learning Reform plan, every university had instituted a different set of teaching practices, the study was then changed to include three universities.

In order to engage more than one university for this study, a personal network of acquaintances was called upon for help. Before studying abroad, having taught English in a university near Shanghai for 4 years, the author had friends and former colleagues with connections to the various universities in Shanghai. Because in China, *guan xi*, or connections, is very important when asking for assistance from strangers, these friends and former colleagues were asked to introduce the author and the study to several universities, via letters or phone calls. Of the five recommended universities, three showed an interest in participating in the research. Based on these connections, formal letters were sent to the deans of the English Department, or English professors who were in charge of the College English division. All replied and granted permission to work with their students. Introductions were arranged through a *Zhong Jian Ren*, or through offers to make direct contacts. Although the original plan did not involve a comparative

study of washback effects at different institutions, the distinct rankings of the three universities promised to provide a far richer picture of the washback mechanisms than the original plan had envisioned.

The three universities in Shanghai were thus selected based on convenience sampling. As the biggest and most populous city in China, Shanghai exerts significant influence on politics, commerce, culture, and education both nationally and globally. Due to its location and rich resources, it is the home of more than 30 universities and colleges. Thousands of students are attracted by its highly-regarded institutions of higher education every year – particularly, science- and engineering-majors – because Shanghai has long been the major industrial and financial center in China. The following brief description will provide further information about the three universities involved in this study, which is the micro context of this research.

Sites for the Study

The universities involved in this study, though selected on the basis of convenience sampling, represent three levels of prestige in the Chinese hierarchy of higher education – highly reputed, intermediate and ordinary. According to “China 2010,” (2010) a website giving the latest ranking of universities in Shanghai city, all three were ranked in the top 25, but each was in a different tier – two were in the top 15 and one in the top 25. On the same website, nationally, two of the three universities were ranked among the top 150 in China in terms of the number of prestigious scholars, research resources, publications, student quality, teacher quality, and availability of tangible resources, such as research funding, library resources, and university size. What is more, all of them are science- and engineering-oriented. Overall, however, University

A is a more comprehensive university. One of the oldest universities in China, University A is a nationally famous engineering university. College English is one of the divisions that falls under the School of Foreign Languages, along with English and several other foreign language departments. It also houses the Administrative Office for the NCETC (National College English Testing Committee) and many key committee members of NCETC either taught, or are still teaching, at University A.

With more than 100 years of history, University B places great emphasis on engineering and science. The branch campus visited by the author is the only one among the three universities to provide only first-year education primarily for engineering, economics, and management majors. After one year's study, qualified students then proceed to corresponding colleges on the main campus to pursue their majors. College English is a subdivision under the School of Foreign Languages, along with English and several other foreign language departments.

University C was at one time an Open University. It did not become a full-time, regular university until the 2000s and it is now well-known for its vocational education program at the tertiary level. With its history of vocational education, it specializes in practical and advanced vocational education such as engineering and electronics. As in University A and B, College English is a division under the umbrella of the School of Foreign Language, along with English and several other foreign language departments.

In all three universities, College English does not stand alone as a separate department, but forms an integral part of the language department which offers fundamental English courses for all freshmen and sophomores who are not English majors. Their undergraduate populations range from 10,000 to 20,000. Moreover,

documents indicate that all three universities applied for, participated in, and successfully passed the evaluation conducted by the MoE after participating in a one-year pilot study of the College English Teaching and Learning Reform from February 2004 to February 2005. During the Reform, University A adopted the textbook New Horizon College English and its software by Foreign Language Teaching and Research Press, while universities B and C adopted the textbook New College English and its software by the Shanghai Foreign Language Education Press.

Because receiving a higher education is still a privilege in China, the competition in matriculation testing is very fierce. The higher the ranking of the university, the higher the matriculation test score required for acceptance at the university. So the three universities, each with its different ranking, have developed distinctive student bodies, and equally distinctive English teaching practices in response to the reform of the CET-4.

University A

In University A, general English courses are offered in the first year. All freshmen are given an English proficiency test during their orientation week. These proficiency tests are designed by the English teachers of each university. Students are then divided into fast classes and ordinary classes based on their grades on the test. A general English course is offered twice a week using the same textbook, New Horizon College English. English Listening is a separate class taught in the language lab, but it is an integral part of the mandatory general English course. English teachers are assigned to students in the first semester; during subsequent semesters, students can pick their own English teacher when registering for classes.

In the second year, instead of continuing the courses with New Horizon College English as most other universities do, English courses on a variety of topics are offered to cater to the interests of students. Students can select among Public Speaking, English Translation, English Literature, English Interpretation, Business English, and Audio-Visual-Oral English unless passing the university's internal English proficiency test. Each student is supposed to take one of these courses and pass it in order to fulfill the language credit requirement for graduation. Each topic may be offered by more than one teacher. However, teaching methods and assessments may vary, as each teacher is free to use his or her own style of teaching in the classroom. Some even write their own textbooks and have them published. Of course, teachers who share the same topic may use the same syllabus.

According to University A's policy, although students from the fast class are taught more content at a faster pace, all students without exception must wait until December of their second year to take the CET-4 for the first time. Actually, the fact that the normal time for students at University A to take the CET-4 is in December, and not in June, became an important factor in the decision to add Phases III and IV to the study. It also enabled me to follow-up with the six students in Phase II who planned to take the December CET-4.

University B

At University B, all freshmen live and study at a branch campus for one year in order to complete all the basic courses before moving onto the main campus in the second year. This branch campus was in a suburb quite far from the downtown area. In fact, the branch campus was part of a large college town with branch campuses of several

other universities established nearby. Students are expected to have a quieter learning environment in the suburbs so that they can concentrate more keenly on studying. Of course, cheaper land prices in the suburbs also enables popular universities to expand and accommodate more students.

All freshmen at University B are also divided into fast classes and ordinary classes based on an English proficiency test taken during orientation week. Although both classes use New Horizon College English as their textbook, unlike University A, students in the fast class at University B are allowed to take the CET-4 in June of their first year, whereas students in the ordinary class must wait until their third semester, or December of their second year. Logically, the university wants to give more time to students in the ordinary classes to build their English foundation and prepare for the CET-4. This also means that students from the ordinary class will take the CET-4 at the main campus in December, because by that time they will have moved to the main campus.

The English courses offered at University B are different from those at University A. At University B, English Listening is not a separate class but it is an integral part of the mandatory English courses. What is more, all students are required to take two years of college English courses. In the second semester of the first year, students can register for elective courses such as Oral English and CET-4 Ability Practice, but the offerings are more limited than that at University A. The elective course is usually a one-credit course offered once a week. Of the three universities, University B is the only one that openly offers a CET-4 preparation course for credit.

University C

At University C, likewise, students are divided into fast classes and ordinary classes based on grades received on the university's own, specially-designed English proficiency test, which is administered before the official start of the first semester. The textbook used in the ordinary class is New Standard English, while the fast class uses New College English. Interestingly, the first textbook is not among the ones recommended by the MoE. The fast class skips general English Level I and starts directly at Level II. Their mandatory English course is composed of Comprehensive English and English Listening classes. The former is held twice a week. The latter is held once every two weeks in the language lab, where students mainly practice English listening independently, using pre-installed software. Teachers do not lecture in Listening classes, but they are available in the classroom to answer questions or sometimes to give and collect assignments.

At University C, like University B, students from the fast classes can take the CET-4 in June, their second semester of their first year, while students from the ordinary classes are required to wait until December, which is the third semester of their second year. Other than the mandatory general English courses, no elective English courses are offered at University C. However, the English Department offers an after-class private intensive CET-4 preparation course. This course is held twice a week in the evenings on campus. Registration is voluntary for this course and the registration fee is 240 RMB (around \$35). It can accommodate more than 100 students.

Questionnaire Construction

Although permission was granted by three universities to conduct this washback study, interviewing all the students who planned to take the CET-4 in June was simply not practical. This problem was solved by making use of surveys, as had been done in prior washback studies. “[S]urveys can [provide] an economical and efficient means of gathering a large amount of data from many students” (Creswell, 2008, p. 87). According to Weisberg, Krosnick and Bowen (1996), surveys are best used to address the following issues, “a) the prevalence of attitudes, beliefs, and behavior; b) changes in them over time; c) differences between groups of people in their attitudes, beliefs and behavior; d) causal propositions about these attitudes, beliefs, and behavior” (p. 15).

As will be demonstrated below, the research questions in this study are well suited for collection of data through the use of surveys. The questions at the core of this study are:

1. What are college students’ beliefs about English learning; what are their expectations and experiences of studying English at college, and what are their perceptions of the reformed CET-4?
2. What plans did the students make to prepare for taking the CET-4, and how did they implement their plans?
3. To what extent has the reformed CET-4 influenced students’ perceptions of the practices of English teaching and learning at the college level?

These questions address students’ attitudes, beliefs, and behavior regarding college English teaching and learning in the context of the reformed CET-4. A well designed survey can capture a general picture of the opinions of student respondents.

Moreover, compared to other methods, an anonymous survey protects students from identity, allowing students who are shy or who hold ideas that are against the norm, to have their voice be heard without feeling pressured or uncomfortable. Another advantage of a survey is that it can help recruit participants for follow-up purposes after completion of the survey. This was not considered the most practical way to recruit for the next phase, since the students were not known to the author, nor was it appropriate for teachers to recommend particular students. Thus for this study, students who wanted to participate further were asked to leave their name and contact information in a designated area on the survey form. Selection of Phase II (interview) participants was made at a later date, using background information that students had provided on the survey form.

Because the purpose of the student surveys was to elicit opinions from a large number of students in the three universities, as related to research questions specific to this study, a great deal of time and effort went into composing the student survey. As mentioned above, a number of researchers have used surveys in their washback studies (e.g. Cheng, 2005; Green, 2007; Shohamy, 1996), so these surveys were first carefully examined as reference. The surveys could not simply be borrowed, since the surveys were each designed to answer different research questions in specific contexts. In spite of this, the rationale, structure, and content of each survey provided useful input for the survey design for this study. For instance, Green (2007) included questions regarding students' prior English learning experience, their future plans, and their general attitudes towards studying. Both Cheng (2005) and Green (2007) asked students about their attitudes towards English learning, English learning strategies, English course taking experience, and classroom teaching and learning activities.

These were questions that were also relevant for this study. Of course, Cheng's survey (2005) was specifically designed to study washback effects of the HKCEE on students in Secondary Five in Hong Kong. Probably partly because her participants were secondary-school students who were not mature enough to express their thoughts in writing, no open-ended questions were included in her survey. Green (2007) and Cheng (2005) also complemented student surveys with teacher surveys. Green (2007) even administered two more brief IELTS Awareness Forms: Form A, which asks mainly about examinees' test knowledge before the IELTS Academic Writing Module (AWM) and Form B – which asks about English learning experience outside the classroom when the course is about to end. In addition, he designed a Test Strategy Report (TSR) given after the IELTS AWM.

However, in this study, participants were all adults with at least 6 years of English learning experience before entering college. Their busy schedules would not allow for surveys to be administered more than once, but they were sophisticated enough to provide brief reasons for their answers in addition to factual information. Hence, I designed a survey that was appropriate to this particular research context, with the core points of all available surveys absorbed, the literature investigated, and consultations arranged with college level English teachers. Finally, the author's own experience as a college English teacher as well as an English learner was incorporated into the mix, as well as her experience as a native Chinese English learner.

The survey was subjected to multiple revisions. The initial survey contained 6 themes: students' demographic information, prior English learning experience, parents' involvement, present in-class and out-of-class college English learning experience,

students' understanding and plans for the CET-4, and personal questions relating to learning English, such as motivation, goals, and future plans. Survey formats included multiple-choice questions, multiple-choice with brief written answers, and open-ended questions. The first draft of the survey was sent to an English teacher at University B for advice, since this teacher had earlier offered to help with survey revision. Having taught College English for almost 20 years, she was very supportive of the study. She suggested adding a brief introduction that explained the purpose of the survey. She showed the author a sample survey she had designed and administered in a previous semester to collect data on college students' oral English proficiency. She stressed that it was necessary to assure the students that their participation was voluntary, and that it would not influence their grades. She was also concerned about teachers' tight class schedule, so she suggested that some short-answer questions should be changed to multiple-choice so that the survey could be completed in 15 minutes. Because of her first-hand experience with survey research on college students, her suggestions resulted in modifications to the original survey.

The teacher then arranged for a meeting with an English professor from University A to further polish the survey. This professor is a well-known expert on research methods and a current research consultant for the College English CET Committee. From October 1992 to December 1995, before the reform of the CET-4, he had helped secure approval for the CET-4 from experts from the University of Reading and the College English CET Committee, a project that was sponsored by the British Council and the China State Education Commission. He offers courses in Survey Research, Qualitative Research Methods and Quantitative Research Methods for both

undergraduate and graduate students. As a result, he is not only experienced with survey design, but he is equipped with solid knowledge about the CET-4 and quantitative research in applied linguistics. The research plans and survey draft was sent to him for further consultation.

After reviewing the survey very carefully, he offered invaluable suggestions for revisions. First, he pointed out that parents no longer played an important role in students' college studies, because students did not live at home. In fact, unlike Green's (2007) students who came from either the day division, night division or the weekend division, the participants of this study were full-time students who lived in the dorm. Furthermore, in the professor's experience, parents in general were more concerned with their children's grades of their major courses, so questions concerning parents' attitudes and behavior should be deleted.

Moreover, the research goals proposed for this study were not directed toward finding out parent factors influencing the CET-4. Therefore, he suggested deleting questions concerning parents' attitudes and behavior. He also suggested that the subtleties of some questions could be discerned through the use of a likert-scale to measure the range of feelings. Also, to save time and avoid incomplete answers, he suggested using questions that required brief answers in the follow-up qualitative study. This way, the data analysis for the survey would be neater.

As a result of the professor's suggestions, all the questions relating to the participants' parents were removed and questions relating to the students' English classroom teaching and learning experiences were added using the likert-scale. These questions addressed the types and frequencies of various activities in the classrooms, the

extent to which English was used in class, and the frequency of out-of-class English activities. Only 4 open-ended questions remained so that students would be more likely to complete the whole survey.

With the revised Chinese version of the survey in hand, the Zhong Jian Ren at University A was contacted to request assistance for finding a few students to take the survey on a trial basis. This was also an opportunity to check whether they understood every question. I also wanted to check whether the students could finish the survey within 15 minutes. The trial was conducted in the back of the classroom during one of her English classes while the other students were taking a dictation quiz. The two volunteers were both engineering-related majors. One was planning to take the CET-4 in June and the other in December. Both finished the survey within 15 minutes. After class, they were individually questioned about their opinions of the survey. Their responses were generally positive, although one student suggested adding ‘participating in English contests’ as one of the out-of-class activities, because he and some of his classmates had recently participated in a contest and had learned a lot from it. He also said that he had already taken the CET-4 training, thus could not find an appropriate choice for the yes-no question “Are you planning on taking a training class?” All the corresponding revisions were made.

After going through these multiple revisions, the final version of the survey was reduced from four to three pages. It was written in Chinese so that misunderstandings caused by language would be minimized. Five major themes were embedded in the final version:

1. Demographic information;

2. Prior English learning experience, such as how long they have been learning English, and their English grades in the entrance exam to college;
3. Understanding, attitudes and plans with respect to the reformed CET;
4. The influence of the reformed CET-4 on the students' college English teaching and learning, both in class and out of class; and
5. General English learning issues, such as attitudes towards English and their major, overall English proficiency, motivation and future plans.

The questions were categorized into four sections according to question format, with mixed themes in each section. Survey formats include multiple choice, likert-scale, open-ended questions, and semi-closed-ended questions (Creswell, 2008). For semi-closed-ended questions, several options were provided for the participants to choose based upon existing literature. However, suspecting that students might need more choices, as times had changed and the CET-4 test had been reformed, the participants were given the opportunity to check an *Other* option and write briefly what he/she meant by other. This is also an important way to improve washback surveys if a number of participants choose *Other* and provide brief explanations. See Appendix D for the final survey instrument.

Overall Procedures

In Phase I, along with the approval provided by the Dean or Vice Dean of the School of Foreign Languages of each of the three universities, the officials released the name of a Zhong Jian Ren (a teacher who served as a liaison) who could be contacted directly to assist with the study. A request was sent to the Zhong Jian Ren in the three universities asking for assistance in distributing a recruitment notice (see Appendix A)

among the teachers in the English department. The notice stated that teachers' questions or concerns relating to the study would be answered by the designer of the study. Those who were willing to allow their classes to participate then contacted the Zhong Jian Ren who provided these teachers' names, their class schedules and classroom locations where the survey would be administered.

With the teachers' permission, the time and location were set. The students were briefed about the study and then a Letter of Information for the Survey (see Appendix B) was read to the whole class, emphasizing the voluntary nature of their participation. The students were also asked to indicate whether they were willing to participate in a follow-up qualitative study at the end of the survey. Those who chose not to participate were allowed to leave at the time when the survey was conducted. While the students were filling out the survey, the author was available to answer any questions that emerged. The completed surveys were collected at the end of the allotted time. If there was a time conflict, the Zhong Jian Ren helped with the survey administration, following the same procedure as above.

In Phase II, ten students from each university were selected from among the students who had indicated their interest in participating in the follow-up qualitative study. The selection was based upon gender, major, English classes registered, and English grades obtained in the college entrance exam. Of the ten students at each university, eight planned to take the CET-4 in June and two in December.

All 30 students were then contacted individually and asked to sign an Informed Consent Form (see Appendix C) for the follow-up qualitative study. Next, one-on-one interviews, essay writing, and email contacts were scheduled at their convenience. Due to

students' busy schedules, the essays were never completed, but the questions were incorporated into the one-on-one interviews. The original study only consisted of these two phases.

In May 2010, the author learned that she was a recipient of The International Research Foundation for English Language Education (TIRF) 2010 Doctoral Dissertation Grant. With the funding provided and permission of her advisor, the study was expanded into four phases. In Phase III and IV, the six students, two from each university, who planned to take the CET-4 in December were contacted again to fill out a slightly modified survey (see Appendix G). They were also interviewed again and they were asked to self-record their English learning experience, their test-taking preparation, and their overall test-taking experiences, allowing for comparison of any changes, in their English study practice and CET-4 preparation process from one semester to the next. In addition, Phase I and II were administered to two new classes from University A who were about to take the December CET-4. This university was chosen for the follow-up research because December, rather than June, is the normal time for the majority of students to take the CET-4 at this university. After the survey, follow-up with two students from each class was conducted using the procedures in Phase II. The same teachers who participated in June also taught the two new classes.

The survey data were analyzed using descriptive statistics through SPSS 18. The follow-up qualitative study data - individual interviews, students' self recordings, and emails regarding their English learning and preparation for the CET-4 – were transcribed and coded based on Grounded Theory approach. When necessary, ambiguous information was further clarified by emails or phone calls to the participants concerned.

Phase I

Once the survey was finalized, the author contacted the three universities and scheduled dates for administering the survey. In universities A and C, the dean of the English Department introduced a Zhong Jian Ren to help with the survey administration. In University B, the Zhong Jian Ren who helped with the survey revision assisted with administering the surveys.

Because different universities have different practices in response to the reform of college English teaching and the CET-4, students of various kinds were included as survey participants. In Chinese universities, there are usually two semesters per year. The fall semester is from September to the end of January. The spring semester is from March to July. In between are winter vacation and summer vacation. In University A, as stated earlier, students are not allowed to take the CET-4 until December of their second year. So at the time of the survey in May 2010 at University A, the students who completed the survey were either going to take the CET-4 in December 2010, or had taken the CET-4 in December 2009 and were going to take the CET-6 in June.

While considering the course of action, it came to the author's attention that for some reason, a few University A students had not taken the CET-4 the previous December. They were now sophomores and most likely would take the CET-4 in June. However, they were now spread throughout different topic-related English classes and they were not easy to locate. So an email was sent by the Zhong Jian Ren to all English teachers who were teaching second-year students to help find these students and administer the survey. Soon one or two students were found in different classes. Eventually, 29 students who were going to take the CET-4 in June 2010 filled out the

survey. They were taught by seven different English teachers and most of these students had not taken the CET-4 the previous December because they had either missed the registration deadline or were late for the test on the testing day. Given that December is the required month for taking the CET-4 for the first time at University A, the decision was made to survey two additional freshman classes that were going to take the CET-4 in December 2010. So later, 73 students – taught by two different English teachers – who were going to take the CET-4 in December 2010 completed the survey, too. Both classes were composed of students from different majors. Although they were not scheduled to take the test until the next semester, their answers could still reveal whether the CET-4 had influenced their college English teaching and learning by that time.

At University B, the survey was administered not only to students who were going to take the CET-4 in June and December 2010, but also to students from all different types of English classes. The survey was conducted in three classes taught by three different teachers: there were 55 students from a CET-4 Ability Practice class, 31 students from one fast class of Comprehensive English, and 39 students from one ordinary class of Comprehensive English. In University C, the survey was conducted in one day. Since they did not have elective English courses, 41 and 39 students, respectively, from two fast English classes, and 36 from the ordinary class completed the survey. The students in ordinary classes in universities B and C were participants who would be taking the CET-4 in December.

The response rate for each university was 100% since none of the students in any of these classes refused to complete the survey. However, after reviewing the completed surveys, fifteen had to be discarded due to incomplete responses on one or more pages;

one of the discarded surveys was from University A, four were from University B, and ten were from University C. The final response rates were 99%, 96.7%, and 91.4% for University A, B and C respectively. The number of male participants was almost twice that of female participants. This was mainly due to the fact that all three universities were engineering- and science-oriented and it is still common to find more males than females in these professions. By contrast, all the teachers involved were female. This was not surprising either, because the majority of English teachers in English Departments nationwide in China are female. The majority of students in the English Departments are female as well. As to training, all of the teachers involved in the study had master's degrees or doctoral degrees in English-related fields, and none were novice teachers. In University B one male teacher's class could have been included in the survey, but unfortunately he turned down the request because he felt his class schedule was too tight.

Phase II

After administering the survey, survey completed from each University was carefully reviewed. A number of students filled in their contact information, indicating that they wanted to participate in the follow-up interviews (see Appendix F). The research design called for the selection of ten students for Phase II from each university. Among those who volunteered, students for Phase II were selected based on demographic information such as gender, birthplace, and major, as well as their general response to the CET-4, their English grade on the matriculation test, CET-4 preparation plans, future plans and their CET-4 test date.

At University A, the ten selected would be taking the CET-4 in June 2010. Since the survey was also administered in two classes that would take the CET-4 in December,

one student from each of these classes was selected as well. In order to maintain this balance, the same process of selection was employed for Universities B and C. At University B, 4 students from the fast class and 4 students from the CET-4 Ability Practice Course were selected. All 8 students would be taking the CET-4 in June 2010. As mentioned above, 2 more students were selected from the ordinary class, both of whom were scheduled to take the CET-4 in December 2010. Eight out of ten students later participated in the follow-up qualitative study since two withdrew before the interview was conducted.

University C, however, did not offer the CET-4 Ability Practice course and the main focus was on students who would take the CET-4 in June. So 4 students in each of the two fast classes and 2 were from the ordinary class were selected. The first set would take the CET-4 in June, and the latter would take it in December 2010. However, at the time that the interviews were being scheduled, one student was unable to participate because he accepted a part-time job. Thus, 9 students in University C participated in Phase II.

The total numbers of students who participated in Phase I and Phase II are as follows:

Table 3

Participants in Phase I and Phase II

University	Class Type	Phase I		Dropped	Phase II	
University A	CET-4 in June	29	3 Female	0	10	2 Female
			26 Male			8 Male
	CET-4 in Dec. Teacher I	38	5 Female	1	1	0 Female
			33 Male			1 Male
	CET-4 in Dec. Teacher II	35	5 Female		1	1 Female
			30 Male			0 Male
_____	Total	102	_____	1	12	_____
University B	CET-4 Ability Training Practice Class	55	21 Female	1	2	0 Female
			34 Male			2 Male
	Fast Class	31	0 Female	1	4	2 Female
			31 Male			2 Male
	Ordinary Class	39	1 Female	2	2	1 Female
			38 Male			1 Male
_____	Total	125	_____	4	8	_____
University C	Fast Class Teacher I	41	9 Female	3	4	3 Female
			9 Male			1 Male
	Fast Class Teacher II	39	8 Female	4	3	1 Female
			31 Male			2 Male
	Ordinary Class	36	11 Female	3	2	1 Female
			25 Male			1 Male
_____	Total	116	_____	10	9	_____
Phase I & II	GrandTotal	343	63 Female	15	29	11 Female
			227 Male			18 Male

Phase III

Due to lack of funding, students' self-recordings were not included in Phase II as planned. In the meantime, it became evident that the main cohort of students would normally be taking the CET-4 in December, rather than in June in University A. Thus, in May at University A, those students who were expected to take the CET-4 in June were second-semester sophomores. They sat in the same classes as students who had passed the CET-4. By this time their teachers were generally more concerned with the CET-6. So most survey participants at University A who had missed the testing time in December 2009 had to prepare for the June 2010 CET-4 on their own. Some of them were too shy to tell their classmates that they had missed the December test. So their recollection of what happened in the previous semester may not be precise because they had to concentrate on their current classwork. In this case, their attitudes towards the CET-4 may have changed since the previous semester. To resolve this situation, expanding the study to Phases III and IV became an option for consideration.

Fortunately, The International Research Foundation Doctoral Dissertation Grant (TIRF DDG) was approved in September 2010. With this financial support, the expansion of the research into Phases III and IV became a possibility. With my advisor's approval, Phases III and IV were implemented in November and December among students from University A who were planning to take the CET-4 in December 2010. In addition, six more students (two from each of the three universities) from Phase II of the study who would take the CET-4 in December 2010 were followed up.

With IRB approval of the expanded study, Phase III got under way. The cost of going back to China again to collect the additional data was prohibitive. However, since

the Zhong Jian Ren at University A had been so helpful throughout Phases I and II, her assistance was again solicited. She agreed without any hesitation after understanding the situation and the plan. She explained that the two teachers who taught freshmen English in the previous semester were now teaching sophomore classes that would be taking the CET-4 at the end of current semester in December; this would be a good opportunity to track these teachers' classroom behaviors. She agreed to contact the two teachers to conduct another survey with their current students. The two teachers kindly allowed the study to be administered to their students once again. Zhong Jian Ren was given a slightly modified version of the previous survey. She scheduled the survey time and classroom location with the two teachers at University A and administered the survey. Forty-five students from the class of teacher I, and twenty-six from the class of teacher II completed the modified survey at this time.

Aside from changing the date that the students were planning to take their CET-4, the modified survey (see p.240) included several substantive changes. An open-ended question was added to the survey because the students in Phase I and II had not given an explicit answer. The new question asked: "What do you think the connection is between the CET-4 and your College English course? Please explain briefly." Also in the section on demographic information, the question on place of birth was eliminated. Instead, because in Phase II some participants indicated that they had matriculated from a different city/province from the one in which they were born, the students were now asked where they had taken the matriculation test.

This distinction was necessary because, depending on the testing location, college admission requirements and educational resources differ sharply. For instance, students

who are not residents of Shanghai may be required to score twice as high on the matriculation test residents of the city, in order to be accepted into top universities and popular majors.

Phase IV and Brief Summary

Four students who expressed their willingness to participate in the qualitative follow-up study were selected from volunteers in each class in Phase III. Two students from the class of teacher I and three from teacher II eventually showed up for the interviews. The Zhong Jian Ren interviewed these five students based on the interview questions provided to her (see Appendix I). Among them, two from each class agreed to do the self-recordings.

In the meantime, the 6 students that had been interviewed in Phase II were contacted through email or by phone. These students were going to take the CET-4 in December 2010. Although this was not part of the original research plan, they all agreed to participate in Phases III and IV. They all had access to the internet, so the study was conducted through QQ, a free Chinese online chatting software which is similar to Yahoo Messenger and very popular among the Chinese. One can chat live online, leave messages, and call each other for free. Almost every student has a QQ account. This is also why, even in Phases I and II, students were given an option of leaving their QQ number in the contact information, since not everyone could afford a cell phone.

So through emails or QQ, online interviews were scheduled for each of the six students in November 2010. At the beginning of each interview, the student received the modified version of the survey which was completed and returned through email. One student, however, did not return the survey as expected. Then the student participated in a

semi-structured interview that was quite similar to the one given in June. Of course, questions about their prior English learning experience before the time of June 2010 were not included. Instead, the focus of this interview was on the student's English learning experience in the current semester and how it compared to the previous semester. Also, the student was asked how the approach of the CET-4 had influenced their English teaching and learning, if at all. Because the online chat required typing for communication, every interview took about two hours. Of course, the webcam and microphone could have been used, to see and talk to each other during the interview, but some students did not have their own computer or laptop and were using a computer in the language lab. Webcams usually were not installed in the open lab and talking aloud through a microphone in the public lab was not acceptable. So to be fair, all the interviews took place in typed form, and all interviews were conducted in Chinese.

Among the six students, by chance, one student at University A applied for the Computer-Based CET-4 Trial Test. The registration was open online for a short time. This student was lucky and fast enough to get registered before all the seats were filled. So she was the only case who took not only the Computer-Based CET-4 Trial Test but the traditional CET-4 in December, 2010. Her final CET-4 score would be the higher of the two scores obtained from taking the two tests. In the interview, six students agreed to do self-recordings and to send their recordings through email for review. Because their schedules were very tight, every student was asked to self-record only six sessions, preferably twice a week. However, on December 17, the day before the CET-4 test date and on December 18, the day of the CET-4 test date, the recording assignments were mandatory, so that their preparations right before the test and their reflections on the test-

taking experience would be recorded. Because none of them had done self-recordings before, they were given a sample of self-recordings (what type of things to say and how), which included transcripts of three recordings in three different scenarios. For the five students interviewed by the Zhong Jian Ren, each was given a hard copy of the self-recorded samples.

In Phase III and IV, the participants who responded were as follows:

Table 4

Participants in Phase III and IV in University A

Teacher	Survey		Phase IV		Self-recordings	
Teacher I	45	9 female	3	0 female	2	0 female
		36 male		3 male		2 male
Teacher II	26	3 female	2	2 female	1	1 female
		23 male		0 male		0 male
Total	71	12 female	5	2 female	3	1 female
		59 male		3 male		2 male

Table 5:

Participants in Phase III and IV in University A, B and C: A Further Follow-up of Participants in Phase II

University	Survey		Survey Returned		Interviews		Self- recordings Submitted	
University A	2	1 female 1 male	2	1 female 1 male	2	1 female 1 male	2	1 female 1 male
University B	2	1 female 1 male	2	1 female 1 male	2	1 female 1 male	1	1 female
University C	2	1 female 1 male	1	1 male	2	1 female 1 male	1	1 female

Table 6

Participants from Phase I to Phase IV

	Survey	Interviewees	Self-recordings
Female	132	13	4
Male	282	21	3
Total	414	34	7

CHAPTER 5
DATA ANALYSIS

Basic Characteristics About the Participants

Number, Age, Gender, Birthplace and Major

A total of 414 students participated in the survey: 173 from University A, 125 from University B and 116 from University C. Table 7 represents the frequency and percentage of each age group from 17 to 22. Approximately one half of them (51.7%) were 20 years old. In terms of gender, Table 8 shows that 68.12% of them were male (282), and 31.88% were female (132). With the exception of University C, where there are 3.4% more female participants than males, male participants far outnumber females at University A and B. This, however, does not suggest bias in the sampling. Instead, it is a true reflection of the student population of science- and engineering-oriented universities in China. It is also not surprising to see that the ratio between female and male participants at University A is as high as 1:4.3, either, because it is highly competitive for female students to get admitted in primarily male-dominant majors at such a prestigious university. According to A. Q. Wang at University A, there were only five female students in her cohort of Microelectronics majors.

Table 7

Age

	Frequency	Percent
17	1	.2
18	26	6.3
19	125	30.2
20	214	51.7
21	45	10.9
22	3	.7
Total	414	100.0

Table 8

Gender

University		Frequency	Percent
University A	Male	141	81.5
	Female	32	18.5
	Total	173	100.0
University B	Male	85	68.0
	Female	40	32.0
	Total	125	100.0
University C	Male	56	48.3
	Female	60	51.7
	Total	116	100.0

Geographically, all of the eight official administrative districts in China are represented. Nevertheless, because all three universities are in Shanghai, which is in Eastern China, 61.4% of the participants in the study were from Eastern China, as you can see in Table 9. 12.6% were from nearby Central China, and the percentages from other districts are all under 10%. What is more, among the 414 subjects, altogether 97 majors are involved in the study. The top three majors are Telecommunications (11.4%), Environmental Studies (9.2%), and Software Engineering (6%).

Table 9

Birthplace

	Frequency	Percent
Missing Data	1	.2
Eastern China	254	61.4
Southern China	15	3.6
Central China	52	12.6
Northern China	23	5.6
Northwestern China	20	4.8
Southwestern China	35	8.5
Northeastern China	13	3.1
Taiwan, Hong Kong and Macao District	1	.2
Total	414	100.0

Prior English Study Lengths

Among participants who come from different administrative districts, the years spent studying English ranged from 6 to 20 years. As shown in Table 10 below, 23.4% had studied English for 7 years. Another 21.5% had studied English for 10 years. Thus, it can be inferred that, at the time this study was conducted, most of these students had not started studying English until their junior high school (7th grade) or the 4th grade in primary school. The mean duration of English study for all the participants in the three universities is, however, 9.5 years. This reflects the fact that in China since the 1990's, English has been offered as one of the mandatory subjects nationwide beginning at least in junior high school. Also, the data suggest that students who engaged in 7 to 10 years of English study came mostly from Eastern China, Southwestern China and Southern China, which implies that studying English prior to the tertiary level was a greater priority in those districts than in others.

Table 10

English Study Lengths

	Frequency	Percent
0	1	.2
6	7	1.7
7	97	23.4
8	53	12.8
9	39	9.4
10	89	21.5
11	61	14.7
12	29	7.0
13	24	5.8
14	5	1.2
15	5	1.2
16	1	.2
17	2	.5
20	1	.2
Total	414	100.0

Matriculation English Exam Grades in Three Universities

As shown in Table 11 below, a gap in participants' English levels existed prior to college. If test scores are converted to a number out of a total of 100 points, the mean scores of the three universities are 85.70, 78.63 and 76.17 for Universities A, B, and C respectively, which matches the corresponding rankings of the three universities. As the most prestigious university among the three, at University A the lowest English score was 67, which is higher than the lowest score at University B (58) and University C (60). Furthermore, at University A, some interviewees revealed that a number of students at University A were exempted from taking the matriculation exam because they had been admitted to the university on the basis of outstanding performance in math, physics, chemistry, or other international contests in senior high school.

Table 11

Matriculation Statistics of English Exam Grades in the Three Universities Respectively

University A	N	Valid	166
		Missing	7
		Mean	85.70
		Std. Deviation	6.069
		Minimum	67
		Maximum	97
University B	N	Valid	125
		Missing	0
		Mean	78.63
		Std. Deviation	7.135
		Minimum	58
		Maximum	93
University C	N	Valid	116
		Missing	0
		Mean	76.17
		Std. Deviation	7.563
		Minimum	60
		Maximum	96

Primary English Learning Goals at College

Students were also asked what their primary English learning goal was at college. Figure 1 below is the visual presentation of the data. As shown in Figure 1, students at University B (38.4%) and University C (40.5%) were more motivated to get high scores on the CET-4 than students at University A (19.7%), because students at these two universities had the same popularity order of choices to Question 2: for University B, the order and the percentage of each goal was: a (38.4%); d (23.2%); b (16.8%); e (11.2%). For University C the order and the percentage of each goal was a (40.5%); d (29.3%); b (23.3%); e (3.4%). [a – to get high scores in the CET-4; b – to pass required English courses and get high scores; c – to learn English cultures further to satisfy one's own learning interest; d – to improve comprehensive communicative ability in English; e – to

lay a good foundation in English to study abroad in the future; f – other] For their counterparts at University A, however, from high to low, the primary purposes of English learning are; d (38.7%); b (21.4%); a (19.7%); e (17.3%).

So, it seems that the majority of students at University A were interested in improving their comprehensive communicative ability in English, whereas the majority of students at University B and University C were more eager to get good scores on the CET-4.

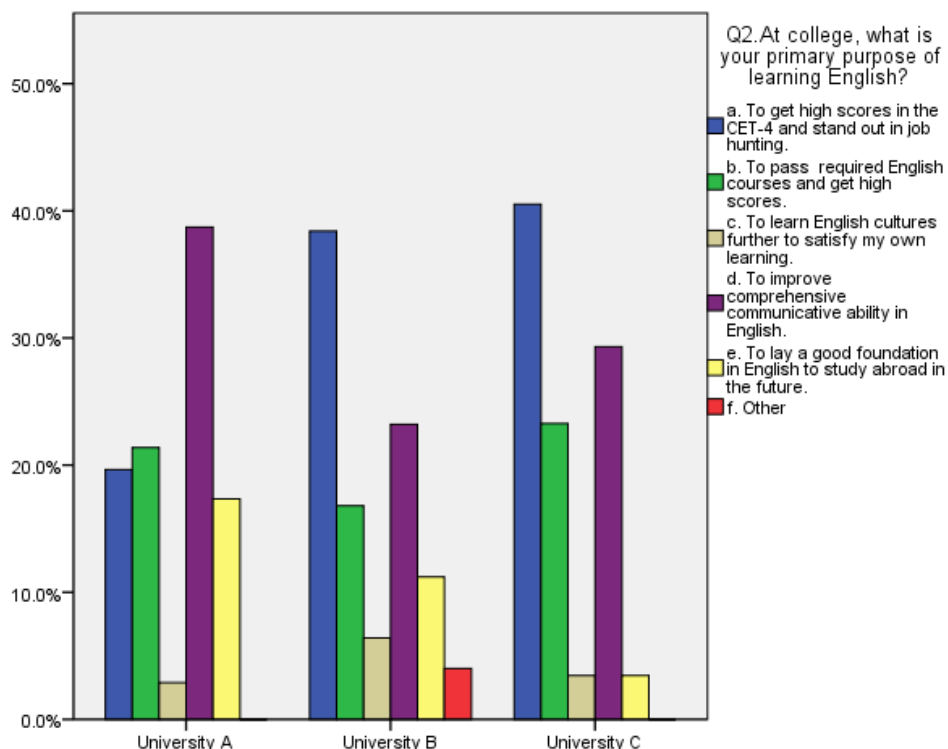


Figure 1. Primary purpose of English learning at college.

College graduation Plans and Perception of the Importance of the CET-4

Student graduation plans at the three universities were also different, corresponding to students' various goals for studying English. Figure 2 suggests that only 26.6% students at University A and 34.4% at University B planned to go directly to work.

However, answers from students at University C, the vocational university, showed that 63.8% preferred going to work directly after getting their bachelor's degree. In contrast, 35.8% of the students at University A and 39.2% of students at University B planned to go to graduate school in China. The fact that at University C, nearly half of the students (40.5%) believed that the CET-4 was more important than the CET-6, supports this finding, because the CET-6 is a test required for graduate level studies. In the other two universities, a significant majority of students (86.7% at University A, and 77.6% at University B) chose the CET-6 over the CET-4 as the more important test.

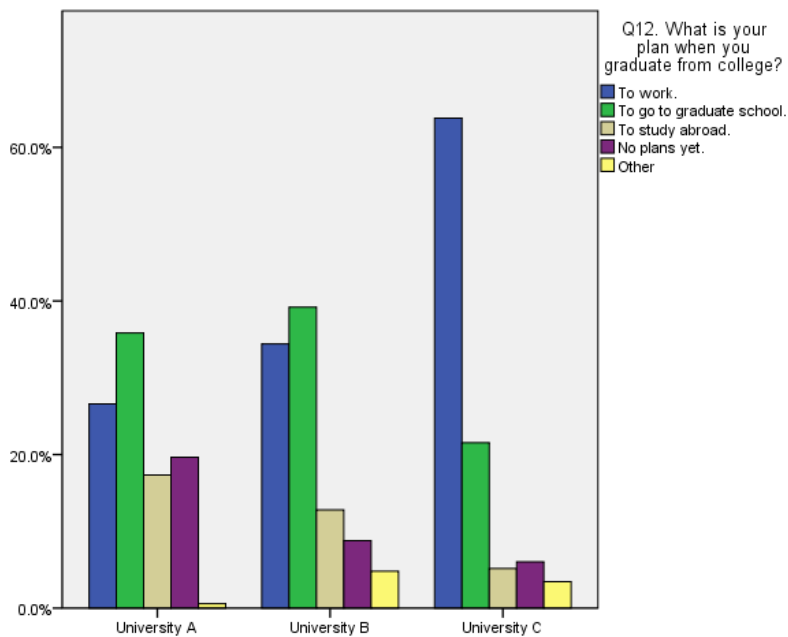


Figure 2. College graduation plans.

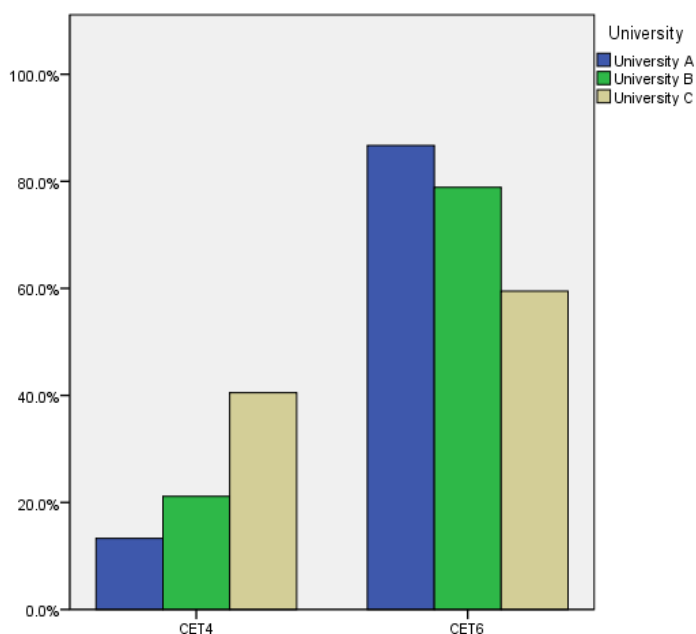


Figure 3. CET-4 and CET-6, which test is more important?

Research Question One

In the rest of this chapter, answers to the three major research questions will be discussed, based on the data collected. Starting with Research Question 1: What are college students' beliefs about English learning; what are their expectations and experiences of studying English at college, and what are their perceptions of the reformed CET-4? The discussion is divided into subtitled sections:

Beliefs About English Learning

According to the qualitative data, almost all the interviewees acknowledged the importance of English as an international language in terms of the status of English. Nevertheless, learning the language was not an interest for all students. Three students (A. Jin, A. B. Zhang, and A. Qin) pointed out that English was simply a tool for them, a means to an end, and the end was primarily utilitarian. A. B. Zhang even claimed that, he had always considered English simply a subject to be learned to pass a test. B. W. Wang

emphasized that English was a stepping stone for students who needed to get a job, particularly jobs in foreign-funded or joint-venture enterprises. A. Huang from University A claimed that science majors were not generally interested in studying English unless they planned to study abroad. Two students (B. Shi and B. W. Wang) from University B and one from University C (C. S. Chen) shared similar opinions. C. Liu from University C said that English was only important when one's future job was associated with English. Otherwise, it was not important. A. Xie, a Chinese Medicine major at University A, stated that among students in her field, English was not important, except for those who planned to practice abroad. Another student from University C (C. Feng) stated that many Chinese people would never make use of English in their lives. Hence, he personally believed that learning English was just a way for people to appear more sophisticated. While some declared that they had little interest in learning English in college, others felt confident that they could rely on the English fundamentals they had learned in senior high school, should they need English in the future.

Table 12

Perception of Closeness Between One's Major and English

University		Frequency	Percent
University A	not close at all	5	2.9
	little close	36	20.8
	middling	57	32.9
	quite close	62	35.8
	very close	13	7.5
	Total	173	100.0
University B	not close at all	10	8.0
	little close	29	23.2
	middling	47	37.6
	quite close	29	23.2
	very close	10	8.0
	Total	125	100.0
University C	not close at all	4	3.4
	little close	27	23.3
	middling	55	47.4
	quite close	27	23.3
	very close	3	2.6
	Total	116	100.0

Table 12 above suggests that the majority of participants thought there was only a middling connection between their majors and their knowledge of English.

Comparatively speaking, about a third of the students at University A (35.8%) considered the connection *quite close* when asked the relationship between their majors and their knowledge of English. The follow-up interviews further suggested that, of the three universities, University A had stronger teachers, better facilities, and more major courses taught in English, which could have prompted students to make the connection. A more detailed explanation of this finding will be provided in Chapter 6.

Expectations and Experiences About College English Courses

In regards to expectations about English learning at the college level, most students were like B. W. Wang, who took for granted the continuity of English course offerings from high school through college. They also expected the College English course to provide more practice and guidance for developing English communicative skills. However, many were disappointed in the College English course. B. W. Wang stated that before college, the primary focus of English classes was passing the test. He expected the college English courses to offer more simulated real-world practice so that students could improve their communication skills. In the past year, he stated, there was little opportunity to apply his English skills in his classes, including an oral English course he had taken as an elective at University B. Nor did the teachers review basic grammar taught at high school. Rather, the teachers mostly focused on vocabulary review and reading comprehension in their textbook. Moreover, he recalled that he and his classmates were very frustrated that they were not allowed to register for the CET-4 that would be administered in the first semester in December. They believed that, had they been allowed to take the test in December, they would have passed easily based on their high school knowledge of English, or at least they would have gotten good scores. He claimed that after one year at college, many students started worrying because they had forgotten so much English vocabulary and other knowledge of the language. He said that both he and his classmates had lost confidence in passing the CET-4 after taking College English. Similarly, in his self-recording shortly after taking the reformed CET-4 in December, A. Qin from University A claimed that he would get a higher score if he were allowed to take the test in June. Other students, like A. Zhu, tended to question the

legitimacy of offering English courses at the college level. According to A. Zhu,

Frankly speaking, I don't know what could be taught in college English courses.

Obviously, all the grammar points have been taught [in senior high]. What could possibly be developed at college is communicative skill. However, I don't think this is a skill that an ordinary English class can teach.

Similarly, A. Z. Li said,

Sometimes I don't think it's meaningful to have college English courses. Many of my classmates share similar opinions. We learned most of our English in high school. We didn't learn anything new now except expanding some vocabulary, doing some more role-plays or presentations at college.

In Phase IV (see p. 103), of the six participants who were interviewed for the second time in November (which will be referred to as “the November cohort”), when asked about their perception of the relationship between the reformed CET-4 and College English courses, B. M. Xu and C. Gao stated explicitly that College English courses mainly prepared students for the CET-4 by teaching vocabulary. B. Song claimed that there was no direct connection between his College English course and CET-4 preparation, except that the teacher supplied model tests to the students, and reminded them about the test now and then in class. A. Z. Li was the only participant who reported that her College English course was completely dedicated to the CET-4 training.

Time spent daily on learning English after class. Given the largely utilitarian attitude towards learning English at college, most participants revealed that they spent little time daily on studying the language after class. As shown in Figure 4 below, 65.5% – the majority of the students in the three universities spent only 0.5 hour after class each

day on English. A Kruskal Wallis test (Table 13) indicated that there was no statistically significant difference between the three universities in terms of time spent on learning English after class, with $\chi^2 (1, n = 414) = 5.692, p = 0.058$.

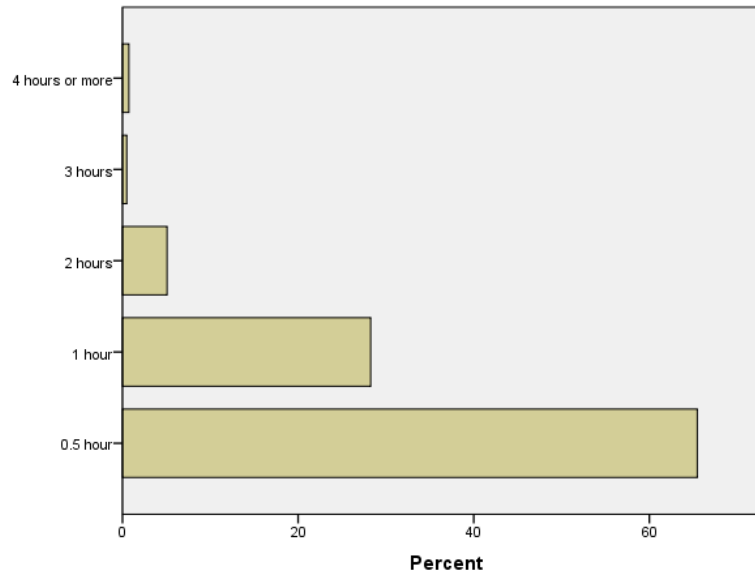


Figure 4. Hours spent daily on studying English after class at college.

Table 13

Test Statistics^{a,b} Regarding Time Spent Daily on Studying English After Class by Students at the Three Universities

	Hours
Chi-Square	5.695
df	2
Asymp. Sig.	.058

Note. a. Kruskal Wallis Test

b. Grouping Variable: University

Relationship between time spent daily on studying English and students' primary purpose for learning English. Even though the majority of students in the three universities spent only 0.5 hours every day on English, a one-way ANOVA test

indicated that there was a significant difference in the amount of time spent daily on English depending on the student's primary purpose for learning the language, $F(5, 408) = 3.03, p = .011$. Post-hoc comparisons showed that if the time spent on studying English was for the purpose of laying a good foundation in English to study abroad in the future, the mean score was significantly different from the mean score of those for whom time spent on English was related to other primary goals: i.e., those whose goal was to get high scores in the CET-4 and stand out in job hunting ($M = 1.3798, SD = .57546, p = .001$), and those whose goal was to pass required English courses and get high scores ($M = 1.3059, SD = .57784, p < .001$), and those whose goal was to improve comprehensive communicative ability in English ($M = 1.4462, SD = .69396, p = .008$). It would appear then, that among the students surveyed, those who planned to study abroad in the future were the students who spent the most time studying English. Also, the largest difference in time attribution existed between students whose goal was to study abroad in the future and those who wanted to pass required English courses and get high grades.

Relationship between time spent daily on English and graduation plans.

Furthermore, an ANOVA test indicated that depending on students' graduation plans, time spent on learning English varied significantly. Comparing the means scores, it became apparent that students who planned to study abroad were significantly more likely to spend time on studying English than those who had other graduation plans. The mean scores of those who planned to go to work directly out of college was $M = 1.3926, SD = .61298, p = .001$; of those who planned to attend graduate school in China, the score was $M = 1.4265, SD = .71629, p = .003$ and of those who did not have plans yet, the

score was $M = 1.2308$, $SD = .46927$, $p < .001$. This data supported the previous finding that students whose primary goal was to study abroad in the future devoted more of their time to English by studying the language every day. Moreover, the largest difference existed between students who planned to study abroad after graduation and those who did not yet have a graduation plan.

Perceptions of the Reformed CET-4

Overall perception of the reformed CET-4. Test takers who want to perform well on a test will usually make an effort to learn about the test, including the types of questions that may be asked. Thus participants were asked about their perceptions of the reformed CET-4. Their answers showed that even though most of the interviewees (23 out of 29 in Phase II) expected to take the CET-4 within a month, they knew little about it. Approximately, 15 of the students had heard of the reform of the test and why the reform had been instituted. Five students who were aware of it were unsure why the reform had been instituted, and five students were totally unaware of the reform. In the interview, altogether, only nine students knew what language skills and what types of questions would be included on the test. The majority of the interviewees, 16 students, had difficulty correctly naming all four sections in the reformed CET-4. Even in Phase IV (p. 103), among the November cohort, three of them (B. M. Xu, A. Qin, and B. Song) were not certain which test sections were included in the reformed CET-4. When the survey was administered for the second time in November in Phase IV, none of the five participants was able to correctly answer all the questions about newly added questions in the reformed CET-4. B. Song in particular claimed that he still knew nothing about the reformed test.

Perceptions about major language skills measured in the CET-4. Being ignorant of the four sections included in the reformed CET-4, the interviewees could not know which major language skills were being measured on the test. Three students thought that listening, reading and writing were the focal areas. Four students thought they would be tested in reading and writing only. Four others expected to be tested in listening and reading, while three thought they would only be tested on listening skills. C. R. Zhao guessed that they would be tested on listening, reading and translating and C. Ye and C. J. Xu thought it was listening, writing and translating. In Phase IV (see p.102), among the November cohort, test sections covering listening (B. M. Xu and C. Liu), reading (B. M. Xu and A. Qin) and writing (C. Liu, A. Z. Li, and B. Song) were the three sections that the participants were still worried about.

While the answers given by students varied, it is worth noting that the students repeatedly mentioned listening skills in their answers. This is an important point because student awareness of the importance of English listening skills is one of the intended positive washback effects targeted by the reformed CET-4.

Perceptions of the grading weight of each section and total score of the reformed CET-4. The grading weight of each of the test sections was adjusted in the reformed CET-4. However, 24 interviewees in Phase II (see p. 97) were unsure about how each section was weighted. Three students did not know the total score of the reformed CET-4 (710), nor the minimum CET-4 score required (usually 425) to qualify for taking the CET-6. Students like C. S. Chen thought that the total score of the CET-4 was still 100 and that a score of 70 was needed to qualify to take the CET-6. Even in Phase IV (see p.102), among the November cohort, there were still four students (B. M. Xu, A. Qin, B.

Song, and C. Gao) who were not aware of how each section was weighted. As a matter of fact, 13 out of the 414 students (3.1%) in the three universities surveyed estimated their eventual CET-4 score on the basis of a 100 point scale, rather than on a scale of 710, which was the updated measure. It is interesting to recall that the new score reporting system had been implemented nationwide in June 2005.

Perceptions of the certificate and the reformed CET-4. In spite of the fact that the CET-4 certificate was officially abolished with the implementation of the new scoring system, some students were not aware of this major change. In fact, 11 students believed that a CET-4 certificate would still be awarded to test takers with a passing score. Three students said that they had never seen a CET-4 certificate, but they firmly believed in its existence. Among them, a very assured A. Jin said, “There must be one. I think many college students are like me, whose goal is to get as many certificates as possible before graduation.” A. Meng replied rhetorically, “Why should I bother taking the CET-4, if there’s no certificate?” Three other students were unsure whether there was a certificate or not. Among all the 29 interviewees in Phase II (see p. 97), only A. Qin explicitly stated without any doubt that, while a CET-4 certificate had indeed been issued in the past, there was now only a score report. Even in Phase IV (see p. 102), among the November cohort, there were still three students (B. M. Xu, A. Z. Li and B. Song) who did not know about the new score reporting system.

Perceptions about requirements for getting a college diploma/degree and the reformed CET-4. In a national news conference at the beginning of the reform in February 2005, the Ministry of Education (MoE) made it clear that, as a matter of national policy, the CET-4 certificate and the college diploma were never to be tied

together. Soon after this clarification was made, many universities, including the universities in this study, publicly announced that they would no longer require their graduates to pass the CET-4 in order to qualify for their diplomas/bachelor's degrees. The English teachers who were among the first contacts for the current research, all assured me that a passing score on the CET-4 was no longer a prerequisite for receiving a diploma/degree from his/her university. Surprisingly, in the interviews, many students knew nothing about the policy.

Ten interviewees believed that there was a connection between a certain CET-4 score and the award of the college diploma/degree. They claimed to have heard this from seniors or friends. Officials did not clarify the rumor. Instead, students like B. J. Li from University B and C. Gao from University C had confirmation to the contrary from sources that they considered reliable: B. J. Li said that a former college English teacher had told the class that every college student had to pass the CET-4 to be granted a diploma. C. Gao recalled that on various occasions at University C, both the university president and the teachers had stressed that it would be impossible to graduate without passing the CET-4. Even in Phase IV (p.103), among the November cohort, B. M. Xu and A. Qin still claimed that students could not get their diploma without the CET-4, while B. Song and C. Gao said students could not get the college degree if their scores did not reach a required minimum.

What is more, additional information regarding graduation requirements was distorted. For instance, B. Song from University B maintained that when he was in senior high school, he had heard that passing the CET-4 was a requirement for getting a diploma. In fact, his college friends told him that at University B, passing both Higher

Mathematics and the CET-4 were required to get the diploma. Likewise, C. Feng said that university administrators at University C had told students that graduation without passing the CET-4 was not possible. His English teacher had said that they would be safe once they passed the CET-4. He claimed that the CET-4 was still a common prerequisite to getting a diploma at college. He even argued that University C had lower requirements, because in other universities, according to what he had heard, the upper-level CET-6 was the required test to take for graduation. It is not clear exactly why some administrators or teachers spread outdated and incorrect information on the subject of CET-4/graduation. As might be expected, the students tended to believe these authorities when faced with conflicting statements.

Perceptions of the CET-Spoken English Test (SET). As mentioned above, the MoE expected to improve student communicative skills through the reform of the College English course. Thus, in addition to listening skills, speaking was also prioritized by the MoE for teaching and learning English at the college level. However, for various reasons, only a selective set of students, i.e., those who get a minimum score of 550 in the reformed CET-4 or a minimum of 520 on the CET-6, are eligible to advance to the CET-SET test which measures oral English proficiency.

About 12 interviewees knew about the CET-SET. They knew that a certain score on the CET-4 was required in order to register, yet none could give the exact required score. When asked whether they would take the test if they qualified, A. Y. Zhao and B. Ding nodded, but they stated that their oral English was too poor to pass. C. Gao was unsure whether she wanted to take the test for the same reason. C. Liu then said he did not want to “embarrass himself by going through it” since he had little practice in oral

English. Among the November cohort, C. Gao and B. Song were the two participants who changed their minds, and said they would take the CET-SET if they qualified for it. In contrast, although they had heard about the CET-SET, A. Z. Li and A. Zhu did not know whether this was a formal test or how valuable the certificate would be. A. Z. Li was certain only that the Intermediate English Interpreter Certification Exam was a formal and nationally-recognized oral English test for everyone in China, and it was this test that she wanted to take. In Phase IV (see p. 102), A. Z. Li still claimed that she did not know how the CET-SET was administered. A. Luo said firmly that he had no plans to take it, because he did not believe that he would truly improve his oral English ability by taking the test. A. Tang and A. Liang who planned to pursue their graduate study abroad said that it was worthless for them to take the CET-SET. Instead, they would take the Test of English as a Foreign Language – Spoken English Test (TOEFL-SET). Two other students, A. B. Zhang from University A and C. R. Zhao from University C, however, confessed that they had never heard of the CET-SET.

Perceived goal of the CET-4 in students' eyes. The goal of both the old and reformed CET-4 is to measure the extent to which college students have met the standards set by the Ministry of Education for the College English curriculum. In the interview, three students said that the goal of the CET-4 was to measure college students' overall English level. B. Ding stated that it was to check a student's English fundamentals. A. Qin said that in addition to this, the score could be used as a reference for future employers. B. M. Xu also stated that it was a way to select more competent employees in the workforce since many employers use the CET-4 certificate to screen job applicants. Three other students held that it was a means by the MoE of pushing college students to

continue studying and improving their English in college. Some students at University B recalled that in a presentation by a guest teacher from the New Oriental Language Training School a month earlier, the teacher told them that the goal of the CET-4 was to eliminate those with the lowest proficiency, but not necessarily to select the best English learners. Although this was not an official interpretation, these students seemed to agree with the speaker. A. B. Zhang from University A, however, honestly admitted that he had no clue what the goal of the CET-4 was.

Perceived skills needed to pass the CET-4 and students' aspirations relating to the test. Without a shared clear idea about the goal of the CET-4, students understandably gave a wide array of responses as to how to pass the test. Five students insisted that extensive practice of the model test questions was the key to passing. Eight interviewees said that the CET-4 tested students' overall English ability. A. Pan stated that he would depend on the accumulation of English knowledge learned in senior high school, because he felt he had not learned much in college. Four students emphasized the importance of memorizing the CET-4 vocabulary, especially the words that appeared with higher frequency. Two students stressed that listening ability was essential, whereas C. Liu claimed that “if you are good at writing, then you will have no problem passing the CET-4.”

Regarding whether students cared about passing the CET-4, eight students mentioned that their immediate family members were concerned about their performance on the test. C. R. Zhao pointed out that her classmates were concerned about her score as well, since her classmates were not only friends, but also competitors. C. Feng and C. J. Xu held that their English teachers shared their concerns about the test. Ten interviewees

insisted that it was the teacher's obligation to help students prepare for the CET-4. Nevertheless, A. Y. Zhao claimed that teachers didn't care whether students passed the test or not. Three other students said that only they cared about the test and no one else did. A. Z. Zhang was the only student among the participants who did not care about the test.

A Short Summary of the Responses to Research Question One

In short, English is largely considered useful as a means to an end, such as getting good grades and finding a satisfactory job. The majority of the college students do not acknowledge a close connection between college English learning and their major. Many students were not impressed with the College English course, nor did they think they had learned much from it. Most students in the three universities spent only 0.5 hour every day studying English after class. However, students whose graduation plans included laying a good foundation in English for future study abroad, spent significantly more time on studying English than students who had other goals. What is more, students whose primary goal was to study abroad devoted more of their time to learning English. The largest difference existed between students who planned to study abroad after graduation and those who did not yet have graduation plans.

As to the reformed CET-4, in general, students did not know much about the format, content or score reporting, etc. despite the fact that the test date was quickly approaching. They also held many misperceptions about related policies.

Research Question Two

Next, the author will move on to Research Question 2: What plans did the students make to prepare for taking the CET-4, and how did they implement their plans? Again, subtitles will be used to discuss the different dimensions of the question.

Books and Materials Needed for CET-4 Preparation

With the approach of the test date, about 15 interviewees bought a CET-4 vocabulary book of some kind. Of these, 13 students had either a copy of a past CET-4 test or simulated tests from past sessions. Six students had both a CET-4 vocabulary book and copies of some model tests, and four students from University A had also purchased a CET-6 vocabulary book. To prepare for the CET-4, A. Z. Li had bought only a CET-6 vocabulary book. Three interviewees from University A had Test of English as a Foreign Language (TOEFL) and Graduate Records Exam (GRE) vocabulary books as well. Two students did not buy any test-related preparation materials. One was A. Zhu, who said that he did not plan to spend extra money on the CET-4 because it looked like an easy test to him. He claimed that if he wanted to practice, he could easily find materials online. Later his CET-4 score was reportedly 623. The other was B. W. Wang, who said his ultimate goal was not to study for the test, but to master the language, so this was just another test for him. Nevertheless, he made some preparations using library resources and CET-4 handouts distributed by his English teacher in class. He obtained a score of 566.

Other than CET-4-related materials, not many students had English materials of other kinds for extensive reading or for entertainment. Six students had English magazines such as Studio Classroom or Crazy English. Five students bought English newspapers like 21st Century or China Daily sometimes. Five students had bilingual books like Classic Essay Appreciation, Greek Mythology or English Classic Stories. So, there seemed to be few English materials other than textbooks available in China for college students to do extensive reading in English in their leisure time.

CET-4 Preparation Plans

Among the 414 students who filled out the survey, 68 (16.42%) claimed that they planned to take the CET-4 without preparation. Out of this number, 19 students (9 at University A, 5 at University B, and 5 at University C) argued that the CET-4 was easy. 29 students (11 at University A, 7 at University, and 11 at University C) replied that they were too busy to make any preparations.

Of the 346 students who claimed that they would prepare for the CET-4, approximately half (171 students, or 49.42%) planned to memorize the CET-4 vocabulary. Approximately a third (110 students, or 31.79%) planned to use simulated CET-4 tests or tests from previous years. Approximately a quarter (86 students, or 24.86%) planned to either practice CET-4 listening, writing or reading (45 students, or 13%; 50 students, or 14.45%; respectively). Less than 5% of the students planned to review grammar, practice translation, or watch English movies or TV programs. Of course, most of the students included several of these practice methods in their preparation plan. One particularly ambitious student from University A wrote that he/she would memorize 100 CET-4 words every day, whereas most students planned to memorize no more than 10 words every day. The majority planned to practice doing one or two sets of simulated tests every week. One student from University B and one from University C wanted to finish one set of simulated tests every day. For listening practice, most of the students planned to use either CET-4-related listening materials or listen to English radios/tapes/CDs every day.

Among the interviewees, 19 had made some preparation for the CET-4 and eight had not, by the time of the interview in May, which was one month before the test. B. M. Xu, a student from Northwestern China at University B, was one of those who fully committed herself to preparation, even though she would not take the test until December.

She said, “As soon as I got to college, I started doing extra listening. I want to improve my English ability, but I practice to pass the CET-4, too.” She did not have any specific plans. However, when she was free in the evenings, she practiced listening or reading some short articles in English. Sometimes she memorized model essays in the mornings. She said the listening section of the test worried her most. Even though she had done a lot of listening as practice for the test, she claimed that it gave her a headache. C. R. Zhao practiced both reading and listening but she too, was nervous about the listening section. A. Jing practiced by using five or six past tests and she memorized some CET-4 words. She said that her listening ability was particularly weak, so she spent time listening to the Voice of America (VOA) Special English listening programs online. C. J. Li practiced listening before going to bed, “Every day when I lay in bed, I listened to English and fell asleep with it.”

Most of the students started their CET-4 preparations at the beginning of the fall semester in February. Rote memorization of the CET-4 vocabulary and practice using simulated or past CET-4 tests played an important role in student preparation. For instance, B. J. Li reported that he had several sets of past tests to use for practice. He also bought a CET-4 vocabulary book after his roommate bought one. His preparation focused primarily on vocabulary. He believed that without understanding new words, he would not be able to comprehend the reading selections. He also mentioned that he needed to read more English articles and memorize well-constructed sentences to prepare for the writing sample.

B. X. Chen, a student from University B, was told by her College English teacher that she should memorize frequently used words for the CET-4, so she followed that

advice. She also used a set of past tests for practice. A. Xie stressed that she had to learn English because of the test. She did some listening, writing, and reading practice using past CET-4 tests, but mainly she memorized new words. (This is why every morning, at colleges or universities throughout China, it is common to see students reading aloud, on campus, in or out of classroom. Many of them read English textbooks or memorize English vocabulary.) B. W. Wang also started memorizing the CET-4 vocabulary. Besides, he had practiced taking approximately 10 past tests using a CET-4 preparation book that he found in the library. Because he did not have the CD for the listening section and had no one to grade his writing, he did not practice the listening and writing sections of those simulated tests. As a result, he was most unsure about those sections of the upcoming CET-4.

A. Jin and A. Zhu did not start the preparation until the time of the interview. They both admitted that they had never taken one whole set of CET-4 simulated tests. A. Zhu was confident that he would pass the test, however, estimating that his score would be somewhere between 620-630. (In fact, he scored 623.) Some students were like B. Shi who organized his test preparation skill by skill: reading, translating, writing and listening. C. J. Li, for instance, started in April with copies of some past tests distributed by her English teacher. She also bought more simulated tests to use for practice. B. Ding said that she mostly memorized CET-4 vocabulary and practiced using past tests.

In addition to listening, writing was a skill that was frequently mentioned by the interviewees. According to the interviewees, the most effective way of preparing for the CET-4 writing section seemed to be by memorizing model sentences, essays or formulaic writing templates. B. Shi said the students in his English class were asked to write timed

essays in class for half an hour, as that was the time allotted during the actual test. Then the students turned in the papers to be rated. C. R. Zhao started preparing in April, memorizing model essays and writing templates. She wanted to collect some good English sentences and become more familiar with the CET-4 writing format. B. Ding planned to memorize good sentences two weeks before the test. To practice writing, C. J. Xu and C. J. Li kept weekly journals as required by their English teacher.

Even though four students (A. Meng, C. Feng, C. J. Xu, and C. Ye) said they made some preparations for the CET-4, they all admitted that they did not strictly follow the plan. One reason for this was that finals were being given shortly before the CET-4 test date (June 19), which created a conflict between preparing for the CET-4 and major-related final tests. A. Meng from University A explained that,

At the beginning of the semester I was very enthusiastic. I did a few pages of exercises on my CET-4 test papers, but I was not persistent. Now I have to review for finals. After finals, I probably will spend two or three days doing some more listening and writing exercises before the test date. You know, if I fail my CET-4 this time, I could retake it with just a small registration fee. However, if I fail the final of a subject in my major, I could end up with a bad GPA and a much more expensive fee for either retaking just the final test or retaking the entire course! So I need to prioritize passing my finals at this moment.

C. Feng and C. J. Xu also memorized some vocabulary and used past CET-4 tests distributed by their teachers during the semester. After finals, they planned to continue expanding their vocabulary and practicing listening.

There were also students who had not started preparing for their CET-4 nor did they have any plans to do so at the time of the interview. C. Y. Zhang, for instance, stated that she had taken three or four difficult courses that semester, and the courses required by her major were hard for her. Earlier, she had been preoccupied with a test for transfer students, which she had failed. She claimed that when finals were over the following week, she would then start preparing for the CET-4. She had purchased some copies of past tests for the purpose, and she planned to finish those and review the errors, but at the time of the interview, she had never finished a complete set of past tests.

A. Luo and B. Dong likewise made no preparations. They both planned to practice using some simulated tests the day before the scheduled test. A. Luo said that, based on his experience, “Cramming is quite important to pass an English test.” Later A. Luo’s CET-4 score was 521. (He did not tell the researcher the scores for each section because he could not find his registration ID card.) B. Dong’s scores by section were Listening 200, Reading 199, Integrated 61 and Writing 84, for a total of 544. C. Gao simply decided to take the test in December, even though she was allowed to register for the test in June.

C. Liu, B. Song and A. Z. Li were not allowed to take the CET-4 until December, yet C. Liu had started expanding his vocabulary and doing model tests. B. Song said he would start memorizing new words and doing listening practice after finals or during summer vacation. A. Z. Li from University A did not think it would be challenging for her to pass the CET-4 but she was concerned about the CET-6. Still, she planned to start memorizing CET-4 vocabulary, and listening to the BBC or practicing some simulated CET-4 tests during the coming summer vacation. In the end, her CET-4 score was 493,

with Listening 187, Reading 166, Integrated 51, and Writing 89, which was much lower than her expected score of 520. However, she reported that she achieved a score of 546 in the internet-based trial test with Listening 192, Speaking 71, Reading 210 and Writing 73.

Other students studied for the CET-4 by preparing either for the CET-6, the Test of English as a Foreign Language (TOEFL) or the Graduate Record Exam (GRE). A. Yan, for instance, wrote in his journal that he prepared for the CET-4 by practicing listening, cloze and fast reading questions from the CET-6. A. B. Zhang prepared the CET-4 by memorizing TOEFL vocabulary, doing TOEFL reading, and practicing vocabulary. He believed that he could prepare for the CET-4 while preparing for his TOEFL, since there were many similarities between them. His goal was to memorize 70% of the TOEFL vocabulary before taking the CET-4 test.

On the other hand, among the interviewees, A. Z. Zhang decided not to prepare for the CET-4. He claimed that he was too busy with finals, and he was competing in a national contest that was related to his major. (His CET-4 score was 529 with Listening: 184; Reading: 191; Integrated: 54; Writing: 100.) A. Liang also said decisively that he had no plans to prepare for the CET-4. Having registered for the GRE, he had devoted all of his spare time to GRE preparation. He was not nervous taking the CET-4, arguing that he could easily retake it if he was not satisfied with his score.

Another student, A. Huang's case was common among students who thought they had built a good foundation in English. He said that from the bottom of his heart, he wanted to spend some time preparing for the CET-4. However, he felt ashamed about studying for it because the upper classmen kept telling him that the test was easy, and that there was no need to prepare for it. A. Q. Wang, another student who had registered for

the GRE, also repeatedly heard that the CET-4 was easy, so she decided not to spend time on it. However, at the same time, she expressed her uneasiness in her journal when she saw students preparing for the CET-4 in the library.

Participants' self-recordings and journals regarding CET-4 preparation. In Phase IV, students were asked to record their preparations for the CET-4 and/or their experience of studying English over a period of three weeks, and then before taking the test, they were to record at least twice a week. They were also asked to record their activities before and after taking the test. There was no limit on the length of the recordings and they were welcome to record more than the required six times.

Eight students submitted self-recordings or journals. Five of them prepared specifically for the CET-4, concentrating primarily on CET-4 vocabulary, listening, and banked cloze. Also mentioned in their recordings/journals, was practice in in-depth reading, cloze, fast reading, model test practice, and model essay memorization.

Anxiety mounts with the approach of the test day. A. Qin said that he could feel the tension among his classmates. He understood that everyone was trying to get a high score, but at the same time he did not want to put too much pressure on himself. B. M. Xu stated that she had to force herself to spend time on practice even though she was tired of it. A. Huang said that he did not feel quite comfortable taking the reformed CET-4 without any preparation, even though many of his senior friends told him not to waste his time on it. He later quit one of his science contests at the last minute in order to do some listening and model test practice. Interestingly, his journal entries were all made after 11:30 p.m. in classrooms that were open to students 24 hours a day.

In the recordings, two students said they did not prepare for the reformed CET-4. A. Tang wrote that she was busy preparing for the Graduate Record Exam (GRE), doing assignments in her major, in addition to working as a Shanghai World Expo volunteer, and she did not have time for anything else. Despite this, she said that she could not help feeling uneasy when she saw students constantly memorizing the CET-4 vocabulary and doing CET-4 model tests in class. B. Song, the other student, chose to cope with the pressure by watching Hollywood movies instead of doing CET-4-related practice. He said this was his way of practicing English listening. He wanted to spend time preparing for the CET-6 in the future.

Almost all the participants mentioned in their recordings/journals that, during that time period, their English teachers seemed to have shifted their attention temporarily to the CET-4 in class. A. Z. Li's teacher played listening materials adapted from CNN News. B. Song's teacher talked about CET-4 writing techniques. B. M. Xu, A. Tang, and A. Z. Li's teachers all had students practice the CET-4 model test in class and discussed the questions and test-taking strategies afterwards.

One day before the test date, according to the recordings/journals, listening was the most commonly practiced skill. Two students continued to review some of the reformed CET-4 vocabulary. B. M. Xu said she would have liked to complete another set of the CET-4 model tests the night before the test, but she was afraid that if she scored poorly on the practice test, it would have a negative influence when she took the test the next day. Instead of studying, she watched a Hollywood movie with her roommates in the dorm. She claimed that even though everyone looked relaxed and seemed to enjoy the movie, she could tell they were nervous inside. When they took the test on December 18,

A. Yan and A. Tang checked the answer keys online as soon as the test ended. The former was confident that he would achieve a score of 600 or more, and the latter did not find that he had made many mistakes.

All the other students reflected on how short the time was: B. Song did not finish writing. A. Z. Li, B. M. Xu and A. Huang said that they had spent too much time on banked cloze, which left them very little time to do cloze, and they simply filled in the bubbles based on guesses. Among them, except for C. Gao, all of them estimated that they would get at least the minimum score of 425. Their estimates all proved to be correct. C. Gao only got 412 with listening: 139; reading: 161; integrated: 39; and writing: 75. Her scores for each section were the lowest in each category among the seven students in Phase IV who gave their final scores to the researcher. The range of scores for each section among the seven participants was as follows: listening 139 ~ 227, reading 161 ~ 249, integrated 39 ~ 68, and writing 75 ~ 126. From among the scores collected, writing was the weakest skill overall. (In the reformed CET-4, the total score for Listening is 249 - 35%, Reading 249 -35%, Integrated 71 - 10%, and Writing and Translating 142 - 20%). Three students from University A achieved the highest possible scores within each section. The highest total score was 646 by A. Huang, who also had the top achievable score in the reading section (249).

Taking CET-4 intensive training courses. Among the 414 surveyed participants, 16.9% of them claimed that they were planning to take a CET-4 intensive training class offered by his/her university or by other agencies, in order to get a satisfactory score on the reformed CET-4. 1.9% said they had taken the course. Of course, a number of them were taking it at the time of the survey. As pointed out above, an elective CET-4 Ability

Training Course open to all students was offered at University B. According to B. J. Li, due to the high demand for this course, two sections were offered, given by two teachers. Nevertheless, 70 – 80 students were registered in each class. This course, as suggested by course title, is dedicated exclusively to the development of skills and to practice various types of questions that had appeared previously on the test. At University C, a separate CET-4 training class was offered on campus by the English Department. It was an evening class that was held twice a week for a fee of 240 RMB (approximately \$35) per person. C. J. Xu signed up for it because she expected that the teacher would concentrate completely on preparation for the CET-4 in this class, without worrying about covering units from the textbooks. There were approximately 120 students in this class. The teacher distributed the CET-4 model tests and tests of previous years and asked them to complete it after class. In class, the teacher taught CET-4 vocabulary, listening, and fast reading, and explained the test questions distributed earlier in a classroom equipped with multi-media. C. J. Xu, however, pointed out that even though the fee was non-refundable; the number of students who attended the class dwindled each time the class met, with only ten students attending the last class.

A Short Summary of the Responses to Research Question Two

In short, almost half of the interviewees bought a CET-4 vocabulary book of some kind and/or copies of the past CET-4 test or simulated tests to prepare for the exam. Approximately 84% of the 414 participants claimed that they planned to prepare for the CET-4. The majority of the students planned to memorize the CET-4 vocabulary and take the simulated tests or tests from previous years. As to specific English skills, listening practice was mentioned most frequently, followed by reading and writing practice. However, not all students followed their plans strictly. A small number of interviewees

still had not started preparing for the CET-4 at the time of the interview, either because they thought the CET-4 was easy, or because they were too busy preparing for their finals or for the TOEFL/GRE, or they were studying for classes in their majors. Participants' self-recordings and journals further showed that the most of the students were nervous before taking the test. If they prepared for it, their focus was mainly on CET-4 vocabulary, listening and banked cloze. At that point the English teachers, too, started talking about the test questions more frequently, according to the self-recording and journal participants.

Research Question Three

In the rest of this chapter, the author will discuss Research Question 3: To what extent has the reformed CET-4 influenced students' perceptions of the practices of English teaching and learning at the college level? Again, subtitles will be used to separate the different aspects that were relevant to this question.

Overall Impact of the Reformed CET-4

According to the survey, as indicated by Table 14, 48 participants (11.6%) thought that the CET-4 had no influence on their English learning. Specifically, as can be seen in Table 15, the total can be broken down by university as follows: 16.8% from University A, 10.4% from University B, and 5.2% from University C. What is more, as shown in Table 16, 51 participants (12.3%) did not think the CET-4 had any influence on how English was taught in their classrooms. Table 17 shows that, 17.3%, 12.8% and 4.3% from University A, B, and C respectively reported that English instruction in their college classrooms had not been affected by the CET-4. In other words, the perceived impact of the reformed CET-4 on English teaching and learning appeared strongest at University C and weakest at University A.

Despite these numbers, the majority of students in the three universities still felt that the CET-4 had made an impact on how they learned English and on how their instructors taught. Out of 34 interviewees, in Phase II and IV (three from University A, five from University B, and five from University C in Phase II, and four from University A in Phase IV) 18 reported that their English teachers helped them prepare for the CET-4 by teaching listening, reading, writing skills, and/or distributing and/or explaining CET-4 model tests, in addition to presenting units from the textbook in class. Other than students who took the CET-4 training course, C. R. Zhao and C. Y. Zhang from University C in particular mentioned that their College English course focused primarily on preparation for the CET-4 during that semester.

Table 14

Collective Student Opinion of the Influence and Relevance of the Reformed CET-4 on Students' English Learning

	Frequency	Percent
Yes	366	88.4
No	48	11.6
Total	414	100.0

Table 15

What is the Respective Student Opinion in Regards to the Relevance of the Reformed CET-4 Influencing Students' English Learning?

University		Frequency	Percent
University A	Yes	144	83.2
	No	29	16.8
	Total	173	100.0
University B	Yes	112	89.6
	No	13	10.4
	Total	125	100.0
University C	Yes	110	94.8
	No	6	5.2
	Total	116	100.0

Table 16

What is the Collective Student Opinion in Regards to the Relevance of the Reformed CET-4 Influencing Teachers' English Teaching in the Three Universities?

	Frequency	Percent
Yes	363	87.7
No	51	12.3
Total	414	100.0

Table 17

What is the Respective Student Opinion in Regards to the Relevance of the Reformed CET-4 Influencing Teachers' English Teaching?

University		Frequency	Percent
University A	Yes	143	82.7
	No	30	17.3
	Total	173	100.0
University B	Yes	109	87.2
	No	16	12.8
	Total	125	100.0
University C	Yes	111	95.7
	No	5	4.3
	Total	116	100.0

Table 18

What is the Collective Student Opinion in Regards to the Influence of the Reformed CET-4 on Students' Learning Content?

	Frequency	Percent
No	205	49.5
Yes	209	50.5
Total	414	100.0

Table 19

What is the Collective Student Opinion in Regards to the Influence of the Reformed CET-4 on Teacher's Teaching Content?

	Frequency	Percent
No	163	39.4
Yes	251	60.6
Total	414	100.0

Table 20

What is the Collective Student Opinion in Regards to the Influencing of the Reformed CET-4 on Teachers' Teaching Methods?

	Frequency	Percent
No	237	57.2
Yes	177	42.8
Total	414	100.0

As indicated in Table 18 above, the aspect of English learning perceived to be most influenced by the CET-4 was content (50.5%). Also, from the student perspective, as shown in Table 19 and Table 20 above, the two aspects of the test that had the greatest influence on teaching, were content (60.6%) and methods (42.8%). Furthermore, in all three universities, the reformed CET-4 appeared to have a positive effect on the language used for instruction and on the adoption of multi-media for teaching English.

Language of Instruction

Table 21

Frequency of Communicative Activities that were Conducted in English During Students' English Classes at College

University		Group discussion/ pair work	Teacher's lecture	Teacher asks questions/ Students answer questions	Students volunteer to express one's ideas
University A	N	173	173	173	173
	Mean	2.91	4.14	3.60	2.82
University B	N	125	125	125	125
	Mean	2.88	4.11	3.31	2.75
University C	N	116	116	116	116
	Mean	2.52	4.28	3.69	2.61

As indicated in Table 21 above, when students were asked to rate communicative activities in their classes, the use of English during lectures was the most frequently

indicated activity, more than the use of group discussion/pair work, questions/answers in class, and volunteering to express ideas in English in class. Using a Likert scale in which 1 = never and 5 = always, the mean scores for frequency of teacher lectures being taught in English were 4.14 for University A, 4.11 for University B, and 4.28 for University C. Compared to the interview data in which many students reported that teachers in their senior and/or junior high school English classes used Chinese as the language of instruction, college teachers seemed to have created a better English learning environment by using English as the language of instruction, thus helping students to experience the language and improve their communication skills.

Adoption of Multi-Media in English Teaching

From both the survey and interview data, multi-media-assisted teaching has reportedly been widely adopted in English language classes. As shown in Table 22 below, with a mean score 4.41, 4.66 and 4.46, teachers at all three universities reportedly used multi-media often in their language classes. The percentages are shown in Table 23 below, with 61.8%, 76.8%, and 62.9% students from University A, B and C respectively, indicating that their English teachers *often* used multi-media. Given that the CET-4 committee encouraged multi-media use, this could be considered an instance of a positive washback effect generated by the reformed CET-4.

Table 22

Frequency of Using Multi-Media Assisted Teaching in English Classes

University		N	Mean	SD
University A	Teach with multi-media	173	4.41	.915
	Valid N	173		
University B	Teach with multi-media	125	4.66	.708
	Valid N	125		
University C	Teach with multi-media	116	4.46	.848
	Valid N	116		

Table 23

Frequency of Using Multi-Media in English Classes

University		Frequency	Percent
University A	Never	4	2.3
	Seldom	3	1.7
	sometimes	18	10.4
	Often	41	23.7
	Always	107	61.8
	Total	173	100.0
University B	never	0	0
	seldom	3	2.4
	sometimes	8	6.4
	Often	18	14.4
	Always	96	76.8
	Total	125	100.0
University C	Never	1	.9
	Seldom	4	3.4
	sometimes	9	7.8
	Often	29	25.0
	Always	73	62.9
	Total	116	100.0

Corresponding to the survey data, student interviews also indicated that their English teachers either used PowerPoint presentations or incorporated English audio/video materials into their English courses. University A, for instance, used the New Horizon textbook, one of the texts that had been designed for the college English reform. The book has a complementary website with listening, speaking, reading, and

writing exercises for each unit. Teachers usually assigned homework from the website, announcing a theme for students to discuss on the online forum. Every student had an account, whose log-in and log-out times and active online time were recorded automatically and checked by the teacher. The software is also designed to grade multiple-choice questions. It largely functions as the WebCT or Moodle in the US colleges or universities.

B. X. Chen said that no textbooks were used in her University B CET-4 Ability Training class. The teacher simply taught through PowerPoint presentations. In College English courses, B. Dong, B. X. Chen, and B. J. Li reported that the teacher often used multi-media facilities, and sometimes showed pre-downloaded movie clips or English songs during the break. However, unlike classes at University A, internet access at University B was not available in classrooms, but only in computer/language labs. To compensate for the situation, teachers at University B often downloaded needed materials ahead of time.

C. R. Zhao at University C stated that in her College English III course, the teacher always used PowerPoint. C. J. Li's teacher always reviewed important words and phrases using PowerPoint. For translation exercises in his College English III course, C. Feng's teacher used a complementary CD. Like University B, University C did not have internet access in class. However, even though this was the case, according to the students, information presented through multi-media assisted college English classes was much more profound and vivid in width and depth than traditional classes taught with a box of chalk and a blackboard.

Students' Learning Methods were Slightly Influenced by the Reformed CET-4

Table 24

What is the Collective Student Opinion in Regards to the Influence of the Reformed CET-4 on Students' Learning Methods/Strategies?

	Frequency	Percent
No	293	70.8
Yes	121	29.2
Total	414	100.0

Table 25

What is the Perspective Student Opinion in Regards to the Influence of the Reformed CET-4 on Students' Learning Methods/Strategies?

University		Frequency	Percent
University A	No	119	68.8
	Yes	54	31.2
	Total	173	100.0
University B	No	96	76.8
	Yes	29	23.2
	Total	125	100.0
University C	No	78	67.2
	Yes	38	32.8
	Total	116	100.0

As indicated in Table 24, and Table 25 above, it seemed that in this study, the reformed CET -4 did not have a strong impact on students' learning methods/strategies, as reported by 70.8% of the 414 students. Broken down by university, 68.8% from University A, 76.8% from University B, and 67.2% from University C claimed that the reformed CET-4 had not had an effect on their English learning methods/strategies.

According to C. Ye, his English learning method remained almost the same as in senior high school, because the focus was still on learning new vocabulary, reading,

listening, and practicing with model tests. B. X. Chen said her English learning method was an extension of the one she adopted while preparing for the entrance exam to college: she memorized new words of high frequency; practiced using model tests, and sometimes read some interesting, simplified English articles if she had time. She said she wanted to improve her ability to apply English at college, but she did not know how to do it. A. Luo from University A stated that he still largely followed the teacher's instructions in class, as he did in high school, and did not spend much extra time on English after class. So he did not think his English learning method had changed either.

Three Perceived Most Influenced Aspects by the Reformed CET-4

As indicated in Table 26, Table 27, and Table 28 below, the three aspects of learning English that had been most influenced by the CET-4 were listening (66.4%), vocabulary (49.8%), and writing (36.2%). Judging from the CET-4 preparation plans reported by the students earlier, students had concentrated on listening practice, vocabulary expansion and memorization of model essays. Additionally, in student eyes, the three aspects that had been influenced most by the CET-4 in their teachers' instructional methods were listening (61.8%), writing (35.7%), and reading (32.6%), as indicated in Table 29, Table 30, and Table 31.

Table 26

Percentage of Listening that was Influenced in Students' Learning

	Frequency	Percent
No	139	33.6
Yes	275	66.4
Total	414	100.0

Table 27

Percentage of Vocabulary that was Influenced in Students' Learning

	Frequency	Percent
No	208	50.2
Yes	206	49.8
Total	414	100.0

Table 28

Percentage of Writing that was Influenced in Students' Learning

	Frequency	Percent
No	264	63.8
Yes	150	36.2
Total	414	100.0

Table 29

Percentage of Listening that was Influenced in Teachers' Teaching

	Frequency	Percent
No	158	38.2
Yes	256	61.8
Total	414	100.0

Table 30

Percentage of Writing that was Influenced in Teachers' Teaching

	Frequency	Percent
No	266	64.3
Yes	148	35.7
Total	414	100.0

Table 31

Percentage of Reading that was Influenced in Teachers' Teaching

	Frequency	Percent
No	279	67.4
Yes	135	32.6
Total	414	100.0

More Emphasis on English Listening at College. In accordance with the survey data, during the interviews students in general claimed that before college, they managed well without reading and writing English. However, they said this was not the case in college where they needed to be proficient in listening. Almost all students in the three universities felt that they had much more listening practice in class in college. At University A, A. Jing and A. Qin said their English teachers always played listening exercises online during their Comprehensive English classes. Their English classroom was always equipped with an overhead projector and a desktop computer with internet access. Besides, all students had bi-weekly listening classes held in a language lab.

At University B, B. Dong reported that the teacher in his College English course always played the CET-4 listening exercises, and explained the answers in class. His CET-4 Ability Training class was planned and delivered in accordance with the four language skills tested in the reformed CET-4. Much time was spent on improving listening skills. The teacher showed English video clips and famous speeches. In a different section of the CET-4 Ability Training course, B. J. Li and B. X. Chen said their teacher not only had CET-4 listening practice in class, but included simple and effective test-taking strategies for tackling the CET-4 listening comprehension questions.

In B. X. Chen's College English course, the teacher always played listening exercises on a tape recorder at the beginning of the class and/or at the end of each unit. In B. J. Li's College English, the teacher showed funny video clips or played popular English songs during breaks. In B. W. Wang's College English III, listening practice was done at the beginning of every other class. Even though there was no separate English listening class at University B, one of B. Shi's classes was dedicated to improving listening skills. The teacher gave students extra listening materials, and then explained each question briefly. In their CET-4 Ability Training class, the teacher had the students make 10 one-minute presentations in every class. After each presentation, the audience could ask questions. English video clips were often shown in class for students to comment on as well. B. Song, the student who would not take the CET-4 until the following June, also mentioned that his teacher held many listening practices in his College English class.

At University C, there was a separate internet-based English class held every other week at the language lab as a component of the College English course. The teacher was simply a facilitator in this class, while students worked independently on listening exercises from the New College English textbook's complementary software. After submitting answers, both teacher and student could see his/her grade. In addition to this listening practice session, four students (C. Ye, C. J. Li, C. J. Xu, C. R. Zhao) reported that their teachers regularly gave them listening practice at the beginning of the Comprehensive English courses.

Listening exercises were more frequently practiced by students out of class.

In line with the interview results, regardless of the university, out-of class listening activities such as watching English movies/TV series and listening to English

speeches/presentations, were more frequent activities than reading, writing, speaking, and doing whole sets of CET-4 mock tests or past CET-4 tests, etc. This information was gleaned from a Likert scale in which 1 = never and 5 = always, where the highest mean score 3.71, 3.63 and 3.67 for University A (Table 32), University B (Table 33), and University C (Table 34) respectively showed the frequency of activities conducted out of class by the participants. In other words, on average, students in all three universities *sometimes* or *often* practiced listening outside of class.

Table 32

Frequency of Practicing English Skills out of Class at University A

	Listening	Reading	Writing	Speaking	Model Tests
N	173	173	173	173	173
Mean	3.71	3.09	1.48	2.39	2.95

Table 33

Frequency of Practicing English Skills out of Class at University B

	Listening	Reading	Writing	Speaking	Model Tests
N	125	125	125	125	125
Mean	3.63	2.63	1.50	2.31	2.33

Table 34

Frequency of Practicing English Skills out of Class at University C

	Listening	Reading	Writing	Speaking	Model Tests
N	116	116	116	116	116
Mean	3.67	2.54	1.91	2.33	2.97

Therefore, this study indicates that the reformed CET-4 has exerted a great deal of influence on the content of student learning, as well as teaching by instructors in the areas of listening, writing and reading. However, on closer examination, the degree of the washback effects differed between the three universities, as illustrated in the next section.

Different Washback Effects of the CET-4 on University A

Frequency of practicing English writing in class at University A. In Figure 5 below, writing, in the students' learning, was ranked as No. 3 at University A, and No. 5 at University B (Figure 6) and No. 4 at University C (Figure 7). This indicated that not only had more time and attention been paid to writing skill at University A in teachers' teaching, but likewise in students' learning. Students interviewed at University A also felt that writing was the aspect that had been more stressed in their English learning, compared to the participants at University B and University C.

An ANOVA test indicated that there was a statistically significant difference in the three universities in terms of the frequency of doing writing practice in class ($p < .001$). Post hoc comparisons indicated that the mean score for University A ($M = 3.08$, $SD = .845$) was significantly different from University B ($M = 2.40$, $SD = .783$), $p < .001$. The mean score for University A was significantly different from University C ($M = 2.84$, $SD = .923$), $p = .044$. The mean score for University B was significantly different from University C, too with $p < .001$.

Therefore, students at University A did significantly more writing practice than their counterparts at University B and University C. Students at University C did significantly more writing practice than that at University B. Students at University B then did significantly less writing practice among the three universities. In the interview, B. J. Li and B. W. Wang and B. J. Li from University B reported that their English teacher never assigned any homework to them. C. J. Li and C. J. Xu from University C, on the other hand, were required to keep English journals bi-weekly besides occasional writing assignments.

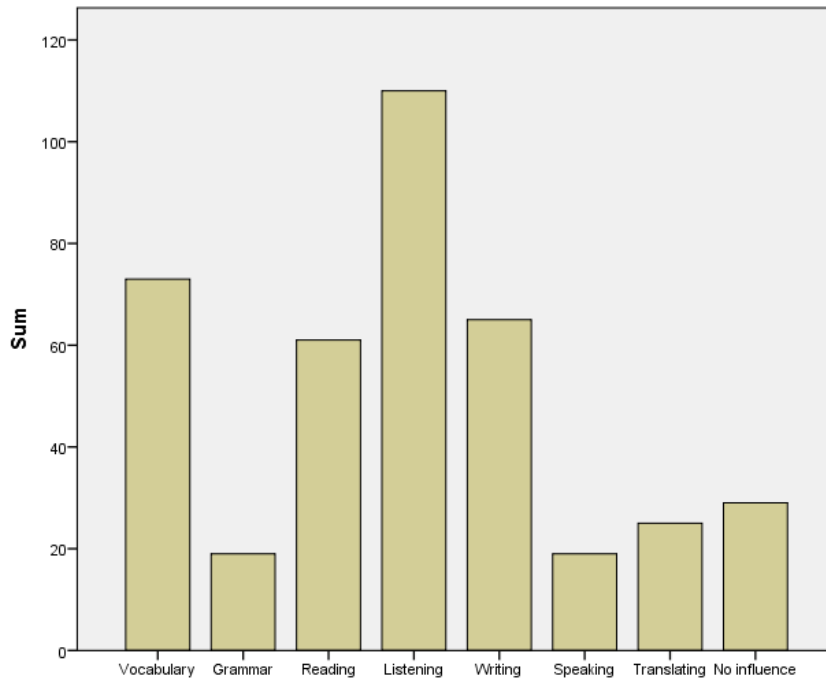


Figure 5. Frequency of practicing different English skills at University A.

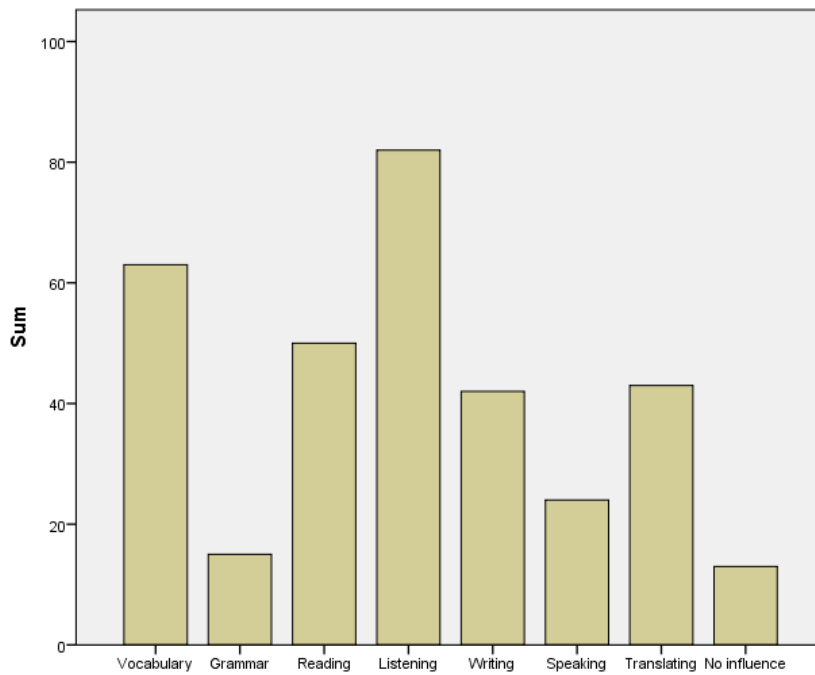


Figure 6. Frequency of practicing different English skills at University B.

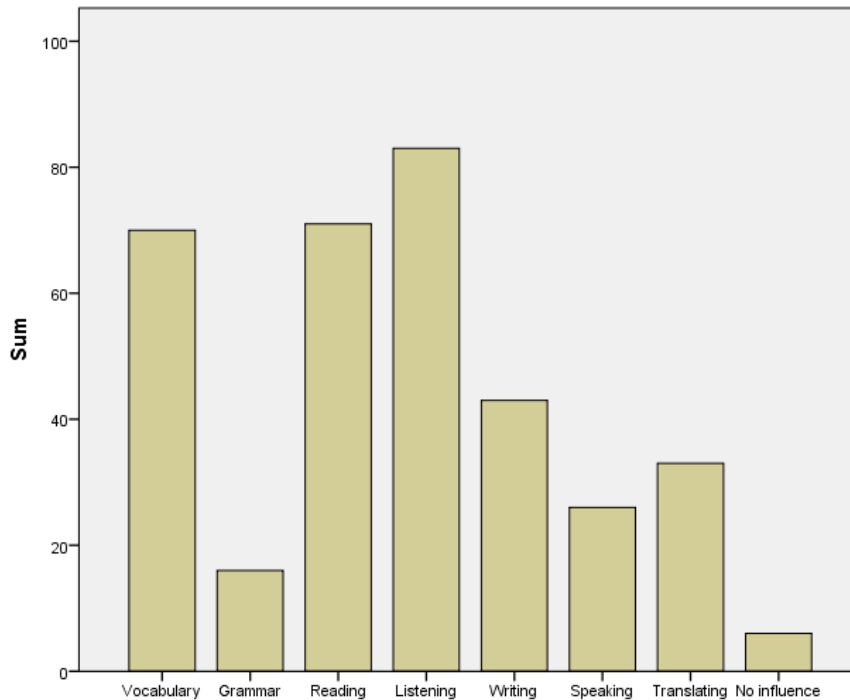


Figure 7. Frequency of practicing different English skills at University C.

Frequency of doing fast reading practice at University A. In class, unlike students at University B or University C, students at University A did not do so much practice in fast reading. For instance, the survey data suggested that only 4.6% students at University A *often* had fast reading practice in class, while 22.4% at University B and 28.4% at University C *often* had fast reading. Also, 0.6% at University A, 4.8% at University B, and 2.6% at University C *always* had fast reading in class.

An ANOVA test indicated that there was a statistically significant difference when it came to the frequency of doing fast reading in class between the three universities ($p < .001$). Post hoc comparisons indicated that the mean score for University A ($M = 2.32$, $SD = .789$) was significantly different from University B ($M = 2.69$, $SD = 1.139$), $p = .003$. The mean score for University A was significantly different from University C ($M = 2.92$, $SD = .988$), $p < .001$. There was no significant difference

between University B and University C. Therefore, students at University A did significantly less fast reading in class than their counterparts at University B and University C.

Frequency of doing intensive reading practice at University A. In terms of intensive reading practice, survey data suggested that only 6.9% of students at University A, but 16.8% at University B, and 27.6% at University C *often* did intensive reading in class. 1.7% at University A, 2.4% at University B, and 1.7% at University C *always* did intensive reading in class.

An ANOVA test indicated that there was a statistically significant difference when it came to the frequency of doing intensive reading in class between the three universities ($p < .001$). Post hoc comparisons indicated that the mean score for University A ($M = 2.36$, $SD = .883$) was significantly different from University B ($M = 2.62$, $SD = .989$), $p = .039$. The mean score for University A was significantly different from University C ($M = 3.01$, $SD = .839$), $p < .001$. The mean score for University B was also significantly different from University C, $p = .003$. Therefore, students at University A did significantly less intensive reading in class than their counterparts at University B and University C. Students at University C, however, did significantly more intensive reading than students at University A and University B. Moreover, University C put more attention and effort towards the reformed CET-4 in some other aspects as follows.

Different Washback Effects of the CET-4 on University C

Frequency of doing intensive reading practice in class in June and December respectively at University C. As it was pointed out above, students at University C did significantly more intensive reading than students at University A and University B. The

survey data suggested that 27.6% at University C, 16.8% at University B, and 6.9% at University A *often* practice intensive reading in class. (In the reformed CET-4, intensive reading section takes 15% of the total score. 25 minutes are given to complete the in-depth reading, which is what intensive reading mainly prepares for.)

An ANOVA test indicated that for students who registered to take the CET-4 in June, there was a statistically significant difference when it came to the frequency of doing intensive reading in class between the three universities ($p < .001$). Post hoc comparisons indicated that the mean score for University A ($M = 2.33$, $SD = .847$) was significantly different from University C ($M = 2.99$, $SD = .868$), $p < .001$. The mean score for University A was significantly different from University B ($M = 2.87$, $SD = .981$), $p = .003$. There was no significant difference between University B and University C. Therefore, in class where students who registered to take the CET-4 in June at University A, they did significantly less intensive reading than their counterparts at University B and University C. This result, nevertheless, was in accordance with the fact that students at University A were not allowed to take the CET-4 until December of their sophomore year.

An ANOVA test indicated that for students who registered to take the CET-4 in December, there was a statistically significant difference when it came to the frequency of doing intensive reading in class between the three universities ($p < .001$). Post hoc comparisons indicated that the mean score for University A ($M = 2.39$, $SD = .910$) was significantly different from University C ($M = 3.05$, $SD = .795$), $p < .001$. The mean score for University C was significantly different from University B ($M = 2.44$, $SD = .963$), $p = .002$. There was no significant difference between University A and

University B. Therefore, in class where students who registered to take the CET-4 in December at University C, students did significantly more intensive reading in class than their counterparts at University A and University B. So, even though students at University A were the ones among the three universities who were required internally to take the test in December, students who registered the test in December at University C did significantly more intensive reading in class among the three universities.

Frequency of doing fast reading practice in class in June and December respectively at University C. An ANOVA test indicated that among students who registered the test in June, there was a statistically significant difference when it came to the impact of the reformed CET-4 on the frequency of doing fast reading in class between the three universities ($p < .001$). Post hoc comparisons indicated that the mean score for University A ($M = 2.27$, $SD = .760$) was significantly different from University B ($M = 3.25$, $SD = 1.142$), $p < .001$. The mean score for University A was significantly different from University C ($M = 3.01$, $SD = 1.000$), $p < .001$. There was no significant difference between University B and University C. Therefore, in classes where students registered to take the CET-4 in June at University A did significantly less fast reading in class than their counterparts at University B and University C.

An ANOVA test indicated that among students who registered the test in December, there was a statistically significant difference when it came to the frequency of doing fast reading in class between the three universities ($p < .005$). Post hoc comparisons indicated that the mean score for University A ($M = 2.35$, $SD = .825$) was significantly different from University C ($M = 2.76$, $SD = .958$), $p = .034$. The mean score for University C was significantly different from University B ($M = 2.28$, SD

= .953), $p < .016$). There was no significant difference between University A and University B. Therefore, in classes where students registered to take the CET-4 in December at University C did significantly more fast reading in class than their counterparts at University A and University B.

So, among students who registered the CET-4 in both June and December, there was a significant statistical difference between University A and University C regarding the frequency of doing fast reading in class. What is more, students at University C did significantly more fast reading in class before both test dates. This suggested that the adding of fast reading to the reformed CET-4 had had strong washback effects on students and English teachers involved at University C.

Frequency of practice listening exercises in class at University C. In terms of listening practice in class, each university was different: 4.3% *always* did listening practice in class at University C, but 32.8% did so at University B, compared to 23.1% at University A. according to students, listening practice occurred *often* in the following percentages: 54.9% at University A, 46.4% at University B, and 37.9% at University C

An ANOVA test indicated that there was a statistically significant difference between the three universities ($p < .001$) in the frequency of practicing listening in class. Post hoc comparisons showed the mean scores for University A to be ($M = 3.93$, $SD = .853$) as opposed to University C ($M = 3.26$, $SD = .896$), $p < .001$. The mean score for University B ($M = 4.06$, $SD = .840$) was significantly different from University C, $p = .003$. However, the mean score for University A was not significantly different from University B. Therefore, among the three universities, students at University C did significantly less English listening practice in class.

Frequency of doing mock CET-4 tests or past CET-4 tests in class at University C. The survey data suggested that 24.1% of teachers at University C, 4% at University B and 3.5% at University A *often* do mock testing in class; and 2.6% at University C *always* do it, while no teachers (0%) at Universities A and B *always* do this type of practice in class.

Frequency of doing mock CET-4 tests/Past CET-4 tests in class in June and December at University C. On the one hand, an ANOVA test indicated that for students who registered to take the CET-4 in June, there was a statistically significant difference when it came to the frequency of doing mock CET-4 tests/past CET-4 tests in class between the three universities ($p < .001$). Post hoc comparisons indicated that the mean score for University A ($M = 2.10$, $SD = .725$) was significantly different from University C ($M = 3.23$, $SD = .820$), $p < .001$. The mean score for University C was significantly different from University B ($M = 1.85$, $SD = .744$), $p < .001$. There was no significant difference between Universities A and B. Therefore, students who registered to take the CET-4 in June at University C did significantly more mock CET-4/past CET-4 tests than their counterparts at Universities A and B.

On the other hand, the same statistical tests indicated that for students who registered to take the CET-4 in December, the mean score for University A ($M = 1.77$, $SD = .866$) was significantly different from University C ($M = 2.29$, $SD = .835$), $p = .003$. The mean score for University C was significantly different from University B ($M = 1.75$, $SD = .835$), $p = .004$. Again, there was no significant difference between University A and University B.

So, students who registered for the reformed CET-4 in both June and December at University C did significantly more mock CET-4 tests/past CET-4 tests in class than those at University A and University B. This further suggested that the reformed CET-4 had had strong washback effects on students and English teachers in classes at University C not only in their first year but also in their second year.

Frequency of doing CET-4 mock tests or past CET-4 tests out of class at University C. In addition, the survey data suggested that out of class, 41.4% at University C, 24.8% at University B and 24.3% at University A *sometimes* do CET-4 mock tests or past CET-4 tests by themselves. 17.2% from University C, 9.6% from University B, and 4% at University A *often* do it; 2.6% at University C, 0.8% at University B and 0.6% at University A *always* do it.

An ANOVA test indicated that there was a statistically significant difference in the three universities in terms of the frequency of doing the CET-4 mock tests or past CET-4 tests out of class ($p < .001$). Post hoc comparisons indicated that the mean score for University A ($M = 2.06$, $SD = .857$) was significantly different from University C ($M = 2.71$, $SD = .987$), $p < .001$. The mean score for University C was significantly different from University B ($M = 2.13$, $SD = 1.016$), $p < .001$, but there was no significant difference between University A and University B.

Given that students who registered the CET-4 in both June and December at University C did significantly more mock CET-4 tests/past CET-4 tests in class than those at University A and University B, the reformed CET-4, therefore, had quite strong washback effects on students and English teachers in class and out of class at University C.

Frequency of mentioning CET-4 in class of June CET-4 test takers at

University C. University C was also the university that more frequently mentioned the CET-4 in its classes because 11%, 20.8%, and 39.7% *often* mentioned the CET-4 in class at University A, University B and University C respectively, while 0%, 4.8% and 6.0% *always* mentioned the CET-4 at Universities A, B and C respectively. The survey data suggested that when it was broken down into the time when students registered to take the CET-4, for students who registered to take the CET-4 in June 10%, 7.5% and 48.6% from University A, B and C respectively *often* mentioned the CET-4 in class. For students who registered to take the CET-4 in December, 11.7%, 30.6% and 23.8% from University A, B and C respectively *often* mentioned the CET-4 in class.

An ANOVA test indicated that among students who registered to take the CET-4 in June, regarding the frequency of teachers' mentioning the CET-4 in class at the three universities ($p < .001$), there was a statistically significant impact due to washback from the reformed CET-4. Post hoc comparisons indicated that the mean score for University A ($M = 2.71$, $SD = .725$) was significantly different from University C ($M = 3.58$, $SD = .811$), $p < .001$. The mean score for University C was significantly different from University B ($M = 2.70$, $SD = .749$), $p < .001$. There was no significant difference between Universities A and B. Therefore, among students who registered to take the CET-4 in June at University C, teachers mentioned the CET-4 significantly more often in class than their counterparts at University A and University B. Given that most students at University C were scheduled to take the CET-4 in June, this finding further indicates that the CET-4 had stronger washback effects on English teachers at University C than at the other two universities.

Frequency of doing group discussion/pair work in English at University C.

An ANOVA test indicated that there was a statistically significant difference class between the three universities ($p = .001$) when measuring the impact of the reformed CET-4 on the frequency of doing group discussion/pair work. Post hoc comparisons indicated that the mean score for University A ($M = 2.91$, $SD = .910$) was significantly different from University C ($M = 2.52$, $SD = .849$), $p = .001$. The mean score for University B ($M = 2.88$, $SD = .972$) was also significantly different from University C, $p = .006$. However, there was no significant difference between Universities A and B. Therefore, students at University C did significantly less group discussion or pair work than those at Universities A and B.

Different Washback Effects of the Reformed CET-4 on University B

Frequency of mentioning CET-4 in class of December CET-4 test takers at University B. An ANOVA test indicated that among students who registered to take the CET-4 in December, there was a statistically significant difference between the three universities ($p < .001$) when measuring the impact of the reformed CET-4 on the frequency of teachers' mentioning the CET-4 in class. Post hoc comparisons indicated that the mean score for University A ($M = 2.64$, $SD = .79$) was significantly different from University B ($M = 3.10$, $SD = 1.037$), $p = .003$. There was no significant difference between Universities A and C ($M = 2.88$, $SD = .832$), or between Universities B and C. Therefore, among the three universities, teachers at University B, where students registered the CET-4 in December, mentioned the CET-4 most frequently in class. The fact that both first-year students and second-year students were allowed to register for the CET-4 Ability Training class could have made the difference here.

Frequency of doing listening exercises in December at University B. Similarly, an ANOVA test was conducted to explore the impact of the reformed CET-4 on the frequency of classroom listening practice for students who registered for the CET-4 in December at the three universities. There was a statistically significant difference at the $p < .001$. Post hoc comparisons indicated that the mean score for University A ($M = 3.89$, $SD = .969$) was significantly different from University B ($M = 4.33$, $SD = .805$), $p = .006$. The mean score for University A was significantly different from University C ($M = 3.17$, $SD = 1.01$), $p < .001$. University B also significantly differed from University C, $p < .001$. Therefore, test takers who registered to take the CET-4 in December at University B did significantly more listening practice in class than their counterparts at University A and University C. It also suggested that students who registered to take the CET-4 in December at University C did significantly less listening practice in class than their counterparts at University A and University B.

Frequency of student participation in class at University B. Group discussion or pair work is a typical communicative activity in language classes. University A had the highest frequency of student participation among the three universities, measuring at 45.1%, whereas University B measured frequency of student participation at 31.2%; and University C had the least. Only 37.1% said that their English classes *often* included student participation. What is more, 16.8% at University A, 7.2% at University B, and 12.1% at University C *always* had student participation in class. Although it seemed that group work was not often practiced in the three universities, 44.5% at University A, 38.4% at University B, and 36.2% at University C *sometimes* conducted group discussions or pair work in English class.

An ANOVA test indicated that there was a statistically significant difference when it came to the impact of the reformed CET-4 on the frequency of having student participation in class activities in the three universities ($p = .001$). Post hoc comparisons indicated that the mean score for University A ($M = 3.63$, $SD = .965$) was significantly different from University B ($M = 3.19$, $SD = .965$), $p < .001$. There was no significant difference between Universities B and C or between Universities A and C. Therefore, students at University B had significantly less class participation than those at University A, and there was no significant difference between Universities A and C in this matter.

Frequency of English study in the college's Self-Access English Center at University B. Students at both University A and University C made use of the resources provided on campus more frequently: the survey data showed that 22.5% at University A and 25.9% at University C *sometimes* studied English at the college's Self-access English Center. During the study, no students claimed that there were too few of these learning facilities on the three university campuses. Each university had language labs in which students could listen to and study English. However, the Center's physical location on University B's campus, coupled with its restricted access to the internet, were both limitations which were likely to inhibit students from utilizing the facilities. It was also not clear whether any of the self-access English centers remained open after class or what types of English learning resources/services were provided after class.

An ANOVA test indicated that there was a statistically significant difference between the frequency of students' learning English at the college's Self-access Centers located at each of the three universities ($p < .001$). Post hoc comparisons showed that the mean score for University A ($M = 1.90$, $SD = .926$) was significantly different from

University B ($M = 1.50$, $SD = .703$), $p < .001$. The mean score for University B was also significantly different from University C ($M = 1.88$, $SD = .876$), $p = .002$. There was no significant difference between University A and University C. However, students at University B spent significantly less time studying in the college's Self-access Center than those at University A and University C.

Pressure Caused by the CET-4

More interviewees at University A seemed to feel little or no pressure regarding the CET-4 and they showed greater confidence that they would pass the first time they took it. Seven students (A. Jin, A. Z. Zhang, A. Luo, A. Pan, A. Z. Li, A. Jing, and A. Zhu) openly declared that they felt little pressure. A. Z. Li said that she was confident that she could pass the CET-4. A. A. Jing claimed that College English was the course that caused the least pressure for her in college. She did not think the CET-4 was a hard test. Everyone around her had said that the CET-6 was hard, not the CET-4 and she believed what they said. On the other hand, three students (A. Meng, A. Y. Zhao and A. Xie), who came from the less developed Northwestern China, stated that they had felt pressure regarding taking the CET-4, and that they had to do extra work to prepare for it.

B. Dong from University B said that he studied English because he was interested in it, so the CET-4 did not pose much of a problem for him. Seven students (B. M. Xu, C. Ye, W. Wang, B. Ding, B. X. Chen, B. J. Li, and B. Shi) at University B said they had felt pressure brought on by the test, yet they all concluded that the test had had a positive influence on them. In fact, they all agreed that taking the reformed CET-4 had motivated them to study English at college. B. W. Wang, another student, said he did not feel much pressure because he had learned the fundamentals of English before coming to college. He also pointed out that very few students would study English in college if there were

no CET-4. According to him, the existence of the CET-4 reminded college students that they needed to move beyond what they had learned in senior high school by continuing to memorize new words and practicing English.

Students at University C, in general, felt more pressure from the test and the preparation it required. Four students (C. J. Xu, C. R. Zhao, C. Liu and C. Y. Zhang) did not believe that the reformed CET-4 had relieved the pressure of learning English. C. J. Xu pointed out that the CET-4 had intensified the pressure. According to her, where there is a test, there is pressure. Yet she also stated that if there were no CET-4, it would be even harder for college students to learn English, because they would be forced to rely on their own personal motivation.

To students like C. Gao and C. J. Li, despite the pressure, they still felt that the influence of the test was more positive than negative. They also believed that if there were no CET-4, many students at college would not care about studying English. B. Song was the only interviewee at University B who claimed that the test exerted primarily negative influence on him, due to his lack of interest in learning English in the first place.

Abolish the CET-4?

The interviewees did not all agree that the reformed CET-4 would reveal their true knowledge of the English language or that it would relieve the pressure they felt to learn English. If this was the case, they were asked, did they believe that the reformed CET-4 should be abolished? Five students (A. Meng, A. Xie, B. Song, B. M. Xu and C. J. Xu) insisted on abolishing the CET-4. A. Meng was the most decisive one, admitting that he didn't like English because he was not good at it. A. Xie and C. J. Xu also wanted to abolish the test because of their poor English. They felt it was painful to take such a high stakes test. B. M. Xu, one of the students from Northwestern China wanted to abolish all

tests, not just English tests. However, in the second interview in Phase IV (see p. 103), she changed her mind and suggested the CET-4 should not be abolished, because it could measure her English ability. It is also worth noting here that these four students had not been required to take the English listening component on their matriculation test, which put them at a comparatively disadvantaged starting point against those who had started listening practice in high school.

The majority of students argued that the reformed CET-4 should be retained. Of these 16 students, five (A. Qin, B. Ding, C. Y. Zhang, A. Jing, and A. Z. Li) believed that it was an official way of determining English levels among college students. They also claimed that a national test was the best way for students and the society to find out one's weaknesses in English learning. Five other students (B. J. Li, C. Gao, C. J. Li, A. Zhu, and C. R. Zhao) wanted to keep the reformed CET-4 because they thought that few students would keep studying English in college if the test were abolished. A. Zhu emphasized that it should be retained because the test was designed to compete with foreign tests such as Test of English as a Foreign Language (TOEFL) or the International English Language Testing System (IELTS). With so many English learners, China should have its own English assessments to cater to the characteristics of English learners in China. Two other students (B. Shi, and A. Luo) argued that as long as a test existed, there was a reason. In light of this logic, the reformed CET-4 should be retained. Four students (C. Ye, A. Z. Zhang, A. Jin, and A. B. Zhang) claimed that the reformed CET-4 score served as proof of English proficiency for college students. To them, one more proof would do no harm, so anyone who wanted to take the CET-4 should be given the opportunity to take it.

A Short Summary of the Responses to Research Question Three

The reformed CET-4 influenced students' perceptions of English teaching and learning at the college level. According to the study, the majority of participants reported that English teaching and learning content had changed based on the reformed CET-4. Of course, the intensity of the washback effects generated by the test varied depending on which of the three universities the students attended. Students and teachers at University C, in particular, spent significantly more time on preparation for the reformed CET-4 than students at the other two universities. Three positive washback effects targeted by the reform of the CET-4 and the new College English courses were: greater use of English as the language of instruction, use of computers to assist teachers in and out of the classroom, and greater emphasis on English listening skills. And although the CET-4 was controversial, the majority of participants did not want to abolish it. Further discussion of the various washback effects and suggestions for future study will be provided in Chapter 6.

CHAPTER 6

CONCLUSIONS, DISCUSSION AND SUGGESTIONS

Conclusions

Almost all the college students in this study acknowledged the essential status of English in the development of the global economy, politics, and cultural exchange. Nevertheless, in relation to their own lives, English is largely considered a means to a utilitarian end such as college graduation, employment, and/or potential assistantship/scholarship. Neither college education nor the society in China has given these students viable opportunities to use the English language. For them, knowledge of English is reduced to a test score which serves as a gatekeeper to future activities. Therefore, the majority of students cannot visualize a genuine connection between English and their major field of study. For those students who want to improve their communicative English skills, the College English course does not meet expectations. On average, they spend only half an hour outside of class studying English on a daily basis. The data indicate that students who plan to study abroad in the future devote significantly more time to studying English than students who have other graduation plans, mostly because the former must maintain an impressive overall GPA and they also have to take other language exams, such as Test of English as a Foreign Language (TOEFL), or International English Language Testing System (IELTS), and/or Graduate Record Exams (GRE).

As to the reformed CET-4, a surprisingly large number of participants know little about the modified policy and the background of the reformed CET-4. Some students do not question outdated or incorrect test information given by teachers and administrators.

So, more effort should be made on the university's part to provide their students the correct, up-to-date and systematic information regarding the reformed CET-4.

Nevertheless, overall, it is still safe to conclude from the results of this study that the reformed CET-4 has had washback effects on college English teaching and learning in China.

The majority of students at the three universities stated that the reformed CET-4 had made an impact on their English studies and on their teachers' English instruction: 88.4 % reported that the CET-4 had an influence on how they studied, and 87.7 % claimed that it had influenced their instruction. Teaching content and teaching methods were the two areas that had changed the most in the classroom. For instance, in all three universities, English teachers initiated test reviews before the testing date. In addition, students reported that college English teachers used English as their language of instruction much more frequently than their teachers at senior high school and middle schools. Multi-media support was also frequently used in College English classes.

As to English learning, the CET-4 appeared to have the most impact on what students study. The majority of participants made plans to prepare for the test, by buying the reformed CET-4 vocabulary books or by taking sample tests or past tests, which were either distributed by their teachers in class or bought from local bookstores. According to the survey, the three components of language learning that were most influenced by the CET-4 were listening, vocabulary, and writing. Similarly, English teachers were influenced in their listening, writing and reading instruction. However, the influence exerted on their reading instruction stood in contrast to the students', who were more motivated by the CET-4 to learn new vocabulary than to apply themselves to the broader

discipline of reading. On the other hand, students did say that their college English teachers taught new vocabulary in the context of reading, to a greater degree than teachers at senior high school or middle school.

It was not clear why writing was considered an area that was most influenced by the CET-4 in both teaching and learning. Students at all three universities did express concern about their English-writing skills; however, the data suggest that only at University A did teachers and students spend substantial time cultivating writing ability and writing skills. The other two universities reportedly took less time on teaching writing. This finding will be discussed further below. On the other hand, almost all the participants pointed out that listening practice was emphasized in college, both in class, supervised by their teachers, and out of class on their own. The increased time spent on listening to English is obviously one of the positive washback effects targeted by designers of the reformed CET-4.

According to the interviewees, memorizing vocabulary, listening, and practicing banked cloze exercises were the areas that they covered most outside of class without teacher supervision. As opposed to the impact made on the instructional methods of English teachers, the data show that the CET -4 influenced only 29.1% of participants to modify their learning methods and strategies. Most interviewees stated that they continued to use learning methods they had used in high school, such as note-taking in class, rote memorization of English vocabulary, and model test practices.

Some even regretted that in college they did not retain the mandatory readings that were a morning ritual in senior high school and middle school. The major difference in how students learned was that, lacking direct parental supervision, they watched

Hollywood movies more often than before. Many considered watching English movies/TV programs as a form of listening practice, even though these movies often had Chinese subtitles.

So this study shows that a test will influence the rate and sequence of leaning (Hypothesis 8, Alderson & Wall, 1993, p. 120), a test will influence the degree and depth of learning (Hypothesis 10, Alderson & Wall, 1993). And a test will influence the degree and depth of teaching (Hypothesis 17, Alderson & Wall, 1993). Besides, a test will influence attitudes to the content, method, etc., of teaching and learning (Hypothesis 11, Alderson & Wall, 1993).

Inferential statistical analysis further indicated that students at University A did significantly less fast reading and intensive reading in class than their counterparts at University B and University C. Of the three universities, University C classes did significantly less group discussion or pair work. On the other hand, students at University A did significantly more writing practice than their counterparts at the other two schools. Students who registered to take the reformed CET-4 in both June and December at University C did significantly more mock/past tests in class than those at Universities A and B. These students also did significantly more practice tests outside of class, and their teachers mentioned the CET-4 significantly more times to students who were registered for the June CET-4.

However, teachers at University B mentioned the test significantly more times in class and did significantly more listening practice with students who were registered to take the reformed CET-4 in December than their counterparts. This conclusion was consistent with data regarding student motivation since the majority of survey

participants at University A reported that they were more interested in improving their comprehensive communicative ability and passing required College English courses, whereas those at Universities B and C were more concerned with getting good grades on the CET-4.

Therefore, the reformed CET-4 had the strongest washback effects on both teachers and students at University C. Thus, this study shows that “Tests will have different amounts and types of washback on some teachers and learners than on other teachers and learners” (Alderson & Hamp-Lyons, 1996, p. 296). However, it does not support Hypothesis 14: Tests will have washback on all learners and teachers (Alderson & Wall, 1993, p. 120). Moreover, this study suggests that tests will have stronger washback effects on teachers and students from schools of lower rank than schools of higher rank. What is more, positive washback effects are more likely to occur in receptive language skills that are direct and easily practiced such as fast reading and listening.

Discussion and Suggestions

Students Spending Less Time and Effort on their English Studies

While the importance of the English language has been endorsed by almost all college students, they appear to spend less time and effort every day studying English than when they were in high school. Based on this study, four factors contribute to this behavior: a) heavy study load at college; b) lack of teacher guidance; c) low status of College English courses, and d) perceived relatively low pressure caused by the reformed CET-4. Each of these factors will be discussed below:

Heavy Study Load

In the study, the interviewees constantly referred to their heavy study load. In

particular, in Phase IV (see p. 105), four interviewees from University A averaged 36.88 credits that semester, or, individually the number of credits they reported were 33.5, 37, 39 and 38. (Students are expected to take a minimum of 24 credits every semester, and a total of 100 – 200 credits to graduate.) In the previous semester, on average they took approximately 36 credits, including a 3 - credit English course. As science majors, they also have complementary lab classes. So many participants were busy taking classes, reviewing lessons, and doing homework and/or writing up lab reports every day.

Lack of Teacher Guidance

The survey data showed that the majority, 62.6% participants in the three universities had the mindset that it was their English teacher's obligation to help students prepare for the reformed CET-4. Almost unanimously, the interviewees reported that their college English teachers were far less strict and provided less guidance than teachers at senior high or junior high school. Students often worked on their own if they wanted to improve their English.

Some even felt that their teachers lacked passion for their jobs. Five interviewees (B. J. Li, A. Luo, B. W. Wang, C. Liu and B. X. Chen) pointed out that their English teachers never assigned any homework. B. J. Li from University B said, "My college English teacher was good. She is nice, too but she never assigns any homework." At University C, according to C. Liu, his English teacher never checked students' homework. Teachers in the three universities did not have office hours according to the school policy. They often came to school when they had classes or meetings and left when they finished teaching. In general, meeting only twice a week for classes, there was no more close relationship between students and English teachers like before.

What is more, at senior high, everyone: teachers, administrators, parents, and students strive to ensure that students score well on their college entrance exams. In contrast, in college, teachers, administrators, and parents would simply remind students not to waste time and to study independently. The participants maintained that it was hard for them to practice independent learning after they had been so closely supervised in senior high school. They had a hard time managing time wisely on their own, especially when faced with large chunks of free time that they had to manage by themselves for the first time. They also mentioned more than once that they saw students who stayed in the dorm most of their time, playing video games, chatting online or watching movies.

Under such circumstances, some students suspected that they were not learning much in college, and their English levels declined compared to what they had learned in senior high school. C. Y. Zhang from University C, for instance, doubted she had learned anything new at college. She wished there was a reformed CET-4 intensive prep class, and continued to take College English classes after taking the CET-4. C. J. Xu also claimed that she did not learn anything from college English courses. According to her,

I was very goal-oriented before college, but not in college. Without a teacher pushing me, I don't even expect to pass the CET-4 this time in June. I wanted to study English more efficiently, but I don't know how to.

Later, she did not show up for the June test even though she had registered for it. She explained that she was busy serving as a volunteer at the Shanghai World Expo, and she was not ready for the test at that time. Students like A. Jing, in contrast, had mixed feelings: "Since we are adults, I don't think our English teachers should push us hard. However, because I'm not diligent enough, my English ability has fallen behind a lot."

With so many students complaining about the lack of teachers' help and guidance, schools should spend more time nurturing autonomy on the part of student learners. On the one hand, the colleges can provide assistance on developing students to be independent learners before letting them sink or swim. Presentations, workshops, or motivational speeches might help push them toward that goal. On the other hand, teachers could be required to offer office hours for students so that student questions or concerns could be solved face to face. Office hours are also a good way for teachers to get to know their students, give more personalized advice, and adjust their teaching accordingly for the whole class.

College English Course Not Challenging/Important Enough

A. Z. Li from University A said that since University A was science- and engineering-oriented, students generally spent more time on courses such as Math, Physics, Linear Algebra, etc. English was not considered an essential subject by many students. She further claimed that it would be incorrect to say that students were not paying attention to their English studies; rather, many students felt they did not need to cram for the CET-4, because they could pass the test with their prior knowledge of English fundamentals. They did not learn Higher Mathematics in high school, and college math was much more difficult than high school math. They could not pass the test of Higher Mathematics if they did not work hard on it. So for herself, she was quite confident that, “I just would not be panic for English tests. The bottom line is that I probably won’t get a very high score if I don’t cram for English finals. I know, however, I won’t fail, either.” C. Ye from University C expressed a similar opinion: In the belief that college English was an easier subject, he would always study for other subjects if a

conflict arose. B. X. Chen from University B also said that as a science major, if he/she skipped science class and fell behind, he/she might not be able to catch up for the rest of the semester. However, skipping an occasional English class would not have an appreciable effect on one's overall English skills.

A. Zhu from University A said that he did not feel he had learned anything in College English. He bet that some of his fellow students would even say that their English was at a lower level now than it was in senior high schools. He stated that College English was not an important course in students' eyes. He admitted that College English was always the course that he had pushed aside when he was busy. However, he emphasized that "it's not because students do not pay attention to English – the language; it's the College English course that we do not care much about."

Due to the low value attached to College English courses, some students simply chose to ignore them completely, or to pay only partial attention, if they did not actually skip class altogether. For instance, even though they might be physically in class for every session, they were inattentive or even sleeping in class. B. Song from University B said he attended College English classes only because of the university attendance policy, which required him to take the class if he wanted to take the English final exam. A. Liang, a participant from University A, admitted that he always used his time in College English to memorize the GRE vocabulary because he had registered for the GRE in June of the following year. The thick red GRE vocabulary book was very eye-catching on the desk, but he said the teacher never said anything about it. He would sometimes do listening practice when the teacher played VOA listening exercises in class. A. Tang, another student who was preparing for the GRE at the time, commented that whether

one's English was good or not had little to do with the teacher. She argued that "Math teachers can teach students creative thinking or new approaches for solving math problems, but English teachers could not." With a busy schedule, she said most of her time in College English class was also devoted to memorizing GRE vocabulary.

Students Feeling Less Pressure Taking the Reformed CET-4

Most students argued that the reformed CET-4 caused less intense pressure when compared to the pressure caused by the college entrance exam. There is no doubt that the entrance exam is one of the most high-stakes tests in China. In 2010, statistics showed that 9,570,000 students registered for the national entrance exam to college and the passing rate was only 68.7% nationwide (Entrance Exam Registration, 2012). With lower stakes, students chose to pay far less attention to the reformed CET-4. C. R. Zhao, for instance, said that she did not think her college English course was as important as the English courses at senior high, because courses were scheduled every day in senior high and in college English classes only met twice a week. With less class time, a busy schedule with classes in their major, students generally spent less time on English. C. J. Li, on the one hand, certainly had felt the pressure of the CET-4, because she remembered that her English teacher had said that the goal for college English learning was to pass the reformed CET-4. Yet she also stated that the entrance exam to college was more stressful since most students only take it once or twice, whereas there are more opportunities for students to take the reformed CET-4. [Theoretically, one could take the reformed CET-4 at least 6 times.] C. S. Chen concurred that the pressure now was not as heavy as passing the entrance exam to college, because the college entrance exam included four other subjects, which had to all be taken at the same time.

So, on the whole, the majority of students have acknowledged the significance of English as a lingua franca and the importance of the reformed CET-4 score in the course of getting the college degree and seeking employment. However, due to the relatively lower level of pressure caused by the reformed CET-4, less attentive English teachers, and heavy major-related study loads at college, many of the students did not think highly of the College English courses and spent less time studying English at college than in high school. Modifying College English courses to motivate students to learn English, and catering to college students' needs, especially their communicative needs, should be evaluated by the College English Test Committee and added to its agenda sooner rather than later.

Why Did the Reformed CET-4 Have Different Washback Effects at University A?

Why did the reformed CET-4 have rather different washback effects on students at University A among the three universities? In other words, why were students at University A more concerned about passing and getting high scores in required English courses than getting good grades in the reformed CET-4? First of all, data indicate that there exists fierce internal competition at University A. As a top-five university in China, most students at University A are top high-school students in each province nationwide. Being accustomed to ranking No. 1 in their high schools, they still want to maintain the high expectations at college that have always been set for them. They know that getting good grades in each course will lead to a high GPA. And getting a higher GPA can prepare them to study abroad and/or obtain a graduate assistantship in the future.

A. Qin reported a well-known unofficial trend at University A: upon graduation, 1/3 of the graduates go directly to work, 1/3 go to graduate school in China, and 1/3 go to

graduate school abroad. Among the interviewees, as mentioned earlier, two sophomores had registered for the Graduate Record Examinations (GRE) in June 2011 and were preparing for the GRE, and another sophomore was preparing for the Test of English as a Foreign Language (TOEFL). Moreover, a number of interviewees had already purchased the TOEFL vocabulary book, International English Language Testing System (IELTS) and/or GRE vocabulary book as an option for future study abroad. At University B, no interviewees reported preparing for these tests. At University C, one participant from Southern China had just taken the TOEFL and was in the process of applying for an undergraduate program in the United States. The survey data correspondingly showed that more students (17.3%) at University A planned to study abroad after graduation. This percentage was above 12.8% at University B and well above 5.2% at University C.

Data analysis in Chapter 5 suggest that students who planned to further their study abroad committed significantly more time to English learning. Nevertheless, when students planned to study abroad, they tended to shift their focus to tests such as the TOEFL, IELTS and GRE – which are required for foreign students who plan to study in English-speaking countries – rather than the reformed CET-4. In terms of breadth and depth, the TOEFL, IELTS and GRE are all more demanding than the reformed CET-4. In this sense, listening attentively in College English courses could help strengthen students' overall English skills. It would enable students to approach the TOEFL, IELTS and GRE with greater confidence in the long term. What is more, obtaining a good grade in their English courses looks more impressive on their college transcript, which will be useful for possible graduate-level study or job hunting when they graduate.

So it is very likely that students' graduation plans are the second major factor that

has weakened the impact of the reformed CET-4 at University A rather than the other two universities. Data indicate that the majority of the students - 35.8% at University A and 39.2% at University B planned to attend graduate school in China. However, in order to be admitted to a graduate school in China, one must take the National Graduate School Examinations (NGSE). To pass the English exam, which is a mandatory component of the NGSE, students' English level should at least reach a level similar to that of the College English Test Band 6 (CET-6), a standard stipulated by the Ministry of Education (MoE). For this reason, more students at University A and University B considered the CET-6 to be more important than the CET-4. Therefore, they set themselves the goal of passing the CET-6 as their ultimate aim for learning English learning in college.

According to the survey, as pointed out in Chapter 5, 86.7% at University A, 77.6% at University B, yet only 59.5% of students at University C chose the CET-6 over the CET-4 as the more important test. When students considered the CET-6 to be more important and were determined to pass it, they tended not to work hard to get a high score on the reformed CET-4. Rather, they would get just the minimum score required on the reformed CET-4 (425) so that they could move on to taking the CET-6. Once they got a desirable score for the CET-6, they no longer cared about their score on the CET-4. This, therefore, may be another reason why the reformed CET-4 has less impact on students' English learning at University A.

Thirdly, University A is the only university among the three that has an established internal English proficiency test, which determines whether sophomores who are not English majors should continue taking college English courses in their second year. The participants said that this internal English proficiency test came into being

because a great number of teachers and administrators at University A insisted that students with sufficient English proficiency should not spend as much time on English as on their major studies. Once students pass the internal proficiency test, they are exempt from taking general English courses offered by the English Department, allowing them to focus mainly on courses required by their majors, and using English as a tool to absorb more knowledge about their major.

The proficiency test was scheduled to be given a few weeks before the reformed CET-4. Some participants at University A had just taken it in Phase II. The test includes note-taking, fast reading, in-depth reading and two timed essays. In Phase IV (page 103), when A. Qin was interviewed for the second time, he pointed out that the internal English proficiency test had been administered just days after students took the reformed CET-4. The score of the internal test, however, would be counted as part of the final grade of their English class that semester. According to A. Qin, his English teacher absolutely paid more attention to the internal English proficiency test than the reformed CET-4 in class, because the teacher was confident that the reformed CET-4 was an easy test for her students. She said that if the students could pass the internal English proficiency test, the reformed CET-4 should be “a piece of cake” for them. Given that the internal English proficiency test was far more difficult than the reformed CET-4, A. Z. Li also claimed in her second interview that, to her, the former was more important than the reformed CET-4.

In University A, teachers of English actually gave students practice with the internal proficiency test. On average, students said that it was almost impossible for students in Ordinary Classes to pass the test. One student claimed that the passing rate for

one of the Fast Classes was as low as 4%. Another student recalled that she almost fell asleep while listening to the long report from the listening section of the test. She only had 10 or 15 minutes to write two essays. What was worse, she said that penalty points were applied in the reading section if the questions answers were answered incorrectly rather than left blank. So in order to avoid the penalties, students preferred to leave the answer sheet blank if they were uncertain of the answers.

Some students at University A said that they believed that the English teachers deliberately set the bar high so that students would have to keep taking college English courses and the teachers could keep their jobs. On the other hand, during the interview, one student who had passed the internal proficiency English test in the past semester still chose to take College English because he was afraid to fall behind if he studied English on his own. Whether it is a good idea to have an internal English proficiency test is a question outside the scope of this study, but it was clear that the implementation of this test motivated students at University A to set themselves more challenging goals regarding English than simply passing the reformed CET-4. Hence, the reformed CET-4 had less impact on students at University A.

A fourth factor influencing the impact of the reformed CET-4's on University A could be the University's distinctive English course offerings for second year students. As mentioned in the last chapter, basic College English courses are not offered at University A after the freshman year. Instead, sophomores are free to choose from a range of more specific English courses, such as English Literature, Public Speaking in English, English Writing, English Translation, and Audio and Visual English, etc. Compared to the one-size-fits-all general College English course offered at the other universities, these tailored

English courses encourage students at University A to spend more time on learning English and developing genuine communicative English skills rather than focusing exclusively on studying for the reformed CET-4.

Besides, students at University A have been offered to take English language courses within their major field of study. For instance, nine participants from University A had either taken semi-bilingual courses in their majors, or courses that were fully bilingual. Their required textbooks and resources were books exported from English-speaking countries. As a result, many interviewees from University A explicitly related English learning to their major course of study. They emphasized that journals which publish up-to-date information in their areas of specialization, are written in English. If they want to present their research, communicate with foreign experts, or import advanced technology from foreign countries in the future, a good command of English is a must. One student in particular, A. Jin, was accepted into a special trial track at University A, in which he was completely immersed in bilingual courses since his freshman year. Another student, A. Jing, reported that in her Digital Electronics Foundation course, the required texts were exported English books, and the PowerPoint presented in class by the teacher are created in English. Homework could be done either in Chinese or in English, but tests are given in English, even though the course is taught in Chinese. Four more students from different majors had bilingual courses taught in similar style. They each took between 2 and 5 bilingual courses in fields such as Physics, Math, Engineering, Introduction to Engineering, Microeconomics, Signal and System, Politics, etc. All the teachers are native Chinese speakers who studied and/or graduated from universities overseas. Students like A. Y. Zhao complained that the teachers'

English pronunciation is not authentic and sometimes they could not make themselves understood in class. Regardless of these complaints, compared to interviewees at Universities B and C whose courses within their majors are never taught in English, students at University A have far more contact with English and experience in using the English language in their areas of concentration.

Therefore, it was not surprising to learn that 35.8% students at University A felt a close connection between English and their major, whereas only 23.2% at University B and 23.3% at University C had similar feelings. The majority of students, nevertheless, still did not acknowledge a close connection between learning college English and their major. As such, problems were likely to emerge in their commitment to learning English, due to a failure to apply their communicative English skills after graduation. Conditions permitting, offering more bilingual courses in all the major disciplines should be encouraged.

Last but certainly not least, the relaxed CET-4 preparation environment at University A could also have played a part in the CET-4 having less impact. By admitting top students from each province, University A was likely to have selected students with a good foundation in English. Six students from University A in Phase I and five students in Phase IV stated in the follow-up interviews that they were certain they would have no difficulty attaining the minimum required CET-4 score. Among them, five students (A. Jin, A. Pan, A. Qin, A. Huang, and A. Q. Wang) said that either English teachers or senior students had told them that the reformed CET-4 was an easy test. Five students (A. Jin, A. Z. Zhang, A. Pan, A. Qin, and A. Tang) claimed that the reformed CET-4 was no more difficult than the English matriculation exam, and that they could have passed it in

senior high school. A. Liang said that around him, he only saw students preparing for the TOEFL test, not the reformed CET-4. In contrast, only two interviewees from University B and one from University C expressed absolute confidence. On the other end of the spectrum, one student (C. Ye) from University C, mentioned that he knew three seniors who had failed to obtain 425 twice.

So it is quite possible that the overall campus environment might have an influence on the time students spent on preparation for the CET-4. As Green (2007) predicted in his washback model, if participants do not perceive the importance of a test, and the test is not particularly difficult – as is the case for the reformed CET-4 at University A – the test will not exert the expected degree of intensity as part of the washback effect at University A as it will at the other two universities.

Further Discussion and Suggestions of the Emerged CET-4 Washback Effects

University A, University B and University C – Multi-Media Teaching

As mentioned in the last chapter, computer assisted teaching has been widely adopted in College English classes in the three universities under consideration. This, according to the Ministry of Education (MoE), is a targeted positive washback effect due to the reform of college English instruction initiated in 2003 by the National Higher Education Undergraduate Level Teaching Quality and Teaching Reform Project of the Ministry of Finance (MoF).

However, even though computer-assisted teaching has been widely used by college English teachers, on closer examination, the services provided by each university are not the same: only teachers at University A had access to the internet in the classroom. Besides, according to the participants, University A was the only university

among the three where all student dorms had access to the internet, as well as to English TV channels, such as CCTV-9 (China Central China Television Channel 9) and The Discovery Channel (programming in English). Moreover, computers were mostly used to show pre-created PowerPoint slides and movie clips, and to play the complementary CDs that accompanied the textbook. The fact that most teachers were still required to teach from the textbook was another reason that teachers had limited time to take advantage of rich online resources. It would be wise to give teachers more autonomy in their classrooms; otherwise, computers will continue to simply replace blackboards and DVD players, and be used for entertainment purposes. The risk is that teachers will pass up the opportunity to use of the virtually unlimited potential of computers for teaching English in their classes.

Different Levels of English Listening Practice at University A

Even though today, English teachers in the three universities put greater emphasis on listening practice, there are different levels of practice among the three universities sampled. For instance, at University A, besides explaining old CET-4 listening questions to the students, one of A. Jing's teachers also gave students listening comprehension and compound dictation tests adapted from Voice of America (VOA) broadcasts every other week. So did A. Qin's English teacher. Among the interviewees from University A in Phase III (p. 100), four students (A. Yan, A. Huang, A. Tang, and A. Liang) pointed out that the listening materials chosen by the teacher were always from either the British Broadcasting Corporation (BBC), Voice of America (VOA), Cable News Network (CNN) or Test of English as a Foreign Language (TOEFL). These exercises took the form of short tests of conversation comprehension, passage comprehension, and

compound dictation, which were all in the formats adopted in the reformed CET-4. A. Tang was very happy about these authentic listening materials. She complained that the complementary listening practice in her New Horizon textbook was not only impractical, but slow. She said, “even senior high school students know how to do those listening comprehension exercises in the textbook.”

So the listening materials from VOA, BBC, CNN, or TOEFL were quite popular at University A. In contrast, they were not commonly used at Universities B and C. Rather, in the latter two universities, listening materials were mostly taken from the CET-4 tests from previous years, or they were simulated CET-4 tests. It should be noted that listening materials from VOA, BBC, CNN, or TOEFL were more difficult than those from the reformed CET-4. In China, listening materials from, or adapted from, the VOA, including its special English programs, were usually used at a level of beginner English majors, not non-English majors.

So, even though the reformed CET-4 exerted positive washback on English listening at the college level, the depth was not the same among students from all the universities. Students from lower-ranking university had a great deal of CET-4 type listening exercises, while students at the higher-ranking university had many listening exercises above the CET-4 level.

No Substantial Two-way Interaction

According to the interviews, students stated that there were more opportunities to interact in class, but this did not appear to amount to substantial two-way interaction. B. M. Xu, A. B. Zhang, and B. Ding stressed that few students actively participated in class activities. The teacher still lectured most of the time, and dynamic interaction in class was

minimal, if there was any at all. As pointed out in Chapter 5, class participation rate at University B was significantly the lowest among the three universities. The experience of B. W. Wang at University B turned out to be a common phenomenon shared by interviewees from that university. B. W. Wang took an Oral English course with the hope of improving his skills. However, he said that most of the time nobody volunteered to answer the teacher's questions in class, be they difficult or easy. Students were too shy to speak English in front of others. If anybody did speak, he/she was considered weird. Sometimes Wang really wanted to break the ice, but he did not want the other students to think he was trying to show off. He always felt the tension of silence when a question was asked. Nevertheless, every student just lowered his/her head, expecting the teacher to give the answer or call on someone quickly to end the awkward silence. A. Qin from University A and C. Liu from University C also expressed concerns about the tension of silence and lack of participation in College English courses. They said sometimes they really felt sorry about the students' non-responsiveness, especially when teachers made an effort to engage them.

At other times the problem lay more with the teacher. For instance, in a CET-4 Ability Training class, B. X. Chen said that the teacher was highly confident about her teaching strategy. She told the students that "if you follow my instructions, and if your English foundations are not too bad, I guarantee that you will pass the reformed CET-4". She had faith in the teacher's confidence. However, she said that the class atmosphere was stifling to her. She sighed at the end of the interview and said, "I do not like classes that are taught to the test, but I guess I have no choice at this point."

B. Shi, another student, pointed out that there were more than 70 students in his

CET-4 Ability Training class. In every class the students gave ten one-minute presentations with a brief question and answer session following each presentation. According to B. Shi, this shortened the time available for actual teaching in a class that met only once a week. Everyone was worried about getting his/her presentation done and few would listen attentively to the others' presentations after finishing their own. He and some of his classmates thought their English did not improve much in this class because of the large class size.

B. J. Li, conversely, mentioned that his teacher lectured most of the time in his CET-4 Ability Training course. The class was quite noisy because students constantly held small conversations on the side. He said he would not have signed up for the class if his roommates had not done so. A. Jin claimed that he had a senior-high-like teacher who stressed grammar from the textbook, and there was not much interaction in class. What is more, among the five students who were surveyed both in Phase II and Phase IV, three (B. M. Xu, A. Z. Li, and C. Liu) pointed out that compared to the teacher they had had during the previous semester, who solo lectured only *sometimes*, the current English teacher lectured *often*. One student rated the change from *never* to *seldom* and another gave the same rating for both semesters: *sometimes*.

Therefore, factors like class size, frequency of class meetings, characteristics of Chinese students, and peer pressure, all lead to a rather low level of participation in class. Even the most commonly adopted class activity, the oral presentation, is in fact, largely a one-way interaction, so it does not provide substantial engagement between students in class.

Furthermore, the interview data indicate that students have many misperceptions

about the CET- SET (Spoken English Test). Some students had never heard of it and did not know how it was conducted. Although at this time, the CET-SET is separate from the College English Test, it is still an essential component of the college English test. Speaking fluently is also, after all, a demonstration of one's language proficiency. Publicizing the CET-SET and making it accessible to college students may be a way to foster development of substantial interaction in college classes, and to strengthen the positive washback effects of the reformed CET-4.

Textbook - Centered Pedagogy Unchanged

At University A, more interaction takes place in class than in senior or junior high school, but most interviewees still felt that their English class largely revolved around textbooks. A. Zhu said that even though there were interactive activities planned in class, the class was primarily focused on textbooks. Students in his class had to submit notes taken in class as part of the final grade. He stated that he benefitted most from the new vocabulary reviewed in class. A. Z. Li said the same thing, but she also liked her teacher's explanations of how the text was organized and usage of figures of speech in each unit. One of the teachers that A. Jing had was also very similar to a senior-high teacher. She recalled that the teacher spent most of the class time going over the textbook, page by page. The teacher very carefully explained new words or phrases in each unit. For the rest of the class, she helped students practice for the reformed CET-4 by listening to simulated listening exercises, practicing reading, and introducing test-taking strategies, always based on the CET-4.

At University B, there seemed to be a more uniform curriculum and syllabus for College English teachers than at University A, since most teachers only taught Section A

of each unit in depth. B. X. Chen described how, in her College English III course, the teacher usually spent most of the class time analyzing the structure of the first text in Section A. Paragraph by paragraph, he/she explained the new words and grammar points. The teacher encouraged students to explore the theme of the text in addition to grasping its literal meaning. After explaining Text One, the teacher continued explaining the exercises that followed. Most of the time, he said, his teacher was the focus of the class, acting as the knowledge transmitter for the students. The homework assignments he gave were usually translation and banked cloze exercises from the textbook.

C. S. Chen from University C said his teacher's instruction technique was mainly textbook-based. His teacher underlined important grammar points, explained new words, and asked students to translate complex sentences. Xu's teacher taught in the same way, but with less emphasis on grammar. This was even truer in the first semester classes at University C. C. Liu, for instance, said that the teacher spent much time introducing the new topic. After dividing the text into paragraphs, the teacher explained important new words, phrases, and grammar, and translated difficult sentences. Unit by unit, this procedure was then repeated. C. Liu did not think it was reasonable to focus so much on the textbook. He expected that, in college, the teacher would venture beyond the textbook more often.

In short, the reformed CET-4 had very limited impact on English teachers' instructional styles in college. Teachers still mostly follow the traditional textbook-centered pedagogy and students are not particularly engaged in the classroom. Ironically, after content (60.6%), teaching methods/strategies were the aspect that students perceived to be most influenced by the CET-4 (42.8%). Further discussion suggested that textbook-

centered teaching pedagogy remains largely unchanged. What has changed compared to earlier practice, is that more teachers use computer support to teach from the textbook. The question that emerges is whether participants equate the adoption of computer-assisted teaching to a change in teaching pedagogy.

Writing and Grading Assistance Should be Provided

As mentioned earlier, almost the same percentage of students perceived writing to be the third most influenced aspect of the CET-4, whether it involves learning to write English, or teaching English writing. (The percentages for learning writing are 36.2% and for teaching it, 35.7% of the 414 subjects respectively) However, the interview data further indicate that students do not receive sufficient English writing practice but are simply being taught writing techniques. For instance, B. J. Li mentioned that in one of the sections of the CET-4 Ability Training course, the teacher was famous for her strategy for acing the CET-4 writing component. An impressive first sentence, she claimed would earn a possible 9 points out of 15. In fact, most of the time, English teachers skipped the writing section entirely when practicing mock CET-4 tests.

B. M. Xu, A. Tang, and A. Z. Li's teachers all had their students practice questions from the CET-4 model tests in class, and afterwards discussed strategies. This held true for every section of the test except the writing component. Even if sometimes teachers gave the students time to write in class (30 minutes), they usually would not grade the essays. B. J. Li emphasized that with 70 -80 students in his CET-4 Ability Training course, frequent writing assignments would dramatically increase the teaching load. B. W. Wang said he never had writing assignment in College English III course.

On the other hand, students found that it was hard to practice essay writing on

their own after class. B. W. Wang practiced by himself after class, taking approximately 10 past CET-4 tests and using a CET-4 preparation book from the library. However, because he had no one to grade his writing, he didn't practice the writing sections of those simulated tests. According to the interviewees, a popular way for students to prepare for the writing section was to memorize model phrases, sentences, essays or formulaic writing templates, rather than actually writing an essay and then revising it based on comments of a teacher or a peer. It comes as no surprise that, from the scores collected in Phase IV, writing was the weakest skill.

The lesson derived from this state of affairs is that writing and grading assistance would be beneficial for both English teachers and students. Funding permitting, teaching assistants could be assigned to teachers to help them grade essays. Teachers, English graduate students, and senior students who are proficient in English could be employed in a writing center to give students extra help with their English writing skills. Student activity fees could be collected to cover the service if needed.

Helping Students Set Clear Goals

Because the survey results indicated that students who did not have graduation plans spent significantly less time studying English every day compared to students who planned either to study abroad, or to attend graduate school or to work after graduation, students at the college level should be encouraged as early as possible to think about their future plans, and to set short-term and long-term goals if possible. Proper guidance should be provided if students are unable to make a decision.

The survey results also indicated that, compared students with other goals, students spent significantly less time on English when their purpose for learning the

language was to pass required courses and to get high grades. Therefore, colleges can do a better job of demonstrating how students can benefit in their major study areas by learning English for communicative purposes.

In addition, teachers who have a sense of humor also seem to help foster students' interest in learning English. B. Ding, B. Dong, and C. J. Li, for example, pointed out that they liked studying English because they liked their English teacher's sense of humor in their middle school. To lighten up the classroom, English teachers could be encouraged to tell an occasional joke. This would help lower students' anxiety and stimulate their interest in the subject.

Students' Learning Methods Remained Largely Unchanged

One of the two aspects most influenced by the reformed CET-4 were teaching methods (42.8%). However, 70.8% of the 414 reported that the reformed CET-4 did not have a strong influence on their learning methods/strategies. According to C. Ye, his methods remained almost the same as in senior high school, because the focus was still on learning new vocabulary, reading, listening, and practicing model tests. B. X. Chen said her learning method was an extension of the one she adopted while preparing for the entrance exam to college: she memorized new words of high frequency, practiced model tests, and sometimes read some interesting simplified English articles if she had time. She said she wanted to improve her English application ability, but she did not know how to do it.

A. Luo from University A stated that he still largely followed the teacher's instructions in class and he did not spend much extra time on studying English after class. So he did not think his methods had changed either. The conclusion to be drawn from

these findings is that a test will influence what learners learn (Hypothesis 5, Alderson & Wall, 1993) but it will not necessarily influence how learners learn (Hypothesis 6, Alderson & Wall, 1993).

Different Access to Resources Prior to College Education

Among the 29 interviewees, students from Eastern China obviously had a head start in terms of English learning: 16 Eastern China students received formal, systematic English education since primary school, if kindergarten English education is not counted. In Shanghai, the earliest grade for a student to begin learning English classes was Grade One. In contrast, 10 students did not start learning English until middle school. Among them, three were from North Western China (Gansu and Ningxia), two were from the Southwest (Yunnan and Guizhou), and five were from Eastern China. However, all these five students came from either the suburbs or from rural Eastern China.

Students from Eastern China also had more private tutoring than students from other parts of the country: Seven Eastern students had private English tutoring in primary school and/or middle school. The rest of the students were from either Central China (2), Northern China (2), Southern China (1) or Southwestern China (1). Some were given extra materials and practice while for others the tutoring was remedial in nature, according to the 14 students who received private tutoring. It is quite common for parents to send their children to private English training schools catering exclusively to children. In fact, more than six of the 14 students attended English training schools. This practice corresponds to the assumption that young learners can learn English quickly. Compared to formal school education, most students recalled that they had a great deal more fun singing and playing games in private English training school than in later classes.

Once a week while in high school, 14 students had English conversation classes taught by native English speakers. Among them, two students had native English teachers in middle school and eight in high school (3-Shanghai, 1-Hainan, 1-Guizhou, 1-Ningxia, and 1-Guizhou). All the students from Shanghai had conversation classes taught by native English teachers for three years in high school. Two students from the Northwest and one student from the Southwest had such classes only in the first year of high school. The student from Hainan, Southern China had the classes in Grade One and Grade Two. Four students (2-Eastern China, 1-Northern China, 1-Southern China) had native English teachers in both middle schools and high schools. Overall, the majority, seven students from Eastern China, had native English teachers either in middle school and/or high school.

In contrast, eight students had never taken any English classes taught by native English speakers: three were from the Northwest (Gansu and Ningxia), one from the Southwest (Yunnan), and four from the suburbs or rural areas of Eastern China (Fujian, Anhui, Shandong and Shanghai). A. Z. Zhang, a student from downtown Shanghai said that his English courses were all taught *in English* at high school. Other students, however, had already studied or travelled in English-speaking countries. For instance, A. B. Zhang from Shanghai travelled abroad during vacations in primary school, middle school and high school. A. Liang from Shanghai was a summer exchange student in America during his freshman year. A. Yan from Hunan in Central China had frequent contact with foreign experts who worked in the company where his parents were employed. A. Tang from Sichuan, Southwestern China, had studied for a month in the United States as an exchange student while she was a middle-school student. Because

they had studied abroad and/or had face-to-face contact with native English speakers, A. Tang, A. Liang, and A. Yan were all motivated to pursue their graduate study in the United States. In fact, they were all preparing for the GRE at the time of the interview.

In addition, each city or province in China has different policies regarding English testing, a fact which could make a difference in students' English levels. A. Jing and A. Qin both from Eastern China said that they were tested as part of their graduation from Grade Six in their primary schools. By the time that students had taken entrance exams for high school, three students from Eastern China (1-Shanghai, 2-Jiangsu) had taken English test that included both listening and speaking components, though their oral skills were tested separately and used only as a reference for admission. C. Y. Zhang from Nantong, Jiangsu province, Eastern China, said that her oral English test score counted 10-20 points in the final English score. A. Z. Li from Henan, Central China, had been tested on listening, too, in her middle school exit test. In the entrance exam to college, six students from Eastern China (3-Shanghai, 1-Fujian, 1-Jiangsu and 2-Anhui) and one from Guangdong, Southern China, had both oral English and listening components in their English tests.

According to the seven students who reported that they had taken an oral English test as a component of their college entrance exams, the oral test was administered before the entrance exams. It was not added to the total English score, but graded as Excellent, Good and Bad by two to three English teachers from other schools. The students were given a pamphlet before the test that contained all the possible passages that would be tested in the oral test. During the test, each student was randomly assigned a passage to read and then answered questions based on that passage. All the students said that the oral

test was easy and no one whom they knew had failed it. On the other hand, three students from the Northwest (Gansu) and one student from Northern China (Shanxi) reported that the listening component was not even included in their English test in the college entrance exam.

What is more, the English learning resources and learning environment varied from province to province. Students from administrative districts other than Shanghai, in particular, described the enormous pressure they faced in their college English classes. Students from Shanghai, on the other hand, had a stronger English learning environment before college. For instance, A. Z. Zhang, from a key high school in Shanghai said that, all his English courses were already taught in English in high school. He felt that he could have passed the CET-4 in high school. Many of his high school classmates were actually preparing for, or had already passed the Intermediate Level Oral English Test, and/or the Scholastic Assessment Test (SAT). A. Liang from Shanghai mentioned that the reading exercises he did in high school were from the CET-4. C. J. Li from Shanghai also said that her high school English teacher played the CET-4 listening exercises and practiced CET-4-type writing. A. Zhu from Shanghai said the students who obtained excellent scores on the English test on the college entrance exam practiced taking the CET-4 rather than doing high school English exercises. He also did some CET-4 practice in high school and he took a TOEFL class in the New Oriental English Training School in high school as well. A. Qin, another student from Shanghai also mentioned that he and many of his high school friends had registered for summer preparation classes at the New Oriental English Training School after graduating from high school. These classes included an English Interpretation class, an Oral English Intermediate or Advanced Level

class, the Test of English as a Foreign Language (TOEFL) or the International English Language Testing System (IELTS) preparation class. Some even took and passed some of these tests in their second year of high school. They also took English classes taught in multi-media classrooms in high school. C. Ye from Fujian, Eastern China and C. S. Chen from Southern China also stated that their high school English classes were taught in multi-media classrooms.

In contrast, A. Y. Zhao from Gansu, Northwestern China recalled that his middle school was not a key school in his county. His teachers were mostly graduates of a vocational normal school. English classes were taught in his local Chinese dialect. His high school was located next to his middle school, and whenever there was a shortage of high school teachers, the middle school teachers were transferred to the high school. In college, at first, he was at a complete loss since he could not understand the listening exercises, nor could he speak English. He did not register for the CET-4 in December as scheduled, because he was not ready for it and he did not want to waste his money. In June when his classmates were preparing for the CET-6, however, he still felt a lot of pressure just to obtain a score of 425, the minimum for students who wanted to take the CET-6. In the end, he scored a 407 with 130 in listening, 147 in reading, 86 in writing, and 44 in integrated. This was the lowest score given on the test among all the students who reported their CET-4 scores for the study. The third lowest score was obtained by C. Liu, a student from another Northwest District - Ningxia Autonomous Region. His overall score was 415, with 105 in listening, 180 in reading, 55 in integrated, and 72 in writing.

B. M. Xu from Gansu, Northwestern China from University B also stated that her

teacher in high school spent no time at all on listening because listening was not going to be tested in the matriculation test. She worked very hard and got a score of 475, with 146 in listening, 174 in reading, 54 in integrated, and 100 in writing. A. Meng, another student from Gansu shared her frustration. He said that he did not understand a single sentence spoken in his first English class in college. For him, the teacher spoke too much English and she spoke very fast, whereas the English teachers he had before college mostly spoke Chinese. However, in his first year of high school in Gansu province, he did have a conversation class taught by a native English speaker from Oxford University, but he felt that it did not help him much in college. He pointed out that his reading skills were good, and his English score on the entrance exam to college was quite good, yet he immediately felt a large gap between him and his classmates at college, in his ability to apply his English. He said that he only got eight out of thirty points in the listening component of the final English test in his freshman year. He blamed the gap for the uneven development between the inner cities and coastal cities in China. However, he said he would not be willing to go back to his underdeveloped hometown to work if he could find a job in Shanghai. Later his score was 484 with 171 in listening, 189 in reading, 52 in integrated, and 72 in writing.

Students from the rural areas or the suburbs of other Eastern Chinese provinces such as Anhui and Zhejiang also had problems catching up with students from the city of Shanghai. A. Xie from the Zhejiang suburbs and C. J. Xu from a rural area in Anhui province, B. X. Chen from a village in Fujian province all said that they could not understand their college English teacher, and the listening levels expected in college were too hard. C. J. Xu stated that her college English teacher's pronunciation was different

from the English teachers she had back home before college. A. Huang from the city of Fuzhou did not have much difficulty in English in college, yet he expressed his admiration toward students from Shanghai, because on the whole, they were better at both speaking and listening.

A. Yan from Hunan province, Central China said his English was good in Hunan, but he saw the gap between him and his peers from Shanghai when he got to college. B. Song from Yunnan province, Southwestern China sensed the gap, too. He believed that big cities like Shanghai, Beijing and Guangzhou had more advantages in developing their education. He believed that students from prosperous big cities studied English for application in their future work, whereas he only studied English to pass the English tests. He said his parents were proud of him being admitted to college. He was also excited about going to college in Shanghai, but he was surprised to learn that the goals for these students were to study abroad. He also mentioned that his English teacher demotivated him in senior high school and he had lost interest in learning English ever since.

A. Z. Li from Henan province, Central China in University A registered for an intermediate level English Interpretation class in the New Oriental English Training School on weekends. She soon noticed that most of her classmates were high school students in Shanghai and she immediately felt out of place competing against them as a college sophomore.

In conclusion, the variety of competences resulting in different levels of English knowledge in college could be attributed to many factors: the availability of educational resources, differences in educational policies in different cities/provinces, the uneven economic development between the inner cities and coastal cities and between Southern

and Northern China, and urban and rural areas. According to this study, these differences were likely to have a bearing on student skills rather than students' innate English learning abilities.

Prior Positive and Negative English Learning Experience

On the other hand, for young learners of English as a Foreign Language (EFL), or learners who start having contact with English at an early age, the learning process is a positive experience which reinforces their interest in English. In other words, students appear to have more and more interest in learning English when they continue to have positive learning experiences, whether or not they like learning English or not. They simply keep studying English when they realize they are good at it. Among interview participants C. Y. Zhang and A. Jing, for instance, the negative and sometimes humiliating experience they had in middle school made them shy away from studying English later. According to C. Y. Zhang, she was always one of the few female students who were forced to stay after school to copy new vocabulary because of her poor dictation grades. A. Jing was often penalized in middle school by being forced to remain standing in class when she could not answer her English teacher's questions. A. Meng from University A simply said that he had no interest in learning English because he was not good at it.

In contrast, A. Jin won second prize in the First Shanghai Children's English Speech Contest as a third grader. A. B. Zhang was the second prize winner in his district English Speech Contest as a first grader. B. J. Li and B. W. Wang, another two students who were still passionate about studying English, were both designated as representatives for English in middle school. In some ways, four of these students declared that finding

that they were good at English and getting public acknowledgement before senior high school highly motivated them to continue studying English even when it became boring and test-oriented.

What they describe here is quite similar to Serena Williams' relationship to tennis. Serena had five Wimbledon titles and 17 Grand Slam singles victories by September 2013. Famous thought she may be as a top tennis player in the world, she confessed in 2012 that she never had a genuine interest in tennis. She told a reporter that she did not like tennis, and "It's not that I've fallen out of love with it. I've actually never liked sports and I never understood how I became an athlete" (Gregory, 2012, para.1). She also said, "I can't live without it – there's a difference between not loving something and not being able to live without it" (Gregory, para. 1). Obviously she did not quit or plan to quit any time soon. She knows that she is good at this sport and that she will succeed in doing it.

Like her, college students who have had a head start with learning English, are likely to continue to practice and will not quit even if they do not really like it. If researchers want to examine washback effects holistically to include these kinds of factors, they will need to consider the impact of extrinsic factors like socioeconomic background, school and education, family, friends and colleagues. In addition to focusing on intrinsic factors like test taker characteristics and test factors, the effects of experiences such as access to resources prior to college education and prior English learning experience, support the need to look into extrinsic factors (Shih, 2007).

To conclude, some positive washback effects have been achieved as the result of the reform of the CET-4, because of the increased weight of listening, and the addition of

fast reading, for instance, in the reformed test. This has led to an incremental emphasis on those practices by both teachers and students. What is more, the lower the ranking of the university/college, the stronger these positive washback effects are, according to this study. Given that students recruited by universities of the first tier in the nation generally have a better foundation in English, a more challenging version of the CET-4 could be offered to them. Of course, students in general need to be informed about the up-to-date information of the reformed CET-4 before they take the test.

Also, despite the abolition of CET-4 certificates, the reformed CET-4 has changed *what* teachers teach and *what* students learn in and out of English classes. However, it did not substantially change *how* teachers teach or *how* students learn. The study suggests that the requirement of using assigned textbooks by the MoE in almost all universities seems to be the major barrier: Teachers have little teaching autonomy as a result of existing policies on the use of textbooks. Nor could students be given adequate opportunity to use English communicatively. Of course, proper training should be provided for teachers to make a smooth transition if textbooks were to be mainly used as references. Students might benefit more from the reform, too, if their workload were reduced and more guidance were given on becoming autonomous learners.

Moreover, the study shows that no substantial two-way interaction took place in English classes and a number of students did not even know the existence of the Spoken English Test (SET). The fact that the oral English test is a separate proficiency test from the CET-4 and only those who obtain certain minimum scores are eligible to advance, suggests that this policy must be changed. On the other hand, as was expected by the CET-4 committee, most students want to develop their oral English ability in college, so

the oral English test should be included as an integral part of the CET-4. It is said that the computer-based CET-4 that is currently in trials is a test of listening, speaking, reading and writing. It is our hope that it will be adopted soon for all college students. In the meantime, mandatory oral English courses of different levels should be offered in every university. If possible, native English teachers could co-teach with non-native English teachers in these classes.

Given the uneven development of English fundamentals acquired prior to college, the variety of motivations for studying English in college, and different English proficiency levels demanded by different majors, it would be wise to offer more diverse English courses other than the College English that is currently offered. University A which uses an array of courses such as English Public Speaking, Translation, Movie Appreciation and etc. could be used as a model for other universities/colleges in order to cater to students' interests and make their learning more meaningful. Resources permitting, major courses should be offered in English as well. If not, workshops or presentations delivered in students' majors should be provided by guest speakers in English each semester. After all, testing is just a tool to stimulate students to learn. At no time should it be the sole reason why students learn.

In addition, none of these positive washback effects would take place easily if support services such as self-access English learning labs, writing centers, social activities in English, improved English education prior to college, and intensive teacher training programs, etc. were not provided. Validating a test and trying to improve language education through a test's positive washback is a long-term task. It cannot be executed successfully without a collective effort of all stake holders.

References

- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13(3), 280 – 297.
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115 – 129.
- Alderson, J. C. (2004). Foreword. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. ix – xii). Mahwah, NJ: Lawrence Erlbaum.
- American Psychological Association (1985). *Standards for educational and psychological testing*. Washington, DC: Author.
- Andrews, S. (1994). The washback effect of examinations: Its impact upon curriculum innovation in English language teaching. *Curriculum Forum*, 4(1), 44 – 58.
- Andrews S., Fullilove J., & Wong Y. (2002). Targeting washback: A case-study. *System*, 30(2), 207 – 223.
- Bachman, F. L. (1990). *Fundamental considerations in language testing*. Oxford, UK: Oxford University Press.
- Bachman F. L. (2007). What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. Turner, & C. Doe (Eds.), *Language testing reconsidered* (pp. 41 – 72). Ottawa, Canada: University of Ottawa Press.
- Bachman, F. L., & Palmer, A. S. (1996). *Language testing in practice*. Oxford, UK: Oxford University Press.

- Bailey, K. M. (1990). The use of diary studies in teacher education programs. In J. C. Richards, & D. Nunan (Eds.), *Second language teacher education* (pp. 215 – 226). Cambridge, UK: Cambridge University Press.
- Bailey, K. M. (1996). Working for washback: A review of the washback concept. *Language Testing*, 13(3), 257 – 279.
- Biggs, J. B. (1995). Assumptions underlying new approaches to educational assessment. *Curriculum Forum*, 4(2), 1 – 22.
- Biggs, J. B. (Ed.) (1996). *Testing: To educate or to select? Education in Hong Kong at the cross-roads*. Hong Kong: Hong Kong Educational.
- Bourdieu, P., & Thompson, J. B. (1991). *Language and symbolic power*. Cambridge, MA: Harvard University Press.
- Bracey, G. W. (1987). Measurement-driven instruction: Catchy phrase, dangerous practice. *Phi Delta Kappa*, 68, 683 – 686.
- Brannen, J. (1992). Combining qualitative and quantitative approaches: An overview. In J. Brannen (Ed.), *Mixing methods: Qualitative and quantitative research* (pp. 3 – 37). Burlington, VT: Ashgate.
- Bray, M., & Steward, L. (1998). *Examination systems in small states: Comparative perspectives on policies, models and operations*. London, UK: Commonwealth Secretariat.
- Buck, G. (1988). Testing listening comprehension in Japanese university entrance examinations. *Japan Association for Language Teaching Journal*, 10(1), 12 – 42.
- Burrows, C. (1998). *Searching for washback: An investigation of the impact on teachers of the implementation into the adult migrant English program of the assessment of*

- the certificates in spoken and written English*. Unpublished doctoral dissertation, Macquarie University, Sydney.
- Cai, J. (2006). *ELT at tertiary level in China: Review, reflection and research*. Fudan, Shanghai: Fudan University Press.
- Canale, M. (1983). On some dimensions of language proficiency. In J.W. Oller, Jr. (Ed.), *Issues in language testing research* (pp. 333 – 342). Rowley, MA: Newbury House.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approach to second language teaching and testing. *Applied Linguistics*, 1(1), 1 – 47.
- Carroll, J. B. (1968). The psychology of language testing. In A. Davies (Ed.), *Language testing symposium: A psycholinguistic approach* (pp. 46 – 69). London, UK: Oxford University Press.
- Celce-Murcia, M. (2001). *Teaching English as a second or foreign language*. (3rd ed.). Boston, MA: Heinle & Heinle.
- Chapman D. W., & Snyder, C. W. Jr. (2000). Can high stakes national testing improve instruction? Reexamining conventional wisdom. *International Journal of Educational Development*, 20(6), 457 – 474.
- Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. Cambridge, UK: Cambridge University Press.
- Cheng, L. (2008). The key to success: English language testing in China. *Language Testing*, 25(1), 15 – 37.

Cheng, L. (2009). The history of examinations: Why, how, what and whom to select? In

L. Cheng, & A. Curtis (Eds.), *English language assessment and the Chinese learner* (pp.13 – 25). New York, NY: Routledge.

Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 19 – 36).

Mahwah, NJ: Lawrence Erlbaum.

Cheng, L., & Curtis, A. (2009). The realities of English language assessment and the Chinese learner in China and beyond. In L. Cheng, & A. Curtis (Eds.), *English Language assessment and the Chinese learner* (pp. 3 – 12). New York, NY: Routledge.

全国约有 3 亿多人在学英语. [China has approximately 30 million English learners].

(2006, April 18). Retrieved from <http://opinion.people.com.cn/GB/8213/62024/430907.html>

中国大学排行榜： 上海. [China 2010 National University Rankings: Shanghai]. (2010).

Retrieved from http://rank2010.netbig.com/citypaihang_9.html

Corbett, H. D., & Wilson, B. L. (1991). *Testing, reform and rebellion*. Norwood, NJ: Ablex.

Creswell, J.W. (2008). *Educational research: Planning, conducting, and evaluating qualitative and quantitative research* (3rd ed.). Upper Saddle River, NJ: Pearson.

Creswell, J. W., Plano Clark, V. L., Gutmann, M., & Hanson, W. (2003). Advanced Mixed Methods Research Designs. In A. Tashakkori, & C. Teddlie (Eds.),

- Handbook of mixed methods in social and behavioral research* (pp. 619 – 637).
Thousand Oaks, CA: Sage.
- Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.
- Crystal, D. (2003). *English as a global language*. (2nd ed.). Cambridge, UK: Cambridge University Press.
- Davies, A. (1977). The construction of language tests. In J. P. B. Allen, & A. Davies (Eds.), *Testing and experimental methods* (pp. 38 – 104). London, UK: Oxford University Press.
- Davies, A. (1990). *Principles of language testing*. Oxford, UK: Basil Blackwell.
- Davies, A. (2007). Assessing academic English language proficiency: 40 years of UK language tests. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. Turner, & C. Doe (Eds.), *Language testing reconsidered* (pp. 73 – 86). Ottawa, Canada: University of Ottawa Press.
- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- Eckstein, M. A., & Noah, H. J. (Eds.). (1992). *Examinations: Comparative and international studies*. New York, NY: Pergamon Press.
- English, F.W. (1992). *Deciding what to teach and test: Developing, aligning, and auditing the curriculum*. Newbury Park, CA: Corwin Press.
- Foster, P. J. (1992). Commentary. In M. A. Eckstein & H. J. Noah (Eds.), *Examinations: Comparative and international studies* (pp. 121 – 126). Oxford, UK: Pergamon Press.

- Frederickson, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychology*, 39(3), 193 – 202.
- Fullan, M. (2007). *The new meaning of educational change* (4th ed.). New York, NY: Teachers College Press.
- Gao, Y. (2003). Empirical study and open-door policy are needed for foreign language teaching reform. *Foreign Language Teaching and Research*, 99(3), 222 – 223.
- Gates, S. (1995). Exploiting washback from standardized tests. In J. D. Brown, & S. O. Yamashita (Eds.), *Language testing in Japan* (pp. 101 – 106). Tokyo, Japan: Japanese Association for Language Teaching.
- Green, A. B. (2007). *IELTS washback in context: Preparation for academic writing in higher education*. Cambridge, UK: Cambridge University Press.
- Gregory, S. (2012, January 4). Why doesn't Serena Williamson love tennis? Retrieved from <http://keepingscore.blogs.time.com/2012/01/04/why-doesnt-serena-williams-love-tennis/>
- Ha, P. L. (2008). *Teaching English as an international language: Identity, resistance and negotiation*. Clevedon, UK: Multilingual Matters.
- Hamp-Lyons, L. (1997). Washback, impact and validity: Ethical concerns. *Language Testing*, 14(3), 295 – 303.
- Herman, J. L., & Golan, S. (1991). *Effects of standardized testing on teachers and learning: Another look* (CSE Technical Report No. 334). Los Angeles, CA: University of California, CRESST.
- Hu, C.T. (1984). The historical background: examinations and control in pre-modern China. *Comparative Education*, 20(1), 7 – 26.

- Hughes, A. (1989). *Testing for language teachers*. Cambridge, UK: Cambridge University Press.
- Hughes, A. (1993). *Backwash and TOEFL 2000*. Unpublished manuscript, University of Reading, Reading, Berkshire, United Kingdom.
- Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Huhta, A., Kalaja, P., & Pitkanen-Huhta, A. (2006). Discursive construction of a high-stakes test: The many faces of a test-taker. *Language Testing*, 23(3), 326 – 350.
- Jin, Y. (2006). Improving validity and washback: A washback study on the CET-4 and CET-6. *Foreign Language World*, (6), 65 – 73.
- Jin, Y. (2010). The national College English Testing Committee. In L. Cheng, & A. Curtis (Eds.), *English language assessment and the Chinese learner* (pp. 44 – 59). New York, NY: Routledge.
- Jin, Y., & Yang, H. (2006). The English proficiency of college and university students in China: As reflected in the CET. *Language, Culture and Curriculum*, 19(1), 21 – 36.
- Lado, R. (1961). *Language testing: The construction and use of foreign language tests*. London, UK: Longman.
- Lam, A. (2002). English in education in China: Policy changes and learners' experiences. *World Englishes*, 21(2), 245 – 256.
- Larsen-Freeman, D. (2000). *Techniques and principles in language teaching: Teaching techniques in English as a second language*. (2nd ed.). Oxford, UK: Oxford University Press.

- Lee, J. F., & VanPatten, B. (1995). *Making communicative language teaching happen*. New York, NY: McGraw-Hill.
- Li, Q., & Wang, X. (2003). Inspiration of basic English education reform for college English. *Foreign Language Teaching and Research*, (5), 382 – 383.
- Littlewood, W. (1981). *Communicative language teaching: An introduction*. Cambridge, UK: Cambridge University Press.
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and criteria. *Educational Researcher*, 20(8), 15 – 21.
- Luxia, Q. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing*, 22(2), 142 – 173.
- Madaus, G. F. (1988). The influence of testing on the curriculum. In L. N. Tanner (Ed.), *Critical issues in curriculum: Eighty-seventh yearbook of the national society for the study of education* (pp. 83 – 121). Chicago, IL: University of Chicago Press.
- Manjarrés, N. B. (2005). Washback of the foreign language test of the state examinations in Colombia: A case study. *Arizona Working Papers in SLAT*, 12, 1 – 19.
- McNamara, T., & Roever, C. (2006). *Language Testing: The social dimension*. Oxford, UK: Blackwell.
- McKay, S. L. (2002). *Teaching English as an international language: Rethinking goals and approaches*. Oxford, UK: Oxford University Press.
- Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer, & H. I. Braun (Eds.), *Test validity* (pp. 33 – 45). Hillsdale, NJ: Lawrence Erlbaum.

- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241 – 256.
- Morrow, K. (1986). The evaluation of tests of communicative performance. In M. Portal (Ed.), *Innovations in language testing: Proceedings of the IUS/NFER Conference* (pp. 1 – 13). London, UK: NFER/Nelson.
- National College English Testing Committee (1985). *Syllabus for College English Test*. [Brochure]. Shanghai, China: Shanghai Language Education Press.
- National College English Testing Committee (1999). *Syllabus for College English Test*. [Brochure]. Shanghai, China: Shanghai Language Education Press.
- National College English Testing Committee (2006). *Syllabus for College English Test*. [Brochure]. Shanghai, China: Shanghai Language Education Press.
- National College English Testing Committee (2006). *College English Test - Spoken English Test (CET-SET) syllabus and sample test*. [Brochure]. Shanghai, China: Shanghai Language Education Press.
- National College English Testing Committee (2006). *CET-4 syllabus and sample test paper (revised version)*. [Brochure]. Shanghai, China: Shanghai Language Education Press.
- Nobel, A. J., & Smith, M. L. (1994). *Measurement-driven reform: Research on policy, practice, repercussion*, (CES Technical Report No. 381). Tempe, Arizona: Arizona State University, CSE.
- Oneill, P., Murphy, S., Huot, B., & Williamson, M. (2006). What high school teachers in three states say about different kinds of mandated state writing tests. *Journal of Writing Assessment*, 2(2), 81 – 108.

- Onwuegbuzie, A. J., & Teddlie, C. (2003). A framework for analyzing data in mixed methods research. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 351 – 383). Thousand Oaks, CA: Sage.
- Pearson, I. (1988). Tests as levers for change. In D. Chamberlain, & R. J. Baumgardner (Eds.), *ESP in the classroom: Practice and evaluation* (pp. 98 – 107). London, UK: Modern English.
- Popham, W. J. (1983). Measurement as an instructional catalyst. In R. B. Ekstrom (Ed.), *Measurement, technology, and individuality in education: Proceedings of the 1982 ETS Invitational Conference: No.17. New directions for testing and measurement* (pp. 19 – 30). San Francisco, CA: Jossey-Bass.
- Popham, W. J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappa*, 68, 679 – 682.
- Popham, W. J. (1993). Measurement-driven instruction as a “quick-fix” reform strategy. *Measurement and Evaluation in Counseling and Development*, 26(1), 31 – 34.
- Potter, J. (2003). Discourse analysis and discursive psychology. In P. M. Camic, J. E. Rhodes, & L. Yardley, (Eds.), *Qualitative research in psychology: Expanding perspectives in methodology and design* (pp. 73 – 94). Washington, DC: American Psychological Association.
- Qi, L. (2004). Has a high-stakes test produced the intended changes? In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 171 – 190). Mahwah, NJ: Lawrence Erlbaum.

- Rea-Dickins, P. (1997). 'So, why do we need relationships with stakeholders in language testing? A view from the UK'. *Language Testing*, 14(3), 304 – 314.
- Richards, J. C., Platt, J., & Platt, H. (1992) *Longman dictionary of language teaching and applied linguistics*. Harlow, Essex: Longman .
- Ross, J. S. (2008). Language testing in Asia: Evolution, innovation, and policy challenges. *Language Testing*, 25(1), 5 – 13.
- Salehi, H., Mustapha, R., & Yunus, M. M. (n.d.). The perceptions and attitudes of Iranian high school English teachers, students, and parents towards the entrance examination of universities (EEOU): Implications on English learning and teaching. *Prosiding Seminar Penyelidikan Siswazah*, 75 – 84.
- Savignon, S. J. (1997). *Communicative competence: Theory and classroom practice*. New York, NY: McGraw-Hill.
- Saville, N., & Hawkey, R. (2004). The IELTS impact study: Investigating washback on teaching materials. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 73 – 96). Mahwah, NJ: Lawrence Erlbaum.
- Shepard, L. A. (1990). Inflated test score gains: Is the problem old norms or teaching the test? *Educational Measurement: Issues and Practice*, 9(3), 15 – 22.
- Shih, C. (2007). A new washback model of students' learning. *The Canadian Modern Language Review*, 64(1), 135 – 161.
- Shohamy, E. (1992). Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning. *Modern Language Journal*, 76(4), 513 – 521.

- Shohamy, E. (1993). The power of tests: The impact of language tests on teaching and learning. *NFLC Occasional Papers*. Washington, D.C.: National Foreign Language Center.
- Shohamy, E. (2001). *The power of tests: A critical perspective of the uses of language tests*. London, UK: Longman.
- Shohamy, E., Donitza-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13(3), 298 – 317.
- Smith, M. L. (1991). Meaning of test preparation. *American Educational Research Journal*, 28(3), 521 – 542.
- Song, X. M. (2005, February 23). 大学英语四级考试应否与学位挂钩. [Should CET-4 be tied to diploma?]. *Guangming Daily*. Retrieved from http://www.gmw.cn/01Gmrb/2005-02/23/content_184192.htm
- Spolsky, B. (1994). The examination of classroom backwash cycle: Some historical cases. In D. Nunan, V. Berry, & R. Berry (Eds.), *Bring about change in language education* (pp. 55 – 66). Hong Kong: University of Hong Kong, Department of Curricular Studies.
- Spratt, M. (2005). Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, 9(1), 5 – 29.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Grounded theory, procedures and techniques*. Newbury Park, CA: Sage.

- 调查显示英语四六级考试成部分学生的心病. [Survey shows that the CET-4 and CET-6 have become students' big burden]. (2005, February 25). *Beijing Youth*. Retrieved from <http://news.sina.com.cn/e/2005-02-25/12575938628.shtml>
- Taylor, L. (2000). Stakeholders in language testing. *Research Notes*, 2, 2 – 4.
- Vernon, P. E. (1956). *The measurement of abilities*. London, UK: University of London Press.
- Wall, D. (1996). Introducing new tests into traditional systems: Insights from general education and from innovation theory. *Language Testing*, 13(3), 334 – 354.
- Wall, D. (2000). The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled? *System*, 28(4), 499 – 509.
- Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10(1), 41– 69.
- Wang, Y. (2004). 大学英语教学改革要有的放矢 —从“聋哑英语”谈起. [Reform of the CET should be targeted: Starting from 'dumb English']. *Sino-US English Teaching*, 7(1), 15 – 22.
- Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, 13(3), 318 – 333.
- Watanabe, Y. (2004). Methodology in washback studies, In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp.19 – 36). Mahwah, NJ: Lawrence Erlbaum.
- Weisberg, H., Krosnick, J. A., & Bowen, B. (1996). *Introduction to survey research, polling and data analysis*. Thousand Oaks, CA: Sage.

- Widdowson, H. G. (2003). *Defining issues in English language teaching*. Oxford, UK: Oxford University Press.
- Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco, CA: Jossey-Bass.
- Williamson, M. (1994). The worship of efficiency: Untangling theoretical and practical considerations in writing assessment. *Assessing Writing*, 1(2), 147 – 173.
- Xue, B., & Wang, D. B. (2006, August 12). 专家指出大学生就业难九大原因 扩招影响仍存在 [Experts analyze 9 reasons for difficulty in job hunting and the impact of college enrollment expansion]. *Information Times*. Retrieved from <http://news.sina.com.cn/c/2006-08-12/025010701124.shtml>
- Yang, H., & Jin, Y. (2001). Score interpretation of CET. *Foreign Language World*, (1), 62 – 68.
- Yalden, J. (1981). *Communicative language teaching: Principles and practice*. Toronto, Canada: The Ontario Institute for Studies in Education Press.
- Zhang, S. (2003). 教一考分离: 大学英语四六级考试改革的必由之路. [Separation of teaching and testing: Unavoidable choice of the reform of CET-4 and CET-6]. *Foreign Language Teaching and Research*, 99(5), 385 – 386.
- Zhang, X. (2004). 大学英语测试质疑评论. [Questioning the CET]. *Foreign Language Teaching and Research*, 100(2), 65 – 69.
- Zheng, Y., & Cheng, L. (2008). College English Test (CET) in China. *Language Testing*, 25(3), 408 – 417.

Appendix A

Recruitment Notice for the Survey

Dear teachers,

Greetings!

My name is Zhiling Wu. I'm a Ph.D. candidate in Composition & TESOL Program, English Department, Indiana University of Pennsylvania. At present, if you are teaching English to non-English-major sophomores who are planning to take the CET-4 in June/December 2010, I would very much like to invite you to give me permission to do my dissertation study with your students regarding their English learning and their understanding towards English teaching, English learning and the National College English Band 4 (CET-4). I realize that it will take some of your students' valuable time, and I sincerely appreciate your willingness to permit me to conduct this study with your students.

The purpose of this study is to examine the washback effects of the reformed CET-4 on college English teaching and learning from students' perspectives. Therefore, your students' answers to the survey are very important to this study. This survey will only take them 15-20 minutes. At the end of the survey, your students will be asked if they are willing to further participate in a follow-up qualitative study (5/19/2010 – 6/30/2010) which primarily includes one-on-one interview, essay writing, self-recordings, focus group interviews and email contacts regarding their English studying and preparation for the CET-4. They will be contacted and be requested to sign an Informed Consent Form if they are chosen to be one of the 10 participants for this follow-up qualitative study in your university. Also, in return, during this time period the researcher will be available to answer any questions or help solve problems that are related to their English learning for free.

There are no known risks or discomforts associated with this study. The information gained from this study will be used in my dissertation and related presentations or publications. Also, since it is an anonymous survey, your student's or your name will not be disclosed in the dissertation or other related presentations or publications.

Whether you are willing to give me permission to do this study with your students is entirely voluntary. I promise it will have no bearing on your academic services in the department or university. If you are willing to give me the permission to do this study with your students, please leave your

Name _____

Email _____

Phone number _____

and I will contact you to set up a time and location for me to conduct the survey among your students.

If you have any questions or concerns please feel free to contact the researcher:

Researcher: Zhiling Wu, PhD candidate,
English Department, Indiana University of Pennsylvania
724-463-7615
fkkm@iup.edu

Dissertation Director: Dr. Michael M. Williamson
351 Sutton Hall
Indiana University of Pennsylvania
Indiana, PA 15701
724-357-2671
Michael.m.williamson@iup.edu

Thank you very much in advance for your consideration of this request and cooperation for this study!

Sincerely,

Zhiling Wu

Appendix B

Letter of Information for the Survey

Dear students,

My name is Zhiling Wu. I'm a Ph.D. candidate in Composition & TESOL Program, English Department, Indiana University of Pennsylvania. I would very much like to invite you to participate in my dissertation study regarding your English learning and your understanding towards English teaching, English learning and the National College English Band 4 (CET-4). I realize that it will take some of your valuable time, and I sincerely appreciate your willingness to share your time and experience with me.

The purpose of this study is to examine the washback effects of the reformed CET-4 on college English teaching and learning from students' perspectives. Therefore, your answers to the survey are very important to this study. This survey will only take you 15-20 minutes. At the end of the survey, you will be asked if you are willing to participate further in a follow-up qualitative study (5/19/2010 – 6/30/2010) which primarily includes one-on-one interview, essay writing, self-recordings, focus group interviews and email contacts regarding their English studying and preparation for the CET-4. If you are willing to, please leave your contact information. You will be contacted and be requested to sign an Informed Consent Form if you are chosen to be one of the 10 participants for this follow-up qualitative study in your university. Also, in return, during this time period the researcher will be available to answer any questions or help solve problems that are related to your English learning for free.

There are no known risks or discomforts associated with this study. The information gained from this study will be used in my dissertation and related presentations or publications. Since it is an anonymous, your name will not be disclosed in the dissertation or other related presentations or publications.

Your participation in this study is entirely voluntary. You are free not to participate or to withdraw at any time without adversely affecting your relationship with the researcher. If you choose to participate, you may also withdraw at any time without reasons by notifying the researcher. Upon your request to withdraw, all information pertaining to you will be destroyed. If you choose to participate, all the information will be held in strict confidence. What is more, I promise your participation in this study will have no bearing on your academic standing or services you receive from the college or university. If you have any questions or concerns please feel free to contact the researcher:

Researcher: Zhiling Wu, PhD candidate,

English Department, Indiana University of Pennsylvania

724-463-7615

fkkm@iup.edu

Dissertation Director: Dr. Michael M. Williamson

351 Sutton Hall

Indiana University of Pennsylvania

Indiana, PA 15701

724-357-2671

Michael.m.williamson@iup.edu

Thank you very much in advance for your consideration of this request and cooperation for this study!

Sincerely,

Zhiling Wu

Appendix C

Informed Consent Form

(for the follow-up qualitative study)

Thank you very much for participating in the survey part of the study on washback effects of the reformed CET-4 on college English teaching and learning from students' perspectives. Here I would very much like to invite you to further participate in the follow-up qualitative study of the same research topic which primarily includes one-on-one interview, an essay on learning philosophy and an essay on English learning experiences & English learning belief, focus group interviews and regular email contacts regarding your English studying and preparation for the CET-4. This follow-up qualitative study will last from May 15th to June 30th. I realize that it will take some of your valuable time, and I sincerely appreciate your willingness to share your time and experience with me.

The purpose of this study is to look into the washback effects of the reformed CET-4 on college English teaching and learning from students' perspectives. Therefore, your participation and your data provided are very important to this study. Your data collected will be then analyzed based on Grounded Theory approach and suggestions as to how to promote positive the washback effects of the reformed CET-4 on college English teaching and learning will be proposed as well.

There are no known risks or discomforts associated with this study. The information gained from this study will be used in my dissertation and related presentations or publications. Your real names will not be disclosed in the dissertation or other related presentations or publications.

Your participation in this study is voluntary. You are free not to participate or to withdraw at any time without adversely affecting your relationship with the researcher. If you choose to participate, you may withdraw at any time by notifying the researcher. Upon your request to withdraw, all information pertaining to you will be destroyed. If you choose to participate, all the information will be held in strict confidence. What is more, I promise your participation in this study will have no bearing on your academic standing or services you receive from the institute or university.

Please feel free to contact me, Zhiling Wu at englishworldabc@126.com or my dissertation committee chair, Dr. Michael M. Williamson at Michael.M.Williamson@iup.edu, should you have any questions concerning this study.

If you are willing to participate in this study, please sign the statement below. Thank you for your consideration.

I have read and understood the information on the form, and I consent to volunteer to be a part of this study. I understand that my responses are completely confidential and that I have the right to withdraw at any time.

Name (Please print): _____

Signature: _____ Date: _____

E-mail address: _____

Phone number: _____

Appendix D

Survey Regarding the Influence of the Reformed National College English Test 4 (CET-4) on Your College English Teaching and Learning (English Version Distributed in Phase I)

Dear Students,

Thank you for participating in this survey research! The main purpose of this survey research is to gather information regarding the influence of the reformed CET-4 on your college English teaching and learning. Please answer the questions in the survey honestly based on your true learning experiences. There are no absolutely right or wrong answers. All the information gathered will be used exclusively for research purposes. I really appreciate your cooperation.

Gender ☐male ☐female Age _____ Major _____

You're from _____ province _____ city _____ county/district

Total years of English learning _____ currently you are taking English courses _____ periods/week

English courses you are taking currently: required _____ , electives _____

I. From Question 1 to Question 16, please circle the letter which represents the choice that is right to you. If it is a multiple-choice question, please then circles all the letters which represent the choices that are right to you.

1. At college, on average how many hours do you spend on English learning in your spare time every day?
a. 0.5 hour b. 1 hours c. 2 hours d. 3 hours e. 4 hours or more
2. At college, your primary purpose of learning English is to _____.
a. get high scores in the CET-4 and stand out in job hunting
b. pass required English courses and get high scores
c. learn English cultures further to satisfy my own learning interest
d. improve comprehensive communicative ability in English
e. lay a good foundation in English to study abroad in the future
f. other _____ (please specify)
3. Do you think the new CET-4 can reflect your comprehensive English ability as well as the practical communicative ability in English?
a. not at all b. a small portion c. a large portion d. very much so
4. Which of the following items are newly added fixed items in the CET-4 ever since December 2006? (multiple answers)
a. long conversation b. compound dictation c. fast reading d. cloze
e. integrative part f. translation

5. In what aspects has the reformed CET-4 made impact on your English study? (multiple answers)
- a. learning purpose b. learning attitude c. learning content selection
d. learning interest e. learning methods/strategies f. learning width and depth
g. learning rate and sequence h. other_____ (please specify)
6. In what aspects has the reformed CET-4 made impact on your English teachers' classroom teaching? (multiple answers)
- a. teaching purpose b. teaching attitude c. teaching content selection
d. teaching interest e. teaching methods/strategies f. teaching width and depth
g. teaching rate and sequence h. other_____ (please specify)
7. How has the reformed CET-4 influenced your English learning content? (multiple answers)
- a. vocabulary b. grammar c. reading d. listening e. writing f. speaking
g. translating h. no influence
8. How has the reformed CET-4 influenced your English teachers' teaching content? (multiple answers)
- a. vocabulary b. grammar c. reading d. listening e. writing f. speaking
g. translating h. no influence
9. What is the overall nature of influence made by the reformed CET-4 on your English study in your opinion?
- a. completely negative b. negative aspects outweighs positive aspects
c. positive aspects outweigh negative aspects d. completely positive
10. How often do your English teachers mention the content, format or scoring standard of the reformed CET-4 in class?
- a. never b. seldom c. sometimes d. often e. always
11. How close is the connection between your major and English in your opinion?
- a. not close at all b. little close c. middling d. quite close e. very close
12. What is your plan when you graduate from college?
- a. to work b. to go to graduate school c. to study abroad d. no plans yet
e. other_____ (please specify)
13. The scoring standard and score reporting system of the reformed CET-4 have relieved my burden of English learning.
- a. correct b. incorrect
14. If your CET-4 score met the minimum requirement for the CET- Spoken English Test (SET), would you plan to take the CET-SET?
- a. Yes b. No

15. In order to get a satisfying score in the reformed CET-4; are you planning to enroll in any CET-4 intensive training class offered by your university or by other agencies?

- a. Yes b. No c. I'm taking it now. d. I have taken it.

16. Which test is more important to you?

- a. CET-4 b. CET-6

II. From Question 17 to Question 19, please grade them on the 5-point scale format where 1=never, 2=seldom, 3=sometimes, 4=often, 5=always and put 1-5 in the brackets in front of each item.

17. How often do you have the following activities in your English classes at college at present?

(1=never, 2=seldom, 3=sometimes, 4=often, 5=always)

- () Fast reading
- () Learning grammar items /vocabulary
- () Speaking
- () Intensive reading
- () using multi-media to teach
- () Writing
- () Students' participation
- () Doing mock CET-4 tests/tests of previous years
- () Listening
- () Teacher's solo lecturing learning
- () Carrying out group discussion/pair work/ language games
- () Translating

18. How often do you do the following activities **in English** in your English class at college at present?

(1=never, 2=seldom, 3=sometimes, 4=often, 5=always)

- () Doing group discussion/pair work
- () teacher's lecturing
- () Asking questions /answering questions in class
- () Volunteering to express your ideas in class

19. How often do you do the following activities out of class this semester?

(1=never, 2=seldom, 3=sometimes, 4=often, 5=always)

- () Reading English books /magazines/newspapers
- () Reading major-related books in English
- () Surfing English websites
- () Watching English movies/TV series
- () Communicating face-to-face or online in English
- () Doing CET-4 mock tests or tests of previous years
- () Learning English through teaching platform online
- () Speaking English to myself
- () Taking initiative to understand/learn English cultures and traditions

- () Keeping English journals
- () Participating in English contests
- () Listening to English speeches/presentations
- () Learning English in College Self-access English Center
- () Practicing English in English Corner

III. From Question 20 to Question 27, please fill out the missing information in the blank.

20. Your grade of the English subject in the entrance exam to college is _____ and the total score of the English exam at that time was _____.
21. The primary purpose for the Ministry of the Education to administer the CET-4 is to _____.
22. Do you think your English teachers have the obligation to help you prepare for the CET-4?
 Yes, brief reason: _____
 No, brief reason: _____
23. Among English listening, speaking, reading, writing and translating, I think the one I am most good at is _____, because _____. I think the one I am least good at is _____, because _____. I think the most important skill is _____, because _____.
24. If you plan to prepare for the CET-4, how do you plan to do it? Please specify as specific as you can (e.g. memorizing 5 words every day, doing a mock test every week, writing an English essay every two weeks, and summarizing test-taking strategies regularly, etc.). If you do not plan to prepare for the CET-4, please also give some brief reason below.
 a) I plan to prepare for the CET-4, and my plans are:
 1. _____
 2. _____
 3. _____
 4. _____
 5. _____
 b) I do not plan to prepare for the CET-4. Brief reason: _____
25. What is the CET-4 total grade that you aim for? _____
26. Are you planning to take the CET-4 in June/December 2010? (Please circle one answer.)
Yes No

27. If your answer to Question 26 is **YES**, the researcher of this study would like to do a follow-up qualitative study (5/29/2010 – 6/30/2010) with 10 students in your university, which

mainly includes one-on-one interview, an essay on learning philosophy and an essay on English learning experiences & English learning belief, focus group interviews, and regular email contacts regarding your English learning and preparation for the CET-4. Everything will be conducted at the convenience of your time and location.

In return, during this time period the researcher will be available to answer any questions or help solve problems that are related to your English learning for free. So if you are interested in participating further in the follow-up case study, please leave your

Name _____

Phone number _____

Email _____

[The End of the Survey.]

-Thank you

Appendix E

Survey Regarding the Influence of the Reformed National College English Test 4 (CET-4) on Your College English Teaching and Learning (Chinese Version Distributed in Phase I)

关于你的大学英语学习和新英语四级考试 (CET-4) 准备情况的问卷调查

亲爱的同学:

你好! 感谢你参与本问卷调查。本问卷旨在收集新CET-4对你大学英语学习的影响(反拨效应)。请根据你的学习感受真实地回答本问卷。所有答案无对错之分。本问卷收集的数据仅供教学研究之用。感谢你的大力合作!

性别: ☐男 ☐女 年龄 _____ 岁 专业: _____

籍贯: _____ 省 _____ 市 _____ 区/县

你英语学习时间大约已总计 _____ 年 目前每周英语课总时数是 _____ 节/周

你目前所修的英语课程是: 必修 _____, 选修 _____

I. 从第 1 题到第 16 题, 请选择一个你认为正确的选项, 并在代表这个选项的字母上打钩。多选题则请在所有你认为正确的选项前的字母上打钩。

1. 上大学后, 你平均每天英语课外自主花在英语学习上的时间是 ____。
a. 0.5 小时 b. 1 小时 c. 2 小时 d. 3 小时 e. 4 小时或更多
2. 大学里, 你学习英语最主要的目的是 ____。
a. 在 CET-4 考试中拿高分, 得到就业优势 b. 通过英语必修课考试, 拿学分
c. 进一步了解英美文化, 满足自己的兴趣爱好 d. 提高英语综合交际能力
e. 打好基础, 为以后出国做准备 f. 其它 _____ (请注明)
3. 你认为新 CET-4 能反映你的实际英语综合水平和应用能力吗?
a. 根本不能反映 b. 小部分反映 c. 大部分反映 d. 完全反映
4. 下面哪些是 2006 年 12 月起 CET-4 新增加的固定考试题型? (可多选)
a. 长对话 b. 复合式听写 c. 快速阅读 d. 选词填空 e. 综合测试
f. 句子翻译 g. 短问题回答 h. 改错
5. 新 CET-4 对你的英语学习的影响体现在哪些方面? (可多选)
a. 学习目的 b. 学习态度 c. 学习内容, 资料选择 d. 学习兴趣
e. 学习方法/策略 f. 学习的广度和深度 g. 学习进度和顺序 h. 其它 _____ (请注明)

6. 新 CET-4 对你的大学英语教师教学影响体现在哪些方面? (可多选)
 a. 教学目的 b. 教学态度 c. 教学内容,资料选择 d. 教学热情
 e. 教学方法/策略 f. 学的广度和深度 g. 教学进度和顺序 h. 其它 ____ (请注明)
7. 新 CET-4 对你学习内容的影响体现在哪些方面? (可多选)
 a. 词汇 b. 语法 c. 阅读 d. 听力 e. 写作 f. 口语 g. 翻译 h. 无影响
8. 你认为新 CET-4 对你的英语老师课堂教学内容影响体现在 ____。(可多选)
 a. 词汇 b. 语法 c. 阅读 d. 听力 e. 写作 f. 口语 g. 翻译 h. 无影响
9. 新 CET-4 考试对你的大学英语学习影响的整体性质是什么?
 a. 完全负面影响 b. 负面大于正面影响 c. 正面大于负面影响 d. 完全正面影响
10. 你的大学英语老师上课时会计提及新 CET-4 的考试内容, 形式, 或评分标准吗?
 a. 从不 b. 很少 c. 有时 d. 经常 e. 总是
11. 你觉得英语和你所学的专业联系紧密吗?
 a. 一点儿不紧密 b. 不太紧密 c. 一般 d. 较紧密 e. 非常紧密
12. 大学毕业后, 你有何打算?
 a. 工作 b. 读研 c. 出国 d. 尚无打算 e. 其它 _____ (请注明)
13. 新四级考试新的评分体制和报道方式减轻了你的英语学习压力。
 a. 正确 b. 不正确
14. 如果你的新 CET-4 分数达到大学英语口语考试标准, 你打算参加大学英语口语考试吗?
 a. 打算 b. 不打算
15. 为在新 CET-4 中取得理想的成绩, 你打算上校内或校外的四级培训班吗?
 a. 打算 b. 不打算 c. 正在上 d. 已上过
16. 大学英语四级考试和大学英语六级考试, 哪一考试对你更重要一些?
 a. 四级 b. 六级

II. 从第 17 题到第 19 题, 请对每一题列出的不同项目按 1-5 五个等级进行评估, 并把你所选择的字母填在每一项目前面的括号里。

17. 你目前的大学英语课堂上开展下列活动情况大致是怎样的?

[1=从不 (never) 2=很少 (seldom) 3=有时 (sometimes) 4=经常 (often) 5=总是 (always)]

- | | | |
|------------|-----------------|----------|
| () 快速阅读 | () 语法/单词教授 | () 口语训练 |
| () 阅读理解 | () 使用多媒体教学 | () 写作训练 |
| () 学生课堂参与 | () 四级真题/模拟考题讲解 | () 听力训练 |

() 老师一言堂 () 分组讨论/游戏/活动 () 翻译训练

18. 你目前的大学英语课堂用英语开展下列活动的情况大致是怎样的?

[1=从不 (never) 2=很少 (seldom) 3= 有时 (sometimes) 4=经常 (often) 5=总是 (always)]

- | | |
|----------------|------------------|
| () 学生小组/双人讨论 | () 老师讲课 |
| () 学生课上问/回答问题 | () 学生课上主动发表自己观点 |

19. 这学期, 课外你做下列与英语有关的活动的情况大致是怎样的?

[1=从不 (never) 2=很少 (seldom) 3= 有时 (sometimes) 4=经常 (often) 5=总是 (always)]

- | | |
|----------------------|----------------|
| () 读英文休闲书籍/杂志/报刊 | () 读英文专业书籍 |
| () 浏览英文网站 | () 看英文电影/电视节目 |
| () 面对面/网上与他人用英语交流 | () 做四级真题/模拟考题 |
| () 到网上教学平台学习英语 | () 用英语与自己假想对话 |
| () 主动了解/学习英美文化和风俗知识 | () 用英语写日记/周记 |
| () 参加英语竞赛 | () 听英语讲座 |
| () 到本校多媒体自主学习中心学习英语 | () 到英语角练习英语 |

III. 第 20 题到第 27 题, 请在空白横线上提供相关信息。

20. 你的高考英语单科成绩是_____分, 当时的英语单科满分是_____分。

21. 你觉得新大学英语四级考试的目的是:_____。

22. 你认为大学英语老师有义务帮助学生备考新 CET-4 吗?

有, 简要原因:_____

没有, 简要原因:_____

23. 英语听, 说, 读, 写, 译五方面, 你觉得你最擅长的是_____, 因为_____; 你最不擅长的是_____, 因为_____; 你觉得最重要的是_____, 因为_____。

24. 如果你将参加新 CET-4 考试, 你会如何备考? 请简要罗列你的备考计划 (如: 每天背单词 5 个, 每周做一次模拟题, 每两周写一篇英文作文, 总结应试技巧等); 如果你打算裸考, 则请给出简要原因。

a) 我有备考计划: 1. _____
2. _____
3. _____
4. _____
5. _____

b) 我准备裸考, 简要原因: _____

你对大学英语四级考试有何改进意见:_____

25. 你为自己定下的新 CET-4 分数总分大致是 _____ 分。

26. 你准备参加 2010 年 6 月/12 月的新大学英语四级考试吗? 回答: (是 否)

27. 如果你对第 26 题的回答是肯定的, 即你准备参加 2010 年 6 月/12 月的新大学英语四级考试, 那么本问卷调查者想诚恳邀请你参加接下来的个案跟踪调查研究。该研究将在 5/29/2010 至 6/30/2010 之间进行。个案调查项目主要包括: 针对你大学英语四级考试复习备考和大学英语学习方面情况的单独采访、小组采访、作文/资料搜集、电子邮件交流等。具体的个案跟踪调研时间和地点将视调研对象的方便而定。

调查者将在所有表示愿意参加的学生中根据相关指标挑选 10 位调研对象。作为回报, 调研期间, 调查者也将义务地为这 10 位调研对象提供英语学习方面的帮助。如果你愿意参加接下来的个案跟踪调查, 请在下面留下你的

姓名: _____

联系电话: _____

电子邮件: _____

~~~~ 谢谢合作! ~~~~

## Appendix F

### Major Interview Questions in Phase II

1. Please briefly introduce your name, major, and your hometown.
2. What kind of students are good students to you?
3. What kind of learners are good English learners?
4. What is the key to successful English learning to you?
5. Why are you learning English? Is it important to you? Why?
6. What are your long-term goal and short-term goal for English learning?
7. When did you start learning English? How was English taught in your primary school, middle school and high school? Please explain one by one in terms of the teacher's, teaching style, class activities, students' motivations, students' learning styles, and class dynamics, etc.
8. What are the major similarities and differences between English learning at the college level and before the college level, if there is any?
9. Do you think your English learning experiences before the college level have any impact on your English learning now? In what way if there is any?
10. Do you think it is fair to divide students to fast class and ordinary class at college?
11. What are the English courses you are taking this semester? How often are they offered per week? In each class, how many students are there? What textbook is used? Does your English teacher use computer/internet to teach? How often does your English teacher use online resources?
12. What is your current English teacher's teaching style? What activities are usually conducted in your English classes? How is the class interaction in your English course? Are there a lot of opportunities for students to practice English in class? What is your English homework like? Please explain.
13. Does your current English teacher pay attention to students' autonomous learning? What roles do teacher and students play in English classes?
14. What do you think you have learned from your college English classes? What do you think of your college English classes?
15. How is your English proficiency evaluated in your English class at other times?
16. Do you communicate with your English teacher - in class and out of class? Why?
17. Besides your English courses, are there any other situations in which you use English? Do you have other subjects taught in English? Please explain.
18. Have you ever talked to a native English speaker in person?
19. How much time do you spend on English out of class this semester? What do you do?
20. What are your major English learning strategies now? What were your English learning strategies before the college level?

21. Does your current English teacher pay much attention to the CET-4? Do you think your English teacher has the responsibility to help you prepare for the CET-4? Please explain.
22. What is your mid-term and final English assessment like respectively? What do you think of them?
23. Besides English textbooks, what other English books do you have? Are there any books related to the CET-4?
24. Do you know the reformed CET-4 well? Please elaborate your knowledge about the reformed CET-4? (e.g. reason for the reform, objective, time, test format, test order, total score, score report, certificate, etc.)
25. Do you know the CET-SET? What is the qualification to take it? If you are qualified, will you take it?
26. When did you start preparing the CET-4? What is your preparation plan? How is it implemented?
27. What is the key to achieving a good score in the CET-4? Why?
28. To what extent has the reformed CET-4 influenced your perceptions of the practices of English teaching and learning at the college level? Is it largely positive or negative?
29. What is the major aspect in the CET-4 that you are worried about, if there is any?
30. Do you want to abolish the CET-4?
31. If the CET-4 was abolished, what were the things related to English learning you would stop doing?
32. What will be trend for college English teaching in the future to you?
33. What are your plans relating to English learning after taking the CET-4?
34. Do you want to share with me your CET-4 score?

## Appendix G

### Survey Regarding the Influence of the Reformed National College English Test 4 (CET-4) on Your College English Teaching and Learning (English Version Distributed in Phase III)

Note: Modifications from the survey distributed in Phase I are in red.

Dear Students,

Thank you for participating in this survey research! The main purpose of this survey research is to gather information regarding the influence of the reformed CET-4 on your college English teaching and learning. Please answer the questions in the survey honestly based on your true learning experiences. There are no absolutely right or wrong answers. All the information gathered will be used exclusively for research purposes. I really appreciate your cooperation.

Gender ☐male ☐female Age \_\_\_\_\_ Major \_\_\_\_\_

You're from \_\_\_\_\_ province \_\_\_\_\_ city \_\_\_\_\_ county/district

Total years of English learning \_\_\_\_\_ currently you are taking English courses \_\_\_\_\_ periods/week

English courses you are taking currently: required \_\_\_\_\_, electives \_\_\_\_\_, English credit \_\_\_\_\_

**IV. From Question 1 to Question 16, please circle the letter which represents the choice that is right to you. If it is a multiple-choice question, please then circles all the letters which represent the choices that are right to you.**

5. At college, on average how many hours do you spend on English learning in your spare time every day?  
a. 0.5 hour   b. 1 hours   c. 2 hours   d. 3 hours   e. 4 hours or more
6. At college, your primary purpose of learning English is to \_\_\_\_\_.  
a. get high scores in the CET-4 and stand out in job hunting  
b. pass required English courses and get high scores  
c. learn English cultures further to satisfy my own learning interest  
d. improve comprehensive communicative ability in English  
e. lay a good foundation in English to study abroad in the future  
f. other \_\_\_\_\_ (please specify)
7. Do you think the new CET-4 can reflect your comprehensive English ability as well as the practical communicative ability in English?  
a. not at all   b. a small portion   c. a large portion   d. very much so
8. Which of the following items are newly added fixed items in the CET-4 ever since December 2006? (multiple answers)  
a. long conversation   b. compound dictation   c. fast reading   d. cloze

e. integrative part      f. translation      g. short question answer      h. Error Correction

5. In what aspects has the reformed CET-4 made impact on your English study? (multiple answers)

- a. learning purpose      b. learning attitude      c. learning content selection  
d. learning interest      e. learning methods/strategies      f. learning width and depth  
g. learning rate and sequence      h. other\_\_\_\_\_ (please specify)

6. In what aspects has the reformed CET-4 made impact on your English teachers' classroom teaching? (multiple answers)

- a. teaching purpose      b. teaching attitude      c. teaching content selection  
d. teaching interest      e. teaching methods/strategies      f. teaching width and depth  
g. teaching rate and sequence      h. other\_\_\_\_\_ (please specify)

7. How has the reformed CET-4 influenced your English learning content? (multiple answers)

- a. vocabulary      b. grammar      c. reading      d. listening      e. writing      f. speaking  
g. translating      h. no influence

8. How has the reformed CET-4 influenced your English teachers' teaching content? (multiple answers)

- a. vocabulary      b. grammar      c. reading      d. listening      e. writing      f. speaking  
g. translating      h. no influence

9. What is the overall nature of influence made by the reformed CET-4 on your English study in your opinion?

- a. completely negative      b. negative aspects outweighs positive aspects  
c. positive aspects outweigh negative aspects      d. completely positive

10. How often do your English teachers mention the content, format or scoring standard of the reformed CET-4 in class?

- a. never      b. seldom      c. sometimes      d. often      e. always

11. How close is the connection between your major and English in your opinion?

- a. not close at all      b. little close      c. middling      d. quite close      e. very close

12. What is your plan when you graduate from college?

- a. to work      b. to go to graduate school      c. to study abroad      d. no plans yet  
e. other\_\_\_\_\_ (please specify)

13. The scoring standard and score reporting system of the reformed CET-4 have relieved my burden of English learning.

- a. correct      b. incorrect

14. If your CET-4 score met the minimum requirement for the CET- Spoken English Test (SET), would you plan to take the CET-SET?  
a. Yes     b. No
15. In order to get a satisfying score in the reformed CET-4; are you planning to enroll in any CET-4 intensive training class offered by your university or by other agencies?  
a. Yes     b. No     c. I'm taking it now.     d. I have taken it.
16. Which test is more important to you?  
a. CET-4     b. CET-6

**V. From Question 17 to Question 19, please grade them on the 5-point scale format where 1=never, 2=seldom, 3=sometimes, 4=often, 5=always and put 1-5 in the brackets in front of each item.**

17. How often do you have the following activities in your English classes at college at present?

(1=never, 2=seldom, 3=sometimes, 4=often, 5=always)

- ( ) Fast reading
- ( ) Learning grammar items /vocabulary
- ( ) Speaking
- ( ) Intensive reading
- ( ) using multi-media to teach
- ( ) Writing
- ( ) Students' participation
- ( ) Doing mock CET-4 tests/tests of previous years
- ( ) Listening
- ( ) Teacher's solo lecturing learning
- ( ) Carrying out group discussion/pair work/ language games
- ( ) Translating

18. How often do you do the following activities **in English** in your English class at college at present?

(1=never, 2=seldom, 3=sometimes, 4=often, 5=always)

- ( ) Doing group discussion/pair work
- ( ) teacher's lecturing
- ( ) Asking questions /answering questions in class
- ( ) Volunteering to express your ideas in class

19. How often do you do the following activities out of class this semester?

(1=never, 2=seldom, 3=sometimes, 4=often, 5=always)

- ( ) Reading English books /magazines/newspapers
- ( ) Reading major-related books in English
- ( ) Surfing English websites
- ( ) Watching English movies/TV series
- ( ) Communicating face-to-face or online in English

- ( ) Doing CET-4 mock tests or tests of previous years
- ( ) Learning English through teaching platform online
- ( ) Speaking English to myself
- ( ) Taking initiative to understand/learn English cultures and traditions
- ( ) Keeping English journals
- ( ) Participating in English contests
- ( ) Listening to English speeches/presentations
- ( ) Learning English in College Self-access English Center
- ( ) Practicing English in English Corner

**III. From Question 20 to Question 28, please fill out the missing information in the blank.**

20. Your grade of the English subject in the entrance exam to college is \_\_\_\_\_ and the total score of the English exam at that time was \_\_\_\_\_.
21. The primary purpose for the Ministry of the Education to administer the CET-4 is to \_\_\_\_\_.
22. Do you think your English teachers have the obligation to help you prepare for the CET-4?  
 Yes, brief reason: \_\_\_\_\_  
 No, brief reason: \_\_\_\_\_
23. Among English listening, speaking, reading, writing and translating, I think the one I am most good at is \_\_\_\_\_, because \_\_\_\_\_. I think the one I am least good at is \_\_\_\_\_, because \_\_\_\_\_. I think the most important skill is \_\_\_\_\_, because \_\_\_\_\_.
24. If you plan to prepare for the CET-4, how do you plan to do it? Please specify as specific as you can (e.g. memorizing 5 words every day, doing a mock test every week, writing an English essay every two weeks, and summarizing test-taking strategies regularly, etc.). If you do not plan to prepare for the CET-4, please also give some brief reason below.
- a) I plan to prepare for the CET-4, and my plans are:
1. \_\_\_\_\_
  2. \_\_\_\_\_
  3. \_\_\_\_\_
  4. \_\_\_\_\_
  5. \_\_\_\_\_
- b) I do not plan to prepare for the CET-4. Brief reason: \_\_\_\_\_
25. What is the CET-4 total grade that you aim for? \_\_\_\_\_

26. In your opinion, what is the relationship between College English course and the reformed CET-4?

---

27. Are you planning to take the CET-4 in **December** 2010? (Please circle one answer.)

Yes      No

28. If your answer to Question 27 is **YES**, the researcher of this study would like to do a follow-up qualitative study (11/21/2010 – 12/31/2010) with 4 students in your university, which mainly includes one-on-one interview, an essay on learning philosophy and an essay on English learning experiences & English learning belief, focus group interviews, and regular email contacts regarding your English learning and preparation for the CET-4. Everything will be conducted at the convenience of your time and location.

In return, during this time period the researcher will be available to answer any questions or help solve problems that are related to your English learning for free. So if you are interested in participating further in the follow-up case study, please leave your

Name \_\_\_\_\_

Phone number \_\_\_\_\_

Email \_\_\_\_\_

[The End of the Survey.]

-Thank you

## Appendix H

### Survey Regarding the Influence of the Reformed National College English Test 4 (CET-4) on Your College English Teaching and Learning (Chinese Version Distributed in Phase III)

#### 关于你的大学英语学习和新英语四级考试 (CET-4)准备情况的问卷调查

亲爱的同学:

你好! 感谢你参与本问卷调查。本问卷旨在收集新CET-4对你大学英语学习的影响(反拨效应)。请根据你的学习感受真实地回答本问卷。所有答案无对错之分。本问卷收集的数据仅供教学研究之用。感谢你的大力合作!

性别: ☐男 ☐女 (请在答题卡选择题部分男填 1,女填 2)

年龄\_\_\_\_\_岁

专业: \_\_\_\_\_ (请在答题卡选择题部分写拼音, 比如: 电信  
请写: dianxin)

高考生源地: \_\_\_\_\_ 省 (请在答题卡选择题部分选择下面相应的数字)

- 1、华东地区 (包括山东、江苏、安徽、浙江、福建、上海);
- 2、华南地区 (包括广东、广西壮族自治区、海南);
- 3、华中地区 (包括湖北、湖南、河南、江西);
- 4、华北地区 (包括北京直辖市、天津直辖市、河北、山西、内蒙古自治区);
- 5、西北地区 (包括宁夏回族自治区、新疆维吾尔自治区、青海、陕西、甘肃);
- 6、西南地区 (包括四川、云南、贵州、西藏自治区、重庆直辖市);
- 7 东北地区 (包括辽宁、吉林、黑龙江);
- 8、台港澳地区 (包括台湾、香港、澳门)。

你英语学习时间大约已总计 \_\_\_\_\_ 年 目前每周英语课总时数是 \_\_\_\_\_ 节/周

这学期你所修的课程一共 \_\_\_\_\_ 门, 学分总数共计: \_\_\_\_\_ 分, 其中英语课的学分数为: \_\_\_\_\_ 分

**I. 从第 1 题到第 16 题, 请选择一个你认为正确的选项, 答题卡上请填该选项前所对应的数字。**

1. 上大学后, 你平均每天英语课外自主花在英语学习上的时间是 \_\_\_\_。

- 1). 0.5 小时    2). 1 小时    3). 2 小时    4). 3 小时    5). 4 小时或更多

2. 大学里,你学习英语最主要的目的是\_\_\_\_\_。
- 1.在 CET-4 考试中拿高分,得到就业优势      2.通过英语必修课考试,拿学分  
3. 进一步了解英美文化,满足自己的兴趣爱好      4. 提高英语综合交际能力  
5. 打好基础,为以后出国做准备      6. 其它\_\_\_\_\_
3. 你认为新 CET-4 能反映你的实际英语综合水平和应用能力吗?
1. 根本不能反映    2. 小部分反映    3. 大部分反映    4. 完全反映
4. 下面哪些是 2006 年 12 月起 CET-4 新增加的固定考试题型?(可多选)
1. 长对话    2. 复合式听写    3. 快速阅读    4. 选词填空    5. 综合测试  
6. 句子翻译    7. 短问题回答    8. 改错
5. 新 CET-4 对你的英语学习的影响体现在哪些方面?(可多选)
1. 学习目的      2. 学习态度      3. 学习内容,资料选择      4. 学习兴趣  
5. 学习方法/策略    6. 学习的广度和深度    7. 学习进度和顺序    8. 其它 \_\_\_\_\_
6. 新 CET-4 对你的大学英语教师教学影响体现在哪些方面?(可多选)
1. 教学目的      2. 教学态度    3. 教学内容,资料选择    4. 教学热情  
5. 教学方法/策略    6. 学的广度和深度    7. 教学进度和顺序    8. 其它 \_\_\_\_\_
7. 新 CET-4 对你学习内容的影响体现在哪些方面?(可多选)
1. 词汇    2. 语法    3. 阅读    4. 听力    5. 写作    6. 口语    7. 翻译    8. 无影响
8. 你认为新 CET-4 对你的英语老师课堂教学内容影响体现在\_\_\_\_。(可多选)
1. 词汇    2. 语法    3. 阅读    4. 听力    5. 写作    6. 口语    7. 翻译    8. 无影响
9. 新 CET-4 考试对你的大学英语学习影响的整体性质是什么?
1. 完全负面影响    2. 负面大于正面影响    3. 正面大于负面影响    4. 完全正面影响
10. 你的大学英语老师上课时会展及新 CET-4 的考试内容,形式,或评分标准吗?
1. 从不    2. 很少    3. 有时    4. 经常    5. 总是

11. 你觉得英语和你所学的专业联系紧密吗?

1. 一点儿不紧密 2. 不太紧密 3. 一般 4. 较紧密 5. 非常紧密

12. 大学毕业后, 你有何打算?

1. 工作 2. 读研 3. 出国 4. 尚无打算 5. 其它\_\_\_\_\_

13. 新四级考试新的评分体制和报道方式减轻了你的英语学习压力。

1. 正确 2. 不正确

14. 如果你的新 CET-4 分数达到大学英语口语考试标准, 你打算参加大学英语口语考试吗?

1. 打算 2. 不打算

15. 为在新 CET-4 中取得理想的成绩, 你打算上校内或校外的四级培训班吗?

1. 打算 2. 不打算 3. 正在上 4. 已上过

16. 大学英语四级考试和大学英语六级考试, 哪一考试对你更重要一些?

1. 四级 2. 六级

**II. 从第 17 题到第 19 题, 请对每一题列出的不同项目按 1-5 五个等级进行评估, 并把你所选择的字母填在每一项目前面的括号里。**

17. 你目前的大学英语课堂上开展下列活动情况大致是怎样的?

[1=从不 (never) 2=很少 (seldom) 3= 有时 (sometimes) 4=经常 (often) 5=总是 (always)] (请横着看题一排一排作答)

- |            |                 |          |
|------------|-----------------|----------|
| ( ) 快速阅读   | ( ) 语法/单词教授     | ( ) 口语训练 |
| ( ) 阅读理解   | ( ) 使用多媒体教学     | ( ) 写作训练 |
| ( ) 学生课堂参与 | ( ) 四级真题/模拟考题讲解 | ( ) 听力训练 |
| ( ) 老师一言堂  | ( ) 分组讨论/游戏/活动  | ( ) 翻译训练 |

18. 你目前的大学英语课堂用英语开展下列活动的情况大致是怎样的?

[1=从不 (never) 2=很少 (seldom) 3= 有时 (sometimes) 4=经常 (often) 5=总是 (always)] (请横着看题一排一排作答)

- |                |                  |
|----------------|------------------|
| ( ) 学生小组/双人讨论  | ( ) 老师讲课         |
| ( ) 学生课上问/回答问题 | ( ) 学生课上主动发表自己观点 |

19. 这学期, 课外你做下列与英语有关的活动的情况大致是怎样的?

[1=从不 (never) 2=很少 (seldom) 3= 有时 (sometimes) 4=经常 (often) 5=总是 (always)] (请横着看题一排一排作答)

- |                    |                |
|--------------------|----------------|
| ( ) 读英文休闲书籍/杂志/报刊  | ( ) 读英文专业书籍    |
| ( ) 浏览英文网站         | ( ) 看英文电影/电视节目 |
| ( ) 面对面/网上与他人用英语交流 | ( ) 做四级真题/模拟考题 |

- |                                           |                                     |
|-------------------------------------------|-------------------------------------|
| <input type="checkbox"/> 到网上教学平台学习英语      | <input type="checkbox"/> 用英语与自己假想对话 |
| <input type="checkbox"/> 主动了解/学习英美文化和风俗知识 | <input type="checkbox"/> 用英语写日记/周记  |
| <input type="checkbox"/> 参加英语竞赛           | <input type="checkbox"/> 听英语讲座      |
| <input type="checkbox"/> 到本校多媒体自主学习中心学习英语 | <input type="checkbox"/> 到英语角练习英语   |

**III. 第 20 题到第 27 题，请在空白横线上提供相关信息。**

20. 你的高考英语单科成绩是\_\_\_\_\_分/当时的英语单科满分是\_\_\_\_\_分。

21. 你觉得新大学英语四级考试的目的是:\_\_\_\_\_。

22. 你认为大学英语老师有义务帮助学生备考新 CET-4 吗?

有(答题卡上请填 1), 简要原因:\_\_\_\_\_

没有(答题卡上请填 2), 简要原因:\_\_\_\_\_

23. 英语(听 1), (说 2), (读 3), (写 4), (译 5) 五方面, 你觉得你最擅长的是\_\_\_\_\_, 因为\_\_\_\_\_; 你最不擅长的是\_\_\_\_\_, 因为\_\_\_\_\_; 你觉得最重要的是\_\_\_\_\_, 因为\_\_\_\_\_。

24. 如果你将参加新 CET-4 考试, 你会如何备考? 请简要罗列你的备考计划(如: 每天背单词 5 个, 每周做一次模拟题, 每两周写一篇英文作文, 总结应试技巧等); 如果你打算裸考, 则请给出简要原因。

a) 我有备考计划(答题卡上请填 1): 1. \_\_\_\_\_  
2. \_\_\_\_\_  
3. \_\_\_\_\_  
4. \_\_\_\_\_  
5. \_\_\_\_\_

b) 我准备裸考(答题卡上请填 2), 简要原因: \_\_\_\_\_

你对大学英语四级考试有何改进意见:\_\_\_\_\_

25. 你为自己定下的新 CET-4 分数总分大致是\_\_\_\_\_分。

26. 你觉得你的大学英语课和 CET-4 考试的关系是:

27. 你准备参加 2010 年 12 月的新大学英语四级考试吗? 回答: (是 1) (否 2)

28. 如果你对第 27 题的回答是肯定的, 即你准备参加 2010 年 12 月的新大学英语四级考试, 那么本问卷调查者想诚恳邀请你参加接下来的个案跟踪调查研究。该研究将在 11/1/2010 至 12/31/2010 之间进行。个案调查项目主要包括: 针对你大学英语

四级考试复习备考和大学英语学习方面情况的单独采访、资料搜集、电子邮件/QQ交流等。具体的个案跟踪调研时间和地点将视调研对象的方便而定。

调查者将在所有表示愿意参加的学生中根据相关指标挑选 4 位调研对象。作为回报，调研期间，调查者也将义务地为这 4 位调研对象提供英语学习方面的帮助。如果你愿意参加接下来的个案跟踪调查，请在下面留下你的

姓名：\_\_\_\_\_ 联系电话：\_\_\_\_\_

电子邮件：\_\_\_\_\_ QQ: \_\_\_\_\_

~~~~~ 谢谢合作！~~~~~

Appendix I

Major Interview Questions for Further Follow-up in Phase IV

1. Why are you learning English? Is it important to you? Why?
2. What are your long-term goal and short-term goal for English learning?
3. How many courses and credits are you taking this semester? How many courses and credits are English related? How about last semester?
4. What is your standard for selecting your English teacher when registering for English courses?
5. What are the English courses you are taking this semester? How often are they offered per week? In each class, how many students are there? What textbook is used? Does your English teacher use computer/internet to teach? How often does your English teacher use online resources?
6. What is your current English teacher's teaching style? What activities are usually conducted in your English classes? How is the class interaction in your English course? Are there a lot of opportunities for students to practice English in class? What is your English homework like? Please explain.
7. Does your current English teacher pay attention to students' autonomous learning? What roles do teacher and students play in English classes now?
8. What do you think you have learned from your college English classes? What do you think of your college English classes?
9. What are the major similarities and differences between your English class this semester and last semester?
10. How is your English proficiency evaluated in your English class at other times?
11. Do you communicate with your English teacher - in class and out of class? Why?
12. Besides your English courses, are there any other situations in which you use English? Do you have other subjects taught in English? Please explain.
13. Have you ever talked to a native English speaker in person?
14. How much time do you spend on English out of class this semester? What do you do?
15. What are your major English learning strategies now? Is there any change compared to last semester?
16. What is the relationship between the CET-4 and your college English courses? Does your current English teacher pay much attention to the CET-4? Do you think your English teacher has the responsibility to help you prepare for the CET-4? Please explain.
17. What is your mid-term and final English assessment like respectively? What do you think of them?

18. Besides English textbooks, what other English books do you have? Are there any books related to the CET-4?
19. Do you know the reformed CET-4 well now? Please elaborate your knowledge about the reformed CET-4? (e.g. reason for the reform, objective, time, test format, test order, total score, score report, certificate, etc.)
20. Do you know the CET-SET? What is the qualification to take it? If you are qualified, will you take it?
21. When did you start preparing the CET-4? What is your preparation plan? How is it implemented?
22. What is the key to achieving a good score in the CET-4? Why?
23. To what extent has the reformed CET-4 influenced your perceptions of the practices of English teaching and learning at the college level? Is it largely positive or negative?
24. Compared to last semester, is there any change in you in terms of the preparation or perception towards the CET-4? Are you feeling the pressure now?
25. What is the major aspect in the CET-4 that you are worried about, if there is any?
26. Do you want to abolish the CET-4?
27. If the CET-4 was abolished, what were the things related to English learning you would stop doing?
28. What will be the trend for college English teaching in the future to you?
29. What are your plans relating to English learning after taking the CET-4?
30. Do you want to share with me your CET-4 score?

Appendix J

An Overall Timeline of the Study

Table 35

Phase I and Phase II

| University | Class Type | Phase I – Survey | Phase II – Interview |
|--------------|---|---------------------|----------------------|
| | | 5/8 – 5/18/2010 | 5/19 – 6/30/2010 |
| | | No. of Participants | |
| University A | CET-4 in June | 29 | 10 |
| | CET-4 in Dec.
Teacher I | 38 | 1 |
| | CET-4 in Dec.
Teacher II | 35 | 1 |
| | Total | 102 | 12 |
| University B | CET-4 Ability
Training
Practice Class | 55 | 2 |
| | Fast Class | 31 | 4 |
| | Ordinary
Class | 39 | 2 |
| | Total | 125 | 8 |
| University C | Fast Class
Teacher I | 41 | 4 |
| | Fast Class
Teacher II | 39 | 3 |
| | Ordinary Class | 36 | 2 |
| | Total | 116 | 9 |
| Phase I & II | Grand Total | 343 | 29 |

Table 36

Phase III and IV in University A

| Phase | Phase III | Phase IV | Phase IV |
|------------|---------------------|-------------------|----------------|
| | Survey | Interview | Self-Recording |
| Time | 11/1 -11/18/2010 | 11/19 -12/20/2010 | |
| Class | No. of Participants | | |
| Teacher I | 45 | 3 | 2 |
| Teacher II | 26 | 2 | 1 |
| Total | 71 | 5 | 3 |

Table 37

Phase III and IV: A Further Follow-up of Participants in Phase II

| Phase | Phase III | Phase IV | Phase IV |
|--------------|-----------------------|------------------------|-----------------|
| | Survey Returned | Interviews | Self-recordings |
| Time | 11/1/2010 -11/18/2010 | 11/19/2010 -12/20/2010 | |
| University | No. of Participants | | |
| University A | 2 | 2 | 2 |
| University B | 2 | 2 | 1 |
| University C | 1 | 2 | 1 |
| Total | 5 | 6 | 4 |