

A STUDY OF THE ROLE OF THE ‘TEACHER FACTOR’ IN WASHBACK

Jing Wang

Department of Integrated Studies in Education  
McGill University

July, 2010

A thesis submitted to McGill University in partial fulfilment of the requirements of the degree of  
Doctor of Philosophy

© Jing Wang, 2010

## Abstract

The current study seeks to explore the washback effects of the CET (College English Test) on teacher beliefs, interpretations and practices, and in particular seeks to discover the way the ‘teacher factor’ is manifested in the washback phenomenon. It also investigates the pedagogical as well as the social and personal complexities influencing teachers’ beliefs and interpretations and practices.

This study addresses the research question: What role does the ‘teacher factor’ play in washback in the Chinese university context? Participants were 195 tertiary-level EFL teachers of the non-English programs. The main purpose of this study was to investigate whether tests constitute a major constraint on CE (College English) instructional innovation in China. In addition, the intent of the study was to find out what aspects pertinent to this factor (e.g., teacher beliefs, teacher knowledge, experiences) present the major barrier to the implementation of instructional change.

A mixed methods approach combining both qualitative and quantitative methods of data collection and data analysis was adopted in this study. A teacher survey and in-depth case studies (through focused group/individual interviews and classroom observations) were used to collect data. Data were analyzed in two phases. Qualitative analysis involved the use of constant comparative method, while quantitative analysis in this study involved descriptive statistics (e.g., frequency counts, means, standard deviations, etc.) and inferential statistics (e.g., exploratory factor analysis, confirmatory factor analysis, and structural equation modeling).

The findings from this study suggest that the CET coupled with various interrelated components of the ‘teacher factor’ is involved in fostering the washback effect. Given the complexities underlying the washback phenomenon, the educational change carried out in curriculum and assessment is not sufficient on its own to entail teacher change in terms of pedagogical strategies. It appears that for fundamental changes in teacher practice to occur, they must be accompanied by other changes in teachers’ knowledge, beliefs, attitudes and thinking that inform such practice. The pedagogical implications from the study include (1) the need to inform EFL teachers of the purpose and benefits of tests to help them maintain a positive attitude toward the role of tests in the educational system; and (2) enhanced teacher training in how to make their teaching more truly communicative in nature. It is hoped that the issues identified in this study can serve to inform educational authorities, test designers and teachers, and serve as an impetus to upgrade EFL teaching in China.

---

## Résumé

La présente étude vise à explorer le effet de retour (*washback*) du test d'anglais au niveau collégiale (TAC) sur les croyances, interprétations et pratiques de l'enseignant. Elle cherche en particulier à découvrir la voie par laquelle le « facteur enseignant » est manifesté dans ce phénomène. Elle enquête aussi sur les complexités pédagogiques de même que sociales et personnelles influençant les croyances, interprétations et pratiques des enseignants.

Cette étude pose la question de recherche : Quel rôle joue le 'facteur enseignant' en *washback* dans le contexte de l'université Chinoise? Les participants étaient 195 enseignants universitaire de l'anglais langue étrangère (ALE) des programmes non-anglais. Le propos principal de cette étude était de chercher à savoir si les tests constituent une contrainte majeure sur l'innovation d'enseignement d'anglais au niveau collégiale (AC) en Chine. De plus, l'intention de cette étude était de trouver quels aspects pertinents de ce facteur (ie., croyances de l'enseignant, connaissance et expériences de l'enseignant) présentent la barrière majeure à la mise en œuvre du changement d'instruction.

Une approche de méthodes mixtes combinant à la fois les méthodes qualitative et quantitative de collection et d'analyse de données fut adoptée dans cette étude. Un sondage auprès d'enseignants et des études de cas en profondeur (à travers des entrevues individuelles et en groupes puis des observations en classes) furent utilisés pour collecter les données. Celles-ci furent analysées en deux phases. L'analyse qualitative impliquait l'utilisation d'une méthode comparative constante, tandis que l'analyse quantitative dans cette étude impliquait des statistiques descriptives (i.e., comptages de fréquence, moyens, déviations standards etc) et des statistiques déductives (i.e., analyse de facteur exploratoire, analyse de facteur confirmatoire, et modelage d'équation structurelle).

Les résultats de cette étude suggèrent que le TAC couplé à diverses composantes interreliées du 'facteur enseignant' participe à engendrer le *washback*. Étant donné les complexités sous-jacentes à ce phénomène, le changement éducationnel apporté au curriculum et à l'évaluation n'est pas suffisant par lui-même pour entraîner un changement chez l'enseignement en termes de stratégies pédagogiques. Il apparaît que pour que les changements fondamentaux puissent avoir lieu dans la pratique des enseignants, ceux-ci doivent être accompagnés d'autres changements dans les connaissances, croyances, attitudes et pensées qui inspirent une telle pratique chez les enseignants.

Les implications pédagogiques à partir de l'étude incluent (1) le besoin d'informer les enseignants de l'ALE de la raison d'être et des bénéfices des tests pour les aider à maintenir une attitude positive envers le rôle des tests dans le système d'éducation; et (2) rehausser l'entraînement des enseignants sur la manière de rendre leur enseignement véritablement plus communicatif. Il est espéré que les enjeux identifiés dans cette études puissent inspirer les autorités éducationnelles, les concepteurs de tests et les enseignants, puis servir d'élan au rehaussement de l'enseignement de l'ALE en Chine.

## Acknowledgements

I would like to thank with all my heart all the people who have helped direct my doctoral program and provide assistance, support and inspiration during the research and writing phases of this dissertation.

Foremost, I want to thank my supervisor, Dr. Carolyn Turner for her foresight and intelligent guidance in helping me sharpen a number of conceptual and methodological issues that shaped my study. I owe her a tremendous debt of gratitude for the innumerable helpful comments and invaluable feedback she provided me from my research conceptualization through its completion.

I want to particularly thank her for being responsive to my needs and making herself readily available whenever I approached her for advice and help in matters concerning my studies at McGill in general and on my dissertation in particular. She has given me support in countless ways over the years. She is a friend, mentor, and role model.

I would also like to thank her for the many discussions I had with her which helped in keeping me going in the right direction. Without her selfless dedication and generous contributions of time and effort, this dissertation would never have become a reality.

I would also like to acknowledge the extent of my intellectual debt to the professors on my advisory committee, Doctor Mela Sarkar and Doctor Janet Donin, whose expertise, advice, insights and encouragement helped to keep me on target in defining and reaching my objectives. A special note of thanks goes to Mela for her concern and support, and insightful comments and critique on my dissertation.

I wish also to express my great appreciation to a number of professors at McGill whose courses have helped equip me with necessary theoretical and methodological tools for conducting my research. I am extremely grateful to Dr. Roy Lyster, whose thoughtful, reflective, and rigorous input and his suggestions for my initial proposal played an important role in the evolution of my thinking about the subject of language teaching and testing, and bolstered my confidence in the design of my research. I would like to thank Dr. Steve Jorden and Dr. Robert Bracewell for introducing me to a wide range of theoretical and methodological perspectives for examining my research problems and doing research.

I am also profoundly grateful to May Tan for constantly sharing with me her

thought-provoking ideas. Her discussions with me on many issues involved in this study, her encouragement through each phase of my study, and her generous help to review my final dissertation were much appreciated.

My deepest gratitude goes to all the members of my Ph.D group (Carolyn's doctoral students) – May, Christian, Talia, Beverly, Heike, Pamela, and Candace. Their inspiration, help and support played an important role in my life.

My gratitude is also extended to every person at McGill and elsewhere who contributed to this dissertation one way or another.

In writing this dissertation, I also benefited enormously from the support and help of Mr. Yong Li, the statistician who provided me ongoing technical advice and assisted me in analyzing the quantitative data of my study. I am indebted to him for having taken tremendous amount of time to tutor me and answer my questions concerning issues of statistical analyses. I also offer my profound thanks to Andrew and Professor Hanfeng Chen for the priceless feedback that they provided me on data analysis and discussions of many of the statistical issues involved in this study.

My thanks and gratitude also go to a number of friends and former colleagues who, while I was collecting my data, rendered me countless help. I also want to express my especial thanks to all the participating teachers in this study who graciously participated in this study and allowed me to visit their classes. For confidential reasons, I shall not mention them by name.

I am grateful to all my family members for their consistent moral support, encouragement and patience during my long absence from them while I was at McGill.

Finally I would like to express my sincere thanks to the professors who are responsible for reviewing my dissertation. Since the reviewing of a PhD dissertation represents the sacrifice of countless hours, their efforts and time are much appreciated.

To all of you, thank you! I could not have made it this far without you.

## Table of Contents

	Page
Abstract.....	i
Résumé.....	ii
Acknowledgements.....	iii
Table of Contents .....	v
List of Tables.....	xii
List of Figure.....	xiv
Abbreviations and Acronyms.....	xvi
 Chapter 1: Introduction.....	 1
1.1 Introduction to the General Context of the Research .....	1
1.2 Statement of the Research Problem.....	2
1.3 Organization of the Study.....	4
1.4 Purpose of the Study.....	5
1.5 Objectives of the Study and Research Questions .....	5
1.6 Significance of and Rationale for the Study .....	6
1.7 Working List of Definitions of Terms.....	8
 Chapter 2: Review of the Literature .....	 11
2.1 Introduction .....	11
2.2 Overview of the Theoretical Advances in Washback Research.....	11
2.2.1 <i>Research on Washback in Language Education and General Education</i> ....	12
2.2.2 <i>Washback Studies in the Chinese Context</i> .....	23
2.2.3 <i>The Teacher Factor Involved in Washback</i> .....	26
2.2.3.1 <i>Recurring Themes in Washback</i> .....	26
2.2.3.2 <i>Limitations of the Research Results on Washback</i> .....	33
2.3 Overview of the Research Methodologies Employed in Washback Research.....	33
2.3.1 <i>The Evolution of Washback Research</i> .....	34
2.3.2 <i>Design Characteristics and Approaches to Washback Research</i> .....	36
2.3.2.1 <i>Characteristics in Research Design of Washback Research</i> .....	37
2.3.2.1.1 <i>Mixed methods design.</i> .....	37
2.3.2.1.2 <i>Methods in studying washback.</i> .....	39
2.3.2.2 <i>Existing Design Patterns in Washback Research</i> .....	41
2.3.3 <i>Positive Qualities in Current Washback Studies</i> .....	43
2.4 Insights from Other Research Areas.....	46
2.4.1 <i>Change and Innovation</i> .....	47

2.4.2 <i>Themes, Rules and Principles Common in Innovation</i> .....	48
2.4.3 <i>Basic Concepts</i> .....	51
2.4.3.1 <i>The ‘Teacher Factor’</i> .....	52
2.4.3.2 <i>Beliefs</i> .....	52
2.4.3.3 <i>Teacher Thinking</i> .....	52
2.4.3.4 <i>Knowledge or Knowledge Base</i> .....	53
2.4.3.5 <i>The Relationship Between Teacher Beliefs and Teacher Practices</i> ..	54
2.4.4 <i>Conceptual and Theoretical Framework Guiding this Study</i> .....	55
2.4.4.1 <i>Shulman’s (1987) Model of Teachers’ Knowledge</i> .....	55
2.4.4.2 <i>Woods’ (1996) Model (A Model of BAK Networks)</i> .....	56
2.4.4.3 <i>Theories Concerning Teacher Conceptual and Behavioral Change</i>	58
2.4.4.4 <i>Change in Teachers’ Pedagogical Beliefs</i> .....	60
2.4.5 <i>Implications of the Above Insights for Washbach Research</i> .....	64
Chapter 3: The Educational, Sociocultural, and Historical Context of the Study .....	67
3.1 Introduction .....	67
3.2 Glimpse of Chinese CE Teaching and Testing.....	67
3.2.1 <i>“College English” Course – Stages of Development</i> .....	67
3.2.2 <i>Curriculum</i> .....	68
3.2.3 <i>The CET</i> .....	69
3.2.4 <i>Relationship between the CEC and the CET</i> .....	71
3.2.5 <i>Earlier Reforms in the CET</i> .....	71
3.3 Opposing Views on the CEC and the CET .....	72
3.4 Current Reforms .....	74
3.4.1 <i>The CECR</i> .....	74
3.4.2 <i>Revised CET</i> .....	75
3.4.3 <i>Textbooks</i> .....	77
3.5 Cultural Roots.....	77
3.6 Chinese Educational Context.....	78
3.6.1 <i>Exam-driven</i> .....	78
3.6.2 <i>Centralized</i> .....	78
3.6.3 <i>Knowledge-focused</i> .....	79
Chapter 4: Methodological Procedures of the Study.....	81
4.1 Introduction .....	81

---

4.2 Purpose of the study .....	81
4.3 Research Questions .....	81
4.4 Rationale for and Overview of the Methodology .....	82
4.5 Methodology.....	85
4.5.1 <i>Research Design</i> .....	86
4.5.2 <i>Participants and Research Sites</i> .....	88
4.5.2.1 <i>How I Selected My Participants</i> .....	88
4.5.2.2 <i>Participants of the Pilot Survey</i> .....	88
4.5.2.3 <i>Questionnaire Participants</i> .....	89
4.5.2.4 <i>Interview and Observation Participants</i> .....	89
4.5.2.5 <i>How I Selected My Research Sites for the Case-study</i> .....	89
4.5.3 <i>Instruments</i> .....	90
4.5.3.1 <i>The Review of Documents</i> .....	90
4.5.3.2 <i>Group Interview</i> .....	90
4.5.3.3 <i>Classroom Observation</i> .....	91
4.5.3.4 <i>In-depth Interview</i> .....	93
4.5.3.5 <i>Questionnaire</i> .....	93
4.5.3.5.1 <i>Constructs of the ‘teacher factor’ and ‘teacher practice</i> ...	95
4.5.3.5.2 <i>Piloting.</i> .....	96
4.5.4 <i>Data Collection Procedure</i> .....	97
4.5.5 <i>Ethical Considerations</i> .....	98
4.5.6 <i>Analysis of Collected Data</i> .....	98
4.5.6.1 <i>Analysis of the Data from Group Interviews and Case Studies</i> .....	99
4.5.6.1.1 <i>Organizing the data.</i> .....	100
4.5.6.1.2 <i>Developing theory.</i> .....	100
4.5.6.1.3 <i>Reporting the outcomes.</i> .....	100
4.5.6.2 <i>Analysis of the Data from Classroom Observations</i> .....	100
4.5.6.3 <i>Analysis of Questionnaires</i> .....	101
4.5.6.3.1 <i>Computer softwares used.</i> .....	101
4.5.6.3.2 <i>Statistical procedures.</i> .....	101
4.5.6.3.3 <i>Descriptive statistics.</i> .....	102
4.5.6.3.4 <i>Inferential statistics.</i> .....	103
4.5.6.3.4.1 <i>Exploratory factor analysis.</i> .....	103
4.5.6.3.4.2 <i>Confirmatory factor analysis.</i> .....	104
4.5.6.3.4.3 <i>Structural equation modeling.</i> .....	104



---

Chapter 5: Presentation of the Findings .....	106
5.1 Introduction .....	106
5.2 Analysis of the Test, the CET .....	106
5.2.1 <i>Task Characteristics</i> .....	107
5.2.1.1 <i>Rubric</i> .....	107
5.2.1.2 <i>Input</i> .....	108
5.2.1.3 <i>The Nature of Language Input of the CET</i> .....	109
5.2.1.4 <i>Expected Response</i> .....	110
5.2.1.5 <i>Facets of the Relationship Between Input and Response</i> .....	111
5.3 Analysis of the Teaching Materials .....	113
5.4 Results from Interviews and Classroom Observations.....	114
5.4.1 <i>Introduction</i> .....	114
5.4.2 <i>Background Information of the Research Sites</i> .....	115
5.4.3 <i>Profile of the Participants</i> .....	115
5.4.4 <i>Washback of the CET in the Case-Study Setting</i> .....	116
5.4.4.1 <i>How Teachers Perceive and React to</i> .....	116
<i>the CET Reform at School One (S1)</i>	
5.4.4.1.1 <i>Teachers' articulated beliefs about</i> .....	116
<i>the CET and its impact</i>	
5.4.4.1.2 <i>Teachers' curriculum knowledge</i> .....	118
5.4.4.1.3 <i>Observational evidence of the CET impact</i> .....	119
5.4.4.1.4 <i>Teachers' beliefs about teaching and</i> .....	119
<i>classroom scenarios of how they teach</i>	
T1 .....	119
T2 .....	120
5.4.4.2 <i>How Teachers Perceive and React to</i> .....	122
<i>the CET Reform at School Two (S2)</i>	
5.4.4.2.1 <i>Teachers' articulated beliefs about</i> .....	122
<i>the CET and its impact</i>	
5.4.4.2.2 <i>Teachers' curriculum knowledge</i> .....	125
5.4.4.2.3 <i>Observational evidence of the CET impact</i> .....	126
5.4.4.2.4 <i>Teachers' beliefs about teaching and</i>	
<i>classroom scenarios of how they teach</i> .....	127
T3 .....	127
T4 .....	128
5.4.4.3 <i>How Teachers Perceive and React to</i> .....	130
<i>the CET Reform at School Three (S3)</i>	

---

5.4.4.3.1 <i>Teachers' articulated beliefs about the CET and its impact</i>	130
5.4.4.3.2 <i>Teachers' curriculum knowledge</i>	134
5.4.4.3.3 <i>Observational evidence of the CET impact</i>	134
5.4.4.3.4 <i>Teachers' beliefs about teaching and classroom scenarios of how they teach</i>	135
T5	135
T6	136
5.4.5 <i>Overview of the Washback Evidence from Case Studies</i>	137
5.4.5.1 <i>Summary of the Results from Case Studies</i>	137
5.4.5.2 <i>Observable Washback Evidence</i>	139
5.5 <i>Results from the Teacher Questionnaire</i>	145
5.5.1 <i>Introduction</i>	145
5.5.2 <i>Results from Descriptive Statistics</i>	145
5.5.2.1 <i>Participants' Background Information</i>	145
5.5.2.2 <i>Major Aspects of the 'Teacher Factor' and the 'Teacher Practice'</i>	146
5.5.2.2.1 <i>Various beliefs as reported by the teachers</i>	146
5.5.2.2.1.1 <i>Teachers' beliefs about the CET and its impact</i>	146
5.5.2.2.1.2 <i>The washback effects of the revised CET on their teaching and learning</i>	147
5.5.2.2.1.3 <i>Teachers' beliefs about teaching, learning, and how to teach</i>	148
5.5.2.2.1.3.1 <i>Beliefs about teaching and learning</i>	148
5.5.2.2.1.3.2 <i>Views on how to teach</i>	149
5.5.2.2.1.3.3 <i>How they think they are teaching (or: teacher practice)</i>	150
5.5.2.2.2 <i>Teachers' knowledge base</i>	152
5.5.2.2.3 <i>Teachers' past experience</i>	153
5.5.3 <i>Results from Inferential Statistics</i>	154
5.5.3.1 <i>Item-Level Exploratory Factor Analysis for the Questionnaire</i>	154
5.5.3.1.1 <i>Items of the 'teacher factor'</i>	158
5.5.3.1.1.1 <i>Distribution</i>	158
5.5.3.1.1.2 <i>Reliabilities</i>	158
5.5.3.1.1.3 <i>Results</i>	159
5.5.3.1.2 <i>Items of teacher practice</i>	163
5.5.3.1.2.1 <i>Distribution</i>	163
5.5.3.1.2.2 <i>Reliabilities</i>	164

5.5.3.1.2.3 Results.....	164
5.5.3.2 Results from Confirmatory Factor Analysis .....	165
5.5.3.2.1 The hypothesized model: Model 1. ....	165
5.5.3.2.2 The results for Model 1.....	166
5.5.3.2.3 The hypothesized model: Model 2. ....	168
5.5.3.2.4 The results for Model 2.....	169
5.5.3.3 Results from Structural Equation Modeling.....	171
5.5.3.3.1 The hypothesized model: Model 3. ....	171
5.5.3.3.2 The results for Model 3.....	172
5.5.3.3.3 The hypothesized model: Model 4. ....	175
5.5.3.3.4 The results for Model 4.....	176
Chapter 6: Discussion of Results.....	180
6.1 Introduction .....	180
6.2 Research question 1: <i>To what extent and in what form does washback.....</i>	181
<i>exist in China in terms of its effect on teacher beliefs</i>	
<i>(e.g., beliefs about the CET and its impact, beliefs</i>	
<i>about teaching and learning), and classroom behaviors</i>	
<i>(e.g., content and particularly teaching methodology)?</i>	
6.2.1 Disjunction Between the Washback Rhetoric and Classroom Reality .....	181
6.2.2 Different Teacher, Different School, Different Effects of Washback .....	184
6.2.3 Negative Views on Tests and Innovation.....	186
6.3 Research question 2: <i>How is the ‘teacher factor’ manifested in such.....</i>	191
<i>a washback effect? What aspects of the ‘teacher factor’</i>	
<i>(e.g., teacher beliefs, teacher knowledge, teacher</i>	
<i>proficiency, teacher past experiences) contribute to</i>	
<i>the way that the courses are delivered?</i>	
6.3.1 Conflicting Views and Contradictions Related to .....	191
<i>the ‘Teacher Factor’ in Washback</i>	
6.3.1.1 Contradictions in Teachers’ Articulated Beliefs .....	191
<i>on Different Occasions</i>	
6.3.1.2 Contradiction Between How Teachers’ Perceive.....	192
<i>They Are Teaching and the Way They Are Teaching</i>	
6.3.1.3 Contradiction between the Teachers’ Perceived.....	193
<i>Teaching Approach and the Approach They Employ</i>	
6.3.2 Different Aspects of the Teacher Factor and Their Interrelationships .....	197
6.3.2.1 Teacher Views on Teaching and Learning .....	198
6.3.2.2 Teachers’ Knowledge .....	201
6.3.2.3 Teacher Proficiency.....	203

6.3.2.4 <i>Teacher Beliefs, Knowledge, Experience and Their Interrelationships</i> .....	205
6.4 Other Factors in the Washback Phenomenon .....	207
6.4.1 <i>Learner Factor</i> .....	207
6.5 Principal Findings .....	210
6.6 Further Analysis of the Findings .....	212
—Washback Nested in Complex Webs of Factors	
 Chapter 7: Conclusion .....	215
7.1 Introduction .....	215
7.2 Major Findings .....	215
7.3 Implications of the Study .....	218
7.4 Recommendations .....	223
7.5 Suggestions for Future Research .....	226
7.6 Limitations of the Study .....	228
7.7 Strengths and Contributions of the Study .....	228
7.8 Concluding Remarks .....	229
 References .....	231

## List of Appendices

Appendix A : English Version of the CECR
Appendix B: Announcement of Postponement of the Administration of the New CET by the NCETC (October, 21, 2005)
Appendix C: A Sample of the Revised CET
Appendix D: An Interview Protocol
Appendix E: Observation Schedule
Appendix F: The Questionnaire
Appendix G: Informed Consent Form

## List of Tables

Table 1 <i>Summary of Major Washback Studies</i> .....	35
Table 2 <i>Comparison between Transmission Teaching and Experiential Teaching</i> .....	63
Table 3 <i>Three Versions of CEC'</i> .....	69
Table 4 <i>Format and Structures of the CET</i> .....	76
Table 5 <i>Number of Hours of Observation</i> .....	91
Table 6 <i>Taxonomy of the 'Teacher Factor' Questionnaire</i> .....	95
Table 7 <i>Taxonomy of the 'Teacher Practice' Questionnaire</i> .....	96
Table 8 <i>Data Collection Procedures</i> .....	98
Table 9 <i>Procedures of Data Analysis</i> .....	105
Table 10 <i>A description of the structure of the revised CET</i> .....	108
Table 11 <i>Course-books Used by the Observed Teachers in Class</i> .....	113
Table 12 <i>General Characteristics of the Case-study Participants</i> .....	116
Table 13 <i>Washback effects of the revised CET based on T1 and T2's accounts</i> .....	116
Table 14 <i>Washback effects of the revised CET based on T3 and T4's accounts</i> .....	125
Table 15 <i>Percentage of the class time devoted to aural aspects of language</i> .....	126
by T3 and T4 (%)	
Table 16 <i>Percentage of the class time devoted to fast reading practice</i> .....	126
by T3 and T4 (%)	
Table 17 <i>Washback effects of the revised CET based on T6's accounts</i> .....	134
Table 18 <i>Percentage of the class time devoted to</i> .....	134
aural aspects of language by T5 and T6 (%)	
Table 19 <i>Percentage of the class time devoted to</i> .....	140
aural aspects of language by round (%)	
Table 20 <i>Percentage of the class time devoted to</i> .....	140
fast reading practice by round (%)	
Table 21 <i>Percentage of the class time devoted to</i> .....	141
reading practice by round (%)	
Table 22 <i>Percentage of the class time devoted to language forms by round (%)</i> .....	142
Table 23 <i>Average percentage of the class time respectively devoted to</i> .....	143
language forms, aural aspects of language, reading practice (%)	
Table 24 <i>Educational background of the survey participants in percentages (%)</i> .....	146
Table 25 <i>Descriptive Statistics 1</i> .....	147
Table 26 <i>Descriptive Statistics 2</i> .....	148

Table 27 <i>Descriptive Statistics 3</i> .....	152
Table 28 <i>Procedures of Inferential Analyses</i> .....	164
Table 29 <i>Descriptive Statistics –</i> .....	158
<i>Distribution of the various aspects of the ‘teacher factor’</i>	
Table 30 <i>Reliability estimates for the ‘teacher factor’</i> .....	159
Table 31 <i>KMO and Bartlett's Test</i> .....	15960
Table 32 <i>Eigenvalues and Percentage of Variance Explained for Each Factor</i> .....	160
Table 33 <i>Results of EFA for the ‘Teacher Factor’: Varimax Rotation</i> .....	161
<i>Extraction Method: Principal Component Analysis</i>	
Table 34 <i>Rotation Sums of Squared Loadings and Internal</i> .....	162
<i>Consistency Estimates for Each Factor</i>	
Table 35 <i>Factor Correlation Matrix the ‘Teacher Factor’</i> .....	162
Table 36 <i>Composites for the ‘Teacher Factor’ Variables (13 Items)</i> .....	163
Table 37 <i>Descriptive Statistics –</i> .....	163
<i>Distribution of the ‘teacher practice’ variable</i>	
Table 38 <i>Reliability Estimates for ‘Teacher Practice’</i> .....	164
Table 39 <i>Results of EFA for Teacher Practice Variables: Oblimin Rotation</i> .....	164
Table 40 <i>Inter-Item Correlation Matrix for Teacher Practice</i> .....	165
Table 41 <i>Results for the Hypothesized 1-Factor Model of ‘Teacher Practice’:</i> .....	167
Table 42 <i>Parameter Estimates for Teacher Practice: Model 1</i> .....	167
Table 43 <i>Results for the Hypothesized 5-Factor Model of the ‘Teacher Factor’</i> .....	169
Table 44 <i>Parameter Estimates for Teacher Practice: Model 2</i> .....	170
Table 45 <i>Results for the Hypothesized Model of the Relationship Between</i> .....	173
<i>the ‘Teacher Factor’ and Teacher Practice: Goodness of Fit Summary for Model 3</i>	
Table 46 <i>Parameter Estimates for the Relationship(s) Between</i> .....	173
<i>‘Teacher Factor’ and Teacher Practice: Model 3</i>	
Table 47 <i>Results for the Hypothesized Model of the Relationship Between</i> .....	177
<i>the Assumed Washback and Some Aspects of the ‘Teacher Factor’:</i>	
<i>Goodness of Fit Summary for Model 4</i>	
Table 48 <i>Washback Effects of the Revised CET Based on the T1 and T6’s Accounts</i> .....	192
Table 49 <i>Comparison Between Participants’ Survey Responses and Observation Data</i> .....	193
<i>in terms of Percentage of the Class Time Devoted to Language Forms (%)</i>	
Table 50 <i>Comparison between What the Participants Claim(x) and</i> .....	204
<i>What They Do (<math>\sqrt{\phantom{x}}</math>) in terms of Their Medium of Instruction</i>	

## List of Figures

<i>Figure 1</i> Washback research design by Watanabe (1996a, 1996b, 2004b).	41
<i>Figure 2</i> Washback research design by Turner (2002, 2005, 2008).	42
<i>Figure 3</i> The CET score scale.	70
<i>Figure 4</i> Hypothesized washback of the CET on teacher classroom practice	82
<i>Figure 5</i> Visual diagram of the research design of the study.	87
<i>Figure 6</i> Hypothesized construct of the ‘teacher factor’	96
<i>Figure 7</i> Hypothesized construct of ‘teacher practice’	96
<i>Figure 8</i> A flow chart of statistical procedures employed in the current study	102
<i>Figure 9</i> Bar chart of time devoted to aural aspects of language across the observation period (Round 1–3)	140
<i>Figure 10</i> Bar chart of time devoted to reading practice across the observation period (Round 1–3)	141
<i>Figure 11</i> Bar chart of time devoted to language forms across the observation period (Round 1–3)	142
<i>Figure 12</i> Time devoted to language form, aural aspects of language, reading practice	143
<i>Figure 13</i> Observed washback effects of the CET on teacher classroom practice	144
<i>Figure 14</i> The percentage of teacher-talk time in the classroom.	150
<i>Figure 15</i> The percentage of student-talk time in the classroom.	150
<i>Figure 16</i> The percentage of time spent on vocabulary	151
<i>Figure 17</i> Histogram of pedagogical knowledge.	153
<i>Figure 18</i> Scree plot of eigenvalues for each factor	160
<i>Figure 19</i> Breakdown of variance for factor analysis	162
<i>Figure 20</i> Hypothesized 1-factor model of ‘teacher practice’: Model 1	166
<i>Figure 21</i> The model of ‘teacher practice’	168
<i>Figure 22</i> Hypothesized 5-factor model of the ‘teacher factor’: Model 2	168
<i>Figure 23</i> The model of the relationships between various aspects of the ‘teacher factor’	170
<i>Figure 24</i> Hypothesized model of the relationship between the ‘teacher factor’ and teacher practice: Model 3	172
<i>Figure 25</i> The model of the relationship between the ‘teacher factor’ and	173
<i>Figure 26</i> Hypothesized model of the relationship between the assumed washback and some aspects of the ‘teacher factor’: Model 4	176
<i>Figure 27</i> The model of the relationship between the assumed washback and some aspects of the ‘teacher factor’: Model 4	177

<i>Figure 28 Teachers' perception of the CET impact on teaching .....</i>	<i>188</i>
---	------------



## ABBREVIATIONS AND ACRONYMS

AGFI	Adjusted Goodness of Fit Index
AMOS	Analysis of MOment Structures
BAK	Beliefs, Assumptions, Knowledge
BKE	Beliefs, Assumptions, Knowledge, Experience
BI	Beliefs of the Impact of the CET on Teaching
BTL	Beliefs of Teaching and Learning
BW	Beliefs of Washback
CD	Compound dictation
CE	College English
CEC	College English Curriculum
CECR	College English Curriculum Requirements
CET	College English Test
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CK	Curriculum Knowledge
CLT	Communicative Language Teaching
EFA	Exploratory Factor Analysis
EFL	English as a Foreign Language
ELT	English Language Teaching
EFL	English as a Foreign Language
ESL	English as a Second Language
GFI	Goodness of Fit Index
HKCEE	Hong Kong Certificate of Education Examination in English
IELTS	International English Language Testing System
KMO	Kaiser-Mayer-Olkin
L2	Second Language
M	Minute
MDI	Measurement-Driven Instruction
MC	Multiple Choice
ME	Chinese Ministry of Education
RMR	Root Mean Square Residual
ML	Maximum Likelihood
MM	Mixed Methods
MMR	Mixed Methods Research
NCETC	National CET Testing Committee
NFI	Normed Fit Index
PK	Pedagogical Knowledge
QUAL	Qualitative
QUAN	Quantitative
R	Round
R <sup>2</sup>	Squared Multiple Correlation Coefficients
RMSEA	Root Mean Square Error of Approximation
S	School
SD	Standard Deviation
SEM	Structural Equation Modeling
SLA	Second Language Acquisition

T	Teacher
TE	Teacher Experience
TOEFL	Test of English as a Foreign Language

## **Chapter 1: Introduction**

### **1.1 Introduction to the General Context of the Research**

Current washback research suggests that the influence of testing on teaching and learning, known as washback, is a highly intricate rather than a monolithic phenomenon. Despite the complex dimensions of washback, there is still a commonly held assumption that changing test formats will lead to enhanced instruction on the part of teachers. Consequently, in the field of language education, testing is often used as one of the most widely favored instruments to innovate teaching. The argument underlying this assumption is that tests may act as levers for instructional change (Pearson, 1988).

Guided by the same assumption, the Chinese Ministry of Education (ME) has launched a new round of massive reforms since the beginning of 2004. As part of the reform process, a revised version of the College English Test (CET) (a national standardized test sponsored by the Higher Education Department of the Chinese Ministry of Education and administered by the National CET Testing Committee to undergraduates of non-English majors in China) was announced to be piloted in June, 2006. Along with this testing reform, a series of other reforms were carried out both in curriculum and textbooks. The goal of the reforms was to bring about a shift of teaching and learning from its previous focus on linguistic knowledge to one on language use (Wu, 2005; CET News, 2005).

However, research into washback to date suggests that changing an exam alone does not result in the desired changes in teachers' instructional practices (Wall & Alderson, 1993; Burrows, 2004; Chapman & Snyder, 2000; Cheng, 1999; Cheng & Qi, 2006; Qi, 2007). Wall and Alderson (1993) accounted for the absence of such direct linkage between what an examination asks and what a teacher does by saying that the washback effect is produced by teachers' perception rather than the tests themselves. Later studies (Burrows, 2004; Chapman & Snyder, 2000) lent support to this view. According to Wall and Alderson (1993), no matter whether the washback is negative or positive, the quality of the washback effect is independent of the quality of a test. Nonetheless, recent research (Davison, 2008; Tavares & Hamp-Lyons, 2008; Urmston & Fang, 2008; Turner, 2008, 2009; Muñoz & Álvarez, 2010) reports that positive washback may occur if ongoing training and constant guidance and support are provided to teachers on assessment and instructional practices over time. Despite the

different connotations of washback in different studies, there is almost a general consensus in the field that among the number of factors that contribute to the washback phenomenon, the ‘teacher factor’ is crucial in mediating the process of washback (Alderson, 2001, 2004; Alderson & Hamp-Lyons, 1996; Andrews, 2004; Burrow, 2004; Cheng, 1999, 2004; Hayes & Read, 2004; Qi, 2004, 2007; Tan, 2008; Turner, 2008, 2009; Wall, 1999; Wall & Alderson, 1993; Wang, 2008; Watanabe, 1996b, 2004b).

It is interesting to note that the crucial role of the ‘teacher factor’ in educational change is not solely emphasized in washback research. It is stressed in research in language education and general education as well (e.g., Borg, 2006; Brumfit, 1991; Freeman & Freeman, 1992; Fullan, 1982, 1993, 2001; Hargreaves et al., 2001; Kennedy, 1987, 1988; Richards, 2008; Richards & Lockhart, 1994; Richards & Farrell, 2005; Markee, 1997; Sato & Kleinsasser, 2004; Solomon, 2002; Stobart, 2003; White, 1987; Woods, 1996, 2003, 2006).

There is obvious agreement among these researchers that teachers are those on whom the success of educational change largely depends. For example, White (1987) views them as one indispensable condition for change. According to English (2000), the classroom teacher can select and teach almost any curriculum he or she decides is appropriate, irrespective of reforms, innovations, and public examinations.

Overall, both the washback and educational research seem to suggest that unless the ‘teacher factor’ (central attributes that influence teacher decision making and behavior in the classroom – i.e., teacher beliefs, knowledge about teaching, learning, language teaching and testing, and experience) is explored, no explanation produced will help account for the washback phenomenon. Nevertheless, although the ‘teacher factor’ is unanimously believed to be a crucial contributing factor in generating washback, its overall picture has not been well documented. In this regard, there has emerged a clear need for research on this factor. There is a need to examine the perceptions or beliefs teachers are likely to bring with them into their English classes and investigate how they influence teachers’ instructional practices. Meanwhile it is also essential to investigate the pedagogical as well as the social and personal complexities influencing teachers’ beliefs and interpretations.

## 1.2 Statement of the Research Problem

This research addresses problems related to the ‘teacher factor’ embedded in the

washback phenomenon in the Chinese university context. It is well known that in China where the educational system is typically test-driven, exams are of exaggerated importance. For years, the two levels of the CET (see details below), Band Four and Band Six (CET4 and CET6) have been criticized for their adverse effects (so-called “negative washback”) on College English teaching and learning (Feng, 2000; Gu & Liu, 2005; Han, 2002; Han et al., 2004; Ma & Jin, 2000; Qian, 2003; Shu, 2004). Chinese English Language Teaching (ELT) research indicates that although changes in ELT curriculum have occurred since 1989 and initial efforts are also reported to have been made to improve the existing test designs (e.g., with a gradual decrease in its linguistic knowledge component), the effects of ELT innovation were minimal and Chinese English as a Foreign Language (EFL) instruction still remained entrenched in traditional practices (Shu, 2004). To be specific, the most dominant methodology employed by EFL teachers is the traditional structural approach (or grammar-translation approach) and the focus of their instruction has never shifted from formal linguistic knowledge to language use (Dai, 2001; Shu, 2004). Previous findings reveal that this approach could exert a negative impact on classroom teaching at the expense of students’ language skills (Brown, 1994; Savignon, 1983). Earlier research evidence also shows that the traditional approach, with its emphasis on close analysis of syntax and discussion of word meanings, can hardly ensure the students’ successful performance on tests (Alderson & Hamp-Lyons, 1996). However, despite its unfavorable reputation, the approach continues to exist and prevail. Consequently, while many Chinese university students have passed this national test, the CET, their actual English communicative competence still remains at a low level (Cai, 2002, 2003, 2005; Dai, 2001; Gu & Liu, 2005; Guo, 2003; Niu, 2001). One of the key reasons put forth by tertiary-level teachers to account for their adherence to the approach is the constraint of the CET (Chen, 1999; Han et al., 2004; Ma & Jin, 2000; Qian, 2003; Tarnopolsky et al., 2001). According to teachers, it is the test that impedes them from their endeavors to innovate their traditional teaching methods. The view is widely shared among Chinese EFL researchers, practitioners and well-known linguists (Cen, 1999; Gu & Liu, 2005; Han, 2002; Han et al., 2004; Ma & Jin, 2000; Qian, 2003; Liu & Dai, 2003; Shu, 2004). As a result, the above-mentioned reforms were introduced in response to the fierce criticisms leveled at the CET.

But what deserves scrutiny is whether there is a direct causal relationship between the test and its alleged ‘negative test impact’. Above all, most of the criticisms

concerning the influence of the CET on College English teaching and learning lack empirical evidence. Under such circumstances, whether exams may impinge on or facilitate methodology innovation needs to be examined empirically.

### 1.3 Organization of the Study

This dissertation consists of seven chapters. Each chapter covers a central theme. Its organization is as follows.

Chapter 1 offers an introduction to the context of the whole study by giving a brief account of the underlying problems that generated this research study, research problem(s), purpose and objectives, and research questions of the study as well as the significance of and rationale for the study.

Chapter 2 provides a detailed overview of the research studies that have helped shape the present study. It is mainly made up of two parts. It starts with an overview of the theoretical and methodological advances pertaining to washback research over the last two decades to identify the features related to the ‘teacher factor’ and then turns to review pertinent literature in other research areas (e.g., language education, general education, psychology and other innovation research) for insights it can offer with respect to test impact on teaching methodology. Following this, it discusses what the multiple sources reviewed imply for washback research. These two parts, taken together, constitute a general framework for looking at the complex features of the ‘teacher factor’ in innovation and washback.

Chapter 3 provides an examination of the overall local educational, sociocultural and historical context in which this study is situated. Specifically, it gives a general description of the development of Chinese CE teaching and the CET over the past three decades as well as the pertinent reforms carried out nationwide, and discusses their existing problems in relation to the Chinese educational, sociocultural and historical context. By so doing, we may gain a better understanding of the overall local educational context where test impact is constructed or nurtured.

Chapter 4 describes the research design used for this study, including sampling procedures, participants, instruments as well as procedures of data collection and data analysis.

Chapter 5 presents the data yielded by document analysis, informant interviews, classroom observations, and survey questionnaires.

Chapter 6 includes a discussion of the major findings related to the research

questions of this study by synthesizing, integrating, and triangulating the results from different data sets. The findings in this chapter are organized and outlined based on the themes and patterns that have emerged from analyses presented in Chapter 5.

Building on earlier chapters, Chapter 7 summarizes the major findings and elaborates on their implications for the fields of language testing, the overarching field of applied linguistics and general education as well. Also, it proposes some possible directions and recommendations for future research. Finally, limitations (and contributions) regarding technical difficulties as well as the overall scope of the study are addressed.

#### 1.4 Purpose of the Study

The purpose of this study is threefold: (1) to investigate the mainstream attitudes Chinese EFL teachers maintain toward the CET and its impact; (2) to find out what aspects pertinent to the ‘teacher factor’ (e.g., teacher beliefs, teacher knowledge, experience, etc.) present the major barrier to the implementation of educational change so as to conclude whether tests constitute a constraint on ELT methodology innovation in China; and (3) to examine how the major components of the ‘teacher factor’ involved in the current study (e.g., teachers’ beliefs about the CET and its impact, teachers’ beliefs about teaching and learning, teachers’ knowledge base, etc.) are interrelated.

#### 1.5 Objectives of the Study and Research Questions

There are two main research objectives for this research. The first objective is to identify the actual impact of the CET on College English (CE) instruction in China (e.g., whether a link can be established between the CET and what CE teachers practice in their classrooms). The second objective is to ascertain the role that the ‘teacher factor’ plays in washback.

One global question and two corollary questions serve as a tentative focus for this study.

The global research question guiding the study is: What role does the ‘teacher factor’ play in washback in the Chinese university context?

To answer this global question, the following two specific questions will be addressed:

1. To what extent and in what form does washback exist in China in terms of its

effect on teacher beliefs (e.g., beliefs about the CET and its impact, beliefs about teaching and learning), and classroom behaviors (e.g., content and particularly teaching methodology)?

2. How is the ‘teacher factor’ manifested in such a washback effect? What aspects of the ‘teacher factor’ (e.g., beliefs, knowledge, past experiences) contribute to the way that teachers interpret and react to washback?

### 1.6 Significance of and Rationale for the Study

The need to study the role of the ‘teacher factor’ in generating washback has been elucidated by a number of researchers (Alderson, 2001, 2004; Alderson & Hamp-Lyons, 1996; Andrews, 2004; Burrow, 2004; Cheng, 2004; Hayes & Read, 2004; Qi, 2004, 2007; Muñoz & Álvarez, 2010; Tan, 2008; Turner, 2008, 2009; Wall, 1999; Wall & Alderson, 1993; Wang, 2008; Watanabe, 1996a, 1996b, 2004b). According to these researchers, it is not sufficient for washback studies to dwell on testing in isolation from the essential role that the ‘teacher factor’ plays in the innovation process as well as the teaching context in which these studies take place. For example, Alderson and Hamp-Lyons (1996) have conducted a study into the impact of the Test of English as a Foreign Language (TOEFL) on teaching. Through this study, they have realized that there is not much that test developers can do to influence how teachers might teach. It is also through this study that they have become aware of the central role of this ‘teacher factor’ in washback. They end the study with a call for research attention to be paid to the reasons why teachers teach the way they do. Alderson (2004) takes this idea a step further and claims that it is at least as much the teacher who brings about washback, be it positive or negative, as it is the test. He claims succinctly that washback is brought about by people in classrooms, but not by test developers. Alderson and Hamp-Lyon’s (1996) concern has been echoed and emphasized by Burrows (2004), Watanabe (1996a, 2004b) and Hayes and Read (2004). Watanabe (1996) argues that teacher factors may outweigh the influence of an examination in inducing washback. In the same vein, Hayes and Read (2004) and Burrows (2004) also underscore the importance of this factor. Hayes and Read (2004) attribute the difference in the ways the courses are delivered to the functioning role of this factor. Nevertheless, despite the common concern shared among these researchers about the ‘teacher factor’, no further efforts are reported to have been exerted to explore the shape it takes as well as the way it functions in the process of washback. It



seems that such insufficient knowledge of the ‘teacher factor’ has delayed the process of our understanding of the nature of washback. Therefore, a more systematic study of this factor is called for. As Alderson (2004) has pointed out, this is a promising area for further research. In this respect, this study will not only help fill the gap in this research area, but the findings may also help shed light on further research questions in this area.

In addition, as mentioned above, the crucial role of the teacher “variable” in educational change has been documented in general education as well (e.g., Borg, 2006; Brumfit, 1991; Freeman & Freeman, 1992; Fullan, 1982, 1993; Hargreaves et al., 2001; Kennedy, 1987, 1988; Richards, 2008; Richards and Lockhart, 1994; Markee, 1997; White, 1987; Wood, 1996, 2003, 2006). The common thread in these studies is the belief that, to a large extent, teachers play a key role in determining what the classroom will be like. In these studies, teachers are perceived as the chief implementers of reforms and viewed as making their own choices in terms of what to perform in the classroom. Although many other individuals also have a stake in the innovation process, teachers are considered to be key players in all instructional innovations (Fullan, 1991; Hargreaves et al., 2001; Markee, 1997; White, 1987). According to English (2000), the classroom teacher can select and teach almost any curriculum he or she decides is appropriate, irrespective of reforms, innovations, and public examinations. What English has asserted seems to justify Markee’s (1997) statement that teachers always “retain ultimate responsibility for what happens in their classroom” (p. 44). In this regard, as illustrated by researchers on innovation, promoting change is not just a matter of technique. Rather, it must seek to “gain teachers’ understanding and support” (Markee, 1997, p. 13). Obviously, more research is needed before their support is secured.

In sum, the rationale underlying this study is significant. An exploratory study on how the teacher role operates in the innovation process is of both immense theoretical and practical value. Although the issue has been touched upon across a range of subjects, it has not been dealt with in depth. Questions with respect to whether tests constitute a facilitator or constraint on instructional innovation, to what degree testing affects teachers’ beliefs about teaching, how various teacher beliefs function, and how and why washback affects some teachers but not others still remain unanswered. In this regard, the theoretical significance of this study lies in that the results of this study will not only add, into the existing washback literature, new insights related to

the above issues, but will also contribute to the ongoing call in the field (Alderson, 2004; Andrews et al., 2002; Turner, 2009) for intensive investigation of the role of the ‘teacher factor’ in washback. Furthermore, the findings of this study may also have important implications for other types of educational research.

### *Practical Rationale*

In addition, the pedagogical and practical value of this study also lies in its potential to provide information on how we can help both language teachers and teachers in general develop positive attitudes toward tests. It is hoped that the insights to be gained from this study can help Chinese EFL teachers reflect on their own practice and trigger a deeper understanding of their teaching process. Such a reflection may not only help teachers increase their awareness of the real problems underlying their teaching practice, but also help them change their perceptions of tests. With a heightened awareness and more accurate perception of the nature of the washback effect and the conditions under which it operates, teachers may question their own roles as teachers and re-envision their ways of teaching. Thus they may be in a position to experiment with alternative teaching approaches to improve their teaching. As a result, teachers may eventually alleviate negative washback (negative consequence of tests on classroom activity) and replace it by positive washback. In view of the possibility that the study may help to direct Chinese EFL teachers towards more effective communicative language teaching, this study is of practical significance.

### 1.7 Working List of Definitions of Terms

Before looking in detail at the research findings reported by other washback researchers and the research designs adopted in their studies, it is important to clarify some basic acronyms and terms. They are defined for the purposes of this study and should be interpreted as such within this dissertation context.

**CE:** “College English” (CE) is a required course for all tertiary-level non-English majors across China.

**The CECR:** The College English Curriculum Requirements (CECR) is a revised national curriculum for CE teaching and learning officially announced and published in 2004.

**The CET:** The CET comprises two levels: CET 4 and CET 6. The CET 4 is

mandatory for non-English majors of a great number of universities in China. The CET 6 represents a higher level of English proficiency, however it is not compulsory for all university students. Only candidates who have passed the CET 4 are permitted to sit for the CET 6. As the CET 6 bears many similarities to the CET 4 and above all, it is not as compulsory a test as the CET 4, this dissertation will only focus on the CET 4. Students are required to pass the CET 4 to get their BA or BSc.

**CLT:** In this research, the definition of Communicative Language Teaching (CLT) correlates with that provided by western ELT theorists (Breen & Candlin, 1980; Ellis, 1990; Savignon, 1991, 2003, 2005; Larsen-Freeman, 1986; Stern, 1992; Brown, 1994; Richards & Rodgers, 2001; Wesche & Skehan, 2002). It refers to a teaching methodology or an approach that focuses primarily on communicative competence comprising both receptive and productive skills (listening, reading, speaking and writing).

It is an approach that (1) is associated with learnercenteredness, (2) encourages learners to use the language to interact and to share real information, and (3) engages them in target language communication, either productive or receptive. This communication-based methodology known as CLT is often used to contrast with the structural approach (or grammar-translation approach) or the grammar-based approach and it represents a departure from the traditional pedagogical practices “before about 1975”.

**The Structural approach (or grammar-translation approach):** This approach focuses on grammatical structures and involves memorization of rules, mechanical manipulation of language forms (i.e., concentrating on the teaching of discrete points of language), introduction of lists of new vocabulary, and correction of the students’ pronunciation (Richards & Rodgers, 2001).

**Washback:** The definition of washback varies from study to study. For the purpose of this study, washback is conceptualized as the instructional changes induced by testing to teaching in terms of teacher beliefs and classroom behaviors. The effects of the test that are expected regarding this study involve both perceptual evidence (change in teacher attitudes and beliefs) and behavioral evidence (change in teaching content and teaching methods that correspond to the test reform).

**High-stakes exams:** High stakes tests refer to those tests whose results are used or misused to make important decisions that may have strong impact on test-takers

(Madaus, 1988). Compared to low-stakes exams, they are considered to have more power to modify teachers' behaviors.

**Task-based:** In Skehan's (1998) definition, there are four criteria for tasks: 1) meaning is primary, 2) it works toward a goal, 3) it is outcome-evaluated, and 4) it is related to the world outside the classroom.

In this chapter, first, the general context and research problem of this study were explained. Then, the organization of the dissertation was outlined and the research questions were articulated. Following that, the purpose and objectives of the study were addressed. Finally, the significance of and rationale for the study were presented, and some terms were clarified.

The next chapter clarifies basic concepts and explores theoretical and methodological advances pertaining to washback research. An extensive discussion of studies in other research areas that influence and shape the present study is also included in the following chapter.

## **Chapter 2: Review of the Literature**

### **2.1 Introduction**

This chapter is a literature review, the purpose of which is to gain insight into the complex dimensions of washback and illuminate the vital role that the ‘teacher factor’ plays in generating washback. It is comprised of three sections. The first section provides an overview of the theoretical advances in washback research over the past two decades. It starts with an examination of the concept of washback as well as the theoretical perspectives that underpin washback research in the literature of language education and general education. It then looks at the washback research conducted in the Chinese context and compares it to that done in other contexts to see what similarities they bear. After that, it examines significant washback research findings in search of patterns and themes related to this ‘teacher factor’. The second section explores pertinent methodological issues which have dominated washback research to date. The third section, however, goes outside washback research per se. Specifically, it draws on ideas from language education, general education, psychology and other innovation research to see whether insights can be gained into the patterns and themes that have recurred in washback research. Here, by reviewing different sources of literature, I endeavor to present the theoretical basis for claims made in washback studies and provide specific explanations for why the claims are made the way they are. This section also discusses how the multiple sources to be reviewed in this chapter relate to my own study.

The above reviews, taken together, constitute a general framework for looking at the research topic in this study.

### **2.2 Overview of the Theoretical Advances in Washback Research**

It is worth noting that language testing researchers have embraced the call from Alderson and Wall (1993) for more intensive research on washback, so that the last two decades have witnessed a substantial volume of research on this topic (Alderson & Hamp-Lyons, 1996; Cheng, 1997, 2004; Cheng & Qi, 2006; Frederiksen & Collins, 1989; Green, 2006, 2007; Muñoz & Álvarez, 2010; Qi, 2004, 2007; Saif, 2006; Shih, 2007; Shohamy, 1993; Shohamy et.,1996; Tan, 2008; Turner, 2001, 2005, 2008, 2009; Wall, 1996, 1999; Wall & Alderson, 1993; Wang, 2008; Wall & Horák, 2007; Watanabe, 1996a, 2004b). Also worth noting is that in recent years, researchers have

been making significant inroads into investigating this phenomenon in different social and educational contexts. As a result, the definition as well as the nature and scope of washback have been extensively discussed and a number of different perspectives have emerged in language testing and ELT research area.

### 2.2.1 *Research on Washback in Language Education and General Education*

To put the research findings in proper perspective, there is a need to make some references to the basic concept of washback.

The term, washback, generally defined as the influence of testing on teaching and learning, has evolved into a multifaceted concept (Turner, 2001). It is based on the notion that tests can drive teaching and learning, namely measurement-driven instruction (MDI) (Popham, 1987). Implicit in this notion is a view of curriculum alignment, a match between the content and format of the test and that of the curriculum (Shepard, 1991, 1993). Proponents of washback (Hughes, 1988, 1989; Messick, 1996; Pearson, 1988) advocate that such alignment can be achieved with the introduction of a new or revised examination. It is interesting to note that this alignment is also referred to as systemic validity by Frederiksen and Collins (1989), consequential validity by Messick (1994, 1996), and test impact by Bachman and Palmer (1996). Although different terms are introduced, they all refer to different facets of the same phenomenon – the influence of testing on teaching and learning. What I need to clarify is that in some studies, the term ‘washback’ is differentiated from the term ‘impact’ in that the meaning of ‘impact’ is more broad in terms of the scope of the effects than is encompassed by ‘washback’ (e.g., it encompasses more widespread societal influences in addition to local classroom influences) (Hamp-Lyons, 1997; Wall, 1997). However, in this dissertation, the two terms will be used interchangeably.

It is important to note that the term ‘washback’, though commonly used, may carry different connotations in different studies (e.g., Alderson & Hamp-Lyons, 1996; Alderson & Wall, 1993; Cheng, 1997; Muñoz & Álvarez, 2010; Shohamy et al., 1996; Turner, 2005, 2008, 2009; Watanabe, 1996a, 2004b). Although there is general agreement in the field as to the existence and importance of the washback phenomenon, there is considerable variety in opinions as to how washback functions (Bailey, 1996). The way it is perceived depends on the theoretical standpoint a researcher takes as well as the educational context that he or she is associated with. A

case in point is that no consensus can be reached in the language testing community as to whether washback effects are positive or negative. Some studies describe washback as negative, while others view it as a combination of positive, negative and/or neutral aspects. Moreover, since the testing research community has not come up with a unified conception of washback, there is considerable variation in researchers' interpretations of the concept. For example, both Alderson and Wall (1993), and Cheng and Curtis (2004) note that language tests are generally criticized for their negative influence on teaching – so-called “negative washback”. Typical of this critical position are Vernon (1956) and Davies (1968). They charge examinations with distorting the curriculum and making the educational experience narrow and uninteresting (see Wall, 2000, and Cheng & Curtis, 2004). According to this line of thinking, the practice of teaching to the test may lead to the narrowing of curriculum and instruction. More support for this position comes from Noble and Smith (1994a), Shohamy (1993), Smith (1991), and Widen et al. (1997). Shohamy (1993) is critical of MDI, for she assumes that in this way the function of tests would exceed their original design (i.e. to provide information on achievement). She argues that using tests to solve educational problems is a simplistic approach to a complex problem (Shohamy, 1993). She also contends that tests encourage learners to learn the language in a narrow sense rather than to learn ways of approaching language use. In Noble and Smith's (1994b) view, teaching test-taking skills and drilling on multiple-choice worksheets are likely to boost scores but unlikely to promote general understanding. The above views are widely shared among ELT practitioners at large. Obviously, the commonly perceived “negative washback” differs significantly from the one defined by proponents of positive washback.

Contrary to the above views, washback proponents (Messick, 1996; Pearson, 1988) hold that changing test formats will lead to enhanced instruction on the part of teachers. As Pearson (1988) puts it, tests may act as levers for instructional change. They place an emphasis on the instrumental value of external examinations for reforming curriculum and teaching. Based on their interpretation, good tests can induce “good” instructional practice, and vice versa. Consequently, in the field of language education, testing is often used as one of the most widely favored instruments to innovate teaching.

What should be noted is that these definitions are being discussed within an educational context and achievement/progress testing, where there is a curriculum to

follow and high stakes exams at some point (e.g., Alderson & Wall, 1993; Andrews, 2004; Messick, 1996; Turner, 2009; Wall, 2000; Watanabe, 1996a, 2004b). We should also note that there are other contexts like proficiency testing (TOEFL & IELTS) where an education system is not always implicated. Although in the latter contexts sometimes test-preparation classes are offered, they are not usually part of a regular school system.

Following an examination of the various definitions of positive or negative washback, Turner (2005, 2009) has developed her definition of the concept. In her definition, the washback effect of a high-stakes external exam will be positive if it represents the curriculum well in terms of its content and procedures. Her point is that positive washback effects can be achieved when the impact of a test can help a teacher change or align some instruction with general concepts represented in the test. She also illustrates that the test's washback effect will be negative if its content and procedures do not represent the curriculum well. The argument underlying this statement is that the teacher might ignore or abandon the curriculum to prepare the students for an unrelated test. Her interpretation helps clarify the misconceptions that have existed in the field of washback research. Among those that have shaped Turner's (2006) definition of washback are Bachman (1990) and Messick (1996). From the point of view of Bachman (1990), positive washback occurs when the exam used reflects the skills and content taught in the classroom. According to Messick (1996), tests should be designed in such a way that activities/methods involved in learning the language are identical to those involved in the test. It implies that positive washback is more likely to occur when a curriculum and test are highly matched. It is interesting to note that Brown and Hudson (1998) also expound on the definition of washback. Their definition is consistent with that of Turner's. In their view, washback effects can be either negative or positive. A positive washback effect occurs when the assessment procedures correspond to the course goals and objectives. An explicit example they used to illustrate their point is that if a program sets a series of communicative performance objectives and tests the students using performance assessments (e.g., role plays, interviews) and personal-response assessments (e.g., self-assessments, conferences), positive washback effect can be produced in favour of the communicative performance objectives. However, if the program ends up assessing the students with multiple-choice structure tests, a negative washback effect will be engineered. They seem to remind us that in addition to the alignment between



testing and curriculum, the task types a test utilizes are also a crucial factor in creating positive washback. Much along the same line, both McNamara (2000) and Prodrômou (1995) argue that to create positive washback, process-based assessments should be introduced. The reason provided by them is that performance-based assessments that require integrated content and skills can better reflect the goals of instruction.

From the standpoint of washback proponents, teachers' tendencies to teach to the test are viewed as an impetus rather than an impediment to instructional reform as long as the test represents the curriculum. To put it simply, teaching to the test is thus justified (Noble & Smith, 1994b).

Alderson and Wall (1993) focuses on a 'neutral' perspective. Drawing on the longitudinal observations they have conducted in the Sri Lankan Impact Study, Alderson and Wall (1993) conclude that no matter whether the washback is negative or positive, the quality of the washback effect is independent of the quality of a test. They claim that tests can be powerful determiners, both positively and negatively, of what happens in classrooms. This is taken to mean that bad tests do not necessarily bring bad effects and vice versa. Another major point made by them is that test impact can be positive or negative depending on how teachers perceive it and react to it. This point is supported by Hamp-Lyons (1997), who also argues that the washback effect can be either beneficial or harmful to the extent that it either promotes or impedes the accomplishment of educational goals held by teachers and learners.

However, despite the different connotations in different studies, tests continue to be used as a vehicle for curriculum innovation (Andrews, 2004). Examples of using tests as levers for change are found in a number of countries – Australia (Burrows, 2004), Canada (Turner, 2001, 2008, 2009; Saif, 2006), Colombia (Muñoz & Álvarez, 2010), China (Qi, 2007; Wang, 2008), England (Green, 2006, 2007), Hong Kong (Andrews, 1994; Andrews et al., 2002; Cheng, 1997, 2004; Davison, 2008; Tavares & Hamp-Lyons, 2008; Urmston & Fang, 2008), Israel (Shohamy et al., 1996), Japan (Watanabe, 1996a, 2004b, 2004b), Malaysia (Tan, 2008), New Zealand (Hayes & Read, 2004) and Sri Lanka (Alderson & Wall, 1993; Wall & Alderson, 1993; Wall, 1999). These studies have given an account of washback from different perspectives. In this thesis, I will cite them along the way.

Empirical studies on washback suggest that changing tests to innovate instructional practices can work in some settings and positive washback may occur via different routes. One potential recently noted and indicated in research (Davison, 2008; Muñoz

& Álvarez, 2010; Tavares & Hamp-Lyons, 2008; Turner, 2008, 2009; Urmston & Fang, 2008) is that positive washback may be fostered by providing ongoing training and constant guidance and support to teachers on assessment and instructional practices over time. It appears that teacher support and understanding may lead to knowledge upgrade and possibly an evolution in perception. The studies conducted in Canada (Turner, 2001, 2002, 2005, 2008, 2009) are significant in this respect, since they offer a generally positive example of how testing can be used to improve instruction. Turner's washback studies emerged from her earlier work which involved developing empirically based rating scales (speaking and writing) for provincial exams. Based on a washback project she has undertaken while tracking the innovations in the provincial English as a Second Language (ESL) Speaking exam in Quebec, she probes into how teachers respond to the testing innovation at the micro classroom level. Her studies (e.g., 2001, 2005, 2008, 2009) differ from others in that they highlight the role of teacher involvement in the process of test development. Unlike most of the research which describes teacher behaviour as 'unpredictable', an important finding from her study is that with sufficient training and guidance, the participating teachers were found to be open to the influence of test effects on their classroom behavior by seeking ways to change or align their instructional practices to correspond to the exam materials and methodology (e.g., they reacted favorably toward the new criteria for speaking ability and toward the new format of oral testing in a group discussion as opposed to the earlier format of a one-on-one interview with the teacher). The results of the study demonstrate that testing can affect not only teacher beliefs, but can also affect some teacher behavior or practice. Hence, the researcher concludes that a "good" test (e.g., the test adequately represents the curriculum) can lead to certain positive aspects of washback in some contexts. The study differs in one important way from much of the research that has been reported in the literature on washback. That is, not only has the researcher herself been involved in devising the rating scales for the exam, but she has also been involved in planning and managing the change process. Specifically, rather than waiting for the washback effect to occur itself, the Quebec Ministry of Education has conducted workshops and prepared materials for teachers prior to the exam to help sensitize them to the changes. The study seems to suggest that teacher training and guidance make a difference in why washback happens to some teachers, but not to others. Turner (2009) elaborates on her research saying that "when invited to engage and

participate in assessment development and procedures, teacher understanding and acceptance is increased; when provided with information and training prior to a new high stakes test, teachers integrate task characteristics into their teaching” (p.106). She seems to remind us that positive washback can be attained when the introduction of a test is accompanied by adequate teacher training and guidance. What she has claimed generates important insights into the washback phenomenon. The study provides us with a good example that test developers consulted with and provided necessary guidance for teachers, students, administrators throughout the test development process in order to maximize positive washback.

Another instance of positive washback worthy of citing is the research by Muñoz and Álvarez (2010). They look at the washback effect of an oral assessment system (OAS) on some areas of EFL teaching and learning. Their participants were non-native EFL teachers at a language center in a small private university in Colombia, South America. The data-collection instruments that were used include surveys and classroom observations. By inspecting the connection between curriculum objectives and instructional tasks and probing into teachers’ capacity to specify classroom objectives and conduct activities that directly targeted those objectives, this longitudinal study intended to investigate whether some of the principles that underlie the development of the OAS could generate positive washback. The research adopted a comparative study design. Two evenly divided groups of students (experimental group and comparison group) were compared using quantitative and qualitative research methods. A conclusion that was drawn by the researchers is that a degree of positive washback occurred, since improvements were observed in the teachers’ teaching and assessment practices and in students’ oral production. Nonetheless, the study is limited in that the researchers are not clear “if the instructional practices found are the consequence of teachers’ own conviction about the benefits of effective assessment procedures or are the product of reform initiatives by the Language Center” (p. 46). Despite its tentative conclusion, the study is significant in that it reinforces the standpoint of Turner (2008, 2009) that the constant guidance teachers receive during the innovation process may help them make appropriate use of the test.

Similar findings are found in the studies conducted in Hong Kong which also include working sessions with teachers (Davison, 2008; Tavares & Hamp-Lyons, 2008; Urmston & Fang, 2008). The common thread in these studies is the belief that teacher

involvement and support should be secured if positive washback is to be promoted. One implication of this is that teacher ongoing support may lead to knowledge evolution and possibly an evolution in perception. In other words, only when teachers achieve a good comprehension of the objectives, content and methodology of a test, will they be in a better position to change their perceptions and behaviors which conform to its innovation.

However, apart from these studies, the growing body of empirical research to date indicates that the extent to which washback promotes pedagogical change is limited. That is, changing an exam format alone does not result in direct and desired changes in teachers' instructional practices as originally intended.

Another study worthy of mention is Alderson and Wall (1993). In comparison to other studies in this domain, Alderson and Wall (1993) have observed and addressed the nature of washback effects more extensively. To highlight the complexity of the notion of 'washback', they put forward 15 Washback Hypotheses. The basic assumption of the Washback Hypotheses is that a test on its own makes all the difference. Their purpose of setting forth such hypotheses is to pinpoint the misleading definitions and assertions related to washback. A crucial result of their study is that the impact of tests has been demonstrated only on what teachers teach but not on how they teach. In their eyes, the 15 different versions of washback hypothesis are undoubtedly simplistic. The findings of the study have led them to conclude that the belief that 'good' tests will automatically have 'good' impact is oversimplified. In Wall and Alderson (1993), they reiterate that the exam is only one of the factors that "affect how innovations succeed or fail" (1993, p. 68). Wall (1996) has listed several different factors which are deemed as impeding tests from inducing positive effects. Among them are teacher resistance to change, teachers' lack of understanding of the exam, and deficiency in well- trained teachers, etc. Their study is of crucial value in that it has spurred a number of empirical studies along these lines.

Findings produced by later studies have been, for the most part, consistent with those of Alderson and Wall (1993). Cheng's (1997) study indicates that there is little evidence of substantial changes in teacher behavior. She drew on the new Hong Kong Certificate of Education Examination in English (HKCEE) to investigate how the effort to change this examination would bring about fundamental changes in classroom teaching in Hong Kong secondary schools. The Hong Kong Examinations Authority intended to create a positive washback effect through the innovation of the

examination. However, it is found that in contrast to the test authority's expectation, teacher talk under the new syllabus still dominated the majority of class time. The lessons observed were still highly teacher-controlled. The findings of the study reveal that the change in the public examination could, to a large extent, change the content of teaching, but neither the interaction pattern followed by the teachers nor the general pattern of their teaching approaches changed. Like Cheng (1997), Andrews' study (1995) was also situated in Hong Kong. It was associated with the addition of an oral component to the Hong Kong Use of English Examination. In his small-scale study, he used a survey to examine how two groups of participants (test designers and teachers) perceive the revised test. The findings indicate that the perceptions of examination designers do not overlap the perceptions of teachers. Apart from the content of teaching and time allocation to the oral aspects of language, there was not much evidence of fundamental shifts in teachers' classroom strategies.

Similarly, Andrews et al. (2002), Burrows (2004), Chapman and Snyder (2000), Cheng (2004), Qi (2004, 2005, 2007), Shohamy (1993), and Watanabe (1996a) provide added evidence that tests could only induce the surface level of washback – change in what teachers teach. It is interesting to note that there is amazing similarity in their findings. The results from these studies show that the change carried out in assessment is not sufficient to induce the change in instructional strategies. What is implied but not overtly stated in these studies is that when it comes to methodology – how teachers teach—factors other than tests come into play. Since researchers to date (Andrews et al., 2002; Chapman & Snyder, 2000; Cheng, 1997; Qi, 2004, 2007) claim that tests are not efficient tools to induce pedagogical changes, then, what are powerful constraints that prevent changes in teachers' instructional practices? Wall and Alderson (1993) account for the absence of such direct causal linkage between what an examination asks and what a teacher does by saying that the washback effect is produced by teachers' perceptions rather than tests themselves. They stress that washback research has to consider both teachers' attitudes and their comprehension of exams so that a deeper understanding can be achieved about the way they teach. Other studies (Andrews, 1994; Burrows, 2004; Chapman & Snyder, 2000; Green, 2006, 2007; Qi, 2004, 2007; Watanabe, 2004b) lend support to this view. For example, Greene (2006, 2007) intends to investigate the influence of IELTS writing test on teaching by setting out to examine how writing classes in IELTS preparation courses differ from other forms of EAP.

Teacher practices were then observed in writing classes of the two types: IELTS preparation classes directed at success on the test and EAP writing classes provided by universities to prepare learners for academic study. Altogether, 197 learners and 20 teachers were observed over 51 classroom hours using a modified version of COLT. Analysis of the frequency data reveals that few differences were noted between the two types of classes in terms of organization, student modality, or content. Thus, a conclusion that was drawn by the researcher is that teacher variables (e.g., teacher beliefs) may encourage practices that cannot be predicted from test design. He posits that many of the differences observed between classes might be linked rather to teacher or institutional variables, such as levels of professional training and beliefs about effective learning, than to the influence of the test. Another example is found in Chapman and Snyder (2000), who drew on their experience to assess how high stakes national testing can improve classroom instruction in a number of countries (e.g., Uganda, Kenya, Chile, Tobago and Trinidad). Their research shows mixed results. In Kenya, the experience was generally successful, while in Uganda and Chile, it was not. The findings reveal that teachers tended to attribute poor assessment results to factors beyond their control, rather than consider possible inadequacies in their own instructional practices. Nevertheless, analyses of the data led the researchers to conclude that at the higher cognitive level, there is no particular linkage between what an examination asks and what a teacher does. The strongest message of this conclusion is that it is not the examination itself that influences teachers' behaviours, but teachers' beliefs about those changes. In their view, the power of tests to influence instruction is a perceptual phenomenon- "if you believe it does, then it does" (p.462). A major point summed up by them is that teachers' perception counts far more than all the other factors in the perpetuation of the use of traditional methods even in the face of overwhelming evidence of their limited effectiveness. Based on their analyses, one likely reason for the failure of tests to function as a means of innovation is that the teachers do not have the necessary content knowledge or pedagogical skills to meet new demands. One can speculate that this may partially be attributed to the fact that teachers are not trained or guided as they are in Turner (2008), (2009), Davison (2008), Tavares and Hamp-Lyons (2008), and Urmston and Fang (2008). Although Chapman and Snyder's study lacks a solid theoretical base, it is still valuable in that the researchers adopt a broad and cross-cultural perspective rather than a mono-cultural perspective when examining the washback phenomenon. Above all, it

highlights many factors that other studies ignore.

Consistent with the view held by Chapman and Snyder, Watanabe (2004b) indicates that a set of teachers' psychological factors may function either as a debilitating factor or as a facilitating factor. What he means is that putting undue blame on the examination may be functioning as a debilitating factor, while familiarity with a wide range of teaching methods may be a facilitating factor. Watanabe (2004b) bases his claim on a washback study he conducted to investigate the effect of the Japanese university entrance examinations on secondary level classroom instructions. The results of the study also reveal that the 'teacher factor', such as personal beliefs and educational background, is important in the process of producing examination effects. He suggests that only after we have developed a fuller understanding of a teacher's perceptions about his or her teaching will we be able to provide a clear picture of the washback phenomenon. On a similar note, Cheng (1999) acutely points out that tests can hardly change teachers in their fundamental beliefs and attitudes about teaching and learning, the roles of teachers and students, and how teaching and learning should be carried out. The above views also correlate with that of Chen (2002). Chen (2002) reports on a study investigating whether the new testing syllabus in Taiwan junior high schools has helped change teachers' perceptions of their curriculum planning and instruction. Her finding is that the educational innovations in curriculum, textbook and exam could hardly effect fundamental changes in teachers' perceptions of how to teach. The above findings are reinforced by Andrews (2004) who maintains that washback may have relatively "predictable quantitative effects, but rather less predictable qualitative effects upon the teaching-learning process and what actually takes place in classrooms" (p. 42). In an attempt to fully understand the phenomena of teacher beliefs, Burrows (2004) has taken this line of thinking one step further. Based on her study that examined the washback effect in the context of classroom-based achievement assessment in Australia, she proposes a new model for accounting for the washback effect. The model, drawn on Woods (1996), has provided us with a framework for studying the complex and interrelated processes of teachers' beliefs and practices. However, one drawback of the study is that no detailed information has been provided as to how to use the model for analyzing washback. Woods' model will be further discussed in detail later on in this chapter (see Section 2.4.4.2).

From the above review of the literature on washback, we can see that much interest has been sparked in this research area over the past fifteen years or so. Although the

studies are carried out in different contexts and there are different connotations of washback in different studies, the conclusions from the majority of empirical studies to date are consistent in that changes in examinations themselves cannot bring about fundamental changes in teachers' instructional practices, but just superficial changes, unless specific training or guidance is provided to the teachers (Alderson & Wall, 1993; Andrews, 2004; Chen, 2002; Cheng, 2004; Davison, 2008; Green, 2006, 2007; Muñoz & Álvarez, 2010; Shohamy, 1993; Tavares & Hamp-Lyons, 2008; Turner, 2008, 2009; Urmston & Fang, 2008). More specifically, while the impact of tests has been demonstrated in the content of teaching, the teaching approaches employed by teachers remain unaffected if no specific guidance is offered to the teachers on teaching methodology.

In summary, the research evidence discussed above illustrates that washback is a highly intricate rather than a simple and a monolithic phenomenon. Teachers, as compared to testing, are deemed by researchers as more powerful in inducing washback. The rationale for broadening this research to include the understanding of the role of the 'teacher factor' has been discussed in the previous chapter. With respect to the 'teacher factor', one key component identified by researchers concerns teachers' perceptions or beliefs about washback as well as about teaching and learning. It is held that teachers' perceptions of washback directly reflect their attitudes toward teaching (Alderson, 2001). The findings from the above studies seem to reflect that without changing teachers' perceptions as well as upgrading their knowledge, which can be provided in training, we can hardly expect a considerable positive effect of washback. Furthermore, it appears that more recently, with experimentation with the 'right' kind of efficient teacher training/guidance, the teacher knowledge base seems to be enhanced and there is potential for positive washback in some contexts (Davison, 2008; Muñoz & Álvarez, 2010; Tavares & Hamp-Lyons, 2008; Turner 2008, 2009; Urmston & Fang, 2008). Therefore, in view of the promise that the issue of the 'teacher factor' holds for understanding the nature of washback, it needs to be explored further.

In addition to teacher perceptions or beliefs, other important components related to the 'teacher factor' (e.g., teacher knowledge, teacher ability, the degree of teachers' familiarity with a range of teaching methods, teachers' background, etc.) are also assumed to have some part to play in the process of washback. Although limitations of space prevent a full description of them, the issues will be highlighted in the



following sections.

As the current study examines the educational context and testing culture of China, to set the stage for the discussion, in the following section we need to turn to the research conducted in the Chinese context.

### *2.2.2 Washback Studies in the Chinese Context*

Inspired by washback research conducted in the ESL context, an increasing number of washback studies have been conducted in the Chinese context over the past decade (Gu, 2005; Jin, 2000; Kong & Nie, 2002; Qi, 2004, 2005, 2007). Of the studies, Qi (2004) ranks as the most extensive and in-depth one in terms of the scope and rigor of the research. It is a longitudinal study which investigated the washback effect of the National Matriculation English Test (NMET) on teaching and learning. The results of her study demonstrate that despite the test designers' efforts to reduce the weighting on linguistic knowledge in the NMET over the past decade, teachers continue to focus on teaching formal linguistic knowledge at the expense of students' language skills. Such a contradiction between the test designers' intention to promote language use through practice of skills and the actual teacher practice, from Qi's perspective, constitutes a central factor that leads to unintended washback.

In addition to Qi (2004), another large-scale study is that of Gu (2005). Her survey participants include 140 conference participants, 984 teachers nationwide, 1563 students nationwide, as well as 96 teachers and 1042 students at the three sampled universities. Unlike other washback studies mentioned above, which generally involve an innovative language test, Gu (2005) examined an existing test, the CET, on which no fundamental reforms had been implemented between 1999 and 2005. Although she claims that the study has found a preponderance of evidence indicating positive washback of the CET on CE teaching and learning, it appears that the researcher has made bigger claims than the data would allow her to.

Even though her questionnaire data are supplemented by data from classroom observations, it is still hard to observe whether there are changes in the way the teachers teach or what aspects of teacher classroom practice can be attributed to the effects of the CET, due to the absence of baseline data<sup>1</sup> in her study.

---

<sup>1</sup>Baseline data are collected before the introduction of an innovation. They can "serve as a point of comparison when attempts are made to determine whether change has indeed occurred" (Wall & Horák, 2007, p.99).

In spite of its limitation(s), Gu's (2005) study is valuable in that it allows us to gain an initial understanding of Chinese EFL teachers' perceptions of the CET. Furthermore, it also alerts us to the importance of the 'teacher factor' entangled in the CET impact. In addition to the CET, teachers' devotion to their work, their English proficiency (or teacher quality), their teaching methods, and attitudes toward teaching and testing, their views on "teaching" and "practice", were also deemed by her as exerting great influence on CE teaching and learning. An interesting finding from the study is that teacher factors may outweigh the effect of the CET on CE classroom teaching and learning. Another interesting finding is that regular classroom teaching and learning with the creative use of textbooks and other teaching materials result in better test performance and higher test scores.

In comparison to the above two studies, Jin (2000) and Ye (1998a, 1998b)<sup>2</sup> are relatively small-scale studies. Jin (2000) has focused on the washback of the CET-Speaking English Test (CET-SET). The instrument she used for her research is a survey of 358 students and 28 examiners of the CET-SET. The results of her study suggest that 100% of the teachers held that the CET-SET could exert a powerful impact on CE teaching and learning and the test could direct CE teaching to the right path of developing the students' ability to use English. However, in the journal article where the study was reported (Jin, 2000), there is little evidence showing in what way or what aspects of teaching and learning the CET-SET has affected the teachers and learners. Furthermore, little explanation is provided as to whether the test has caused teachers to change their content of teaching and methodology or not. It might be possible that the data concerning the washback effects of this test on teachers and learners is reported elsewhere.

Ye (1998a, 1998b) presents the results of two surveys. One was administered to 74 EFL teachers from 18 institutions of higher learning, and the other was administered to 174 students at Shanghai Jiaotong University. Based on the results of her questionnaires, she claims that the CET has not only brought changes to teaching content and teaching methods, but has also changed the phenomenon of lecture-based instruction, and increased students' learning initiative and independent thinking. However, her study does not provide sufficient evidence or data to justify her claim.

---

<sup>2</sup> Ye (1998a) and Ye (1998b) appear to be the same article published in different journals. See References for details.

In spite of this claim, she admits that grammar and vocabulary continued to constitute a considerable portion in CE teaching. It seems that these conclusions are contradictory.

Contrary to the findings reported by the above researchers (Gu, 2005; Jin, 2000; Ye, 2008a and 2008b), Gu and Liu (2005) and Liu and Dai (2003) have produced a different picture of the impact of the CET on teaching and learning. Liu and Dai (2003) report on a nationwide large-scale survey on teacher perceptions of teaching methods, teacher pedagogical knowledge and potential for conducting research, and issues related to instructional innovations and testing. The results reveal that more than 90% of the CE instructors maintained that the CET cannot objectively reflect students' communicative competence. They attribute the negligence of aural/oral aspects of language in instruction to the phenomenon of teaching test-related items. They argue that as a test which measures students' linguistic knowledge rather than their abilities in language use, the CET could only encourage students to focus their attention on language knowledge. This, according to them, has led to the test's negative impact. They end their paper with a call for devising the CET as a criterion-referenced test. They further suggest that subjective questions be increased, and commercialization of the test be avoided. While the data presented in this study may not be taken as evidence of washback, for it is not associated with "the introduction of an innovation intended to cause change" as described by Wall and Horák (2007, p.99), the study provides us a window on how Chinese EFL teachers perceive the CET.

An overview of the washback research in the Chinese context suggests that despite the increase in the number of studies carried out in this area, empirical studies are scant. There is remarkably little research in Chinese ELT field that can be said to have investigated and established what washback is or how it works. Through the above analysis, three big limitations are found to exist in current washback research in China. The first limitation relates to the research methods employed by Chinese ELT researchers. A number of studies only rely on survey instruments such as questionnaires. The second limitation is associated with definition or conceptualization of the term, washback. While many studies address the issue of washback, they have not achieved a thorough understanding of the theoretical underpinnings of this concept. In other words, the term 'washback' has often been misinterpreted. For example, Ye (1998a, 1998b) has made the assertion that all tests that are related to teaching and learning can engineer washback effects on teaching

and learning. Another drawback of these studies is that the research evidence they have yielded seems far from adequate, since there are occasions on which researchers seem not to be able to provide adequate data to back up their claims.

In view of the above limitations, more rigorous research is called for in the Chinese College English (CE) context.

To put the findings discussed in the first two sections in perspective, below, I will probe more deeply into the ‘teacher factor’ under study.

### *2.2.3 The Teacher Factor Involved in Washback*

Considering the difficulties encountered in inducing positive washback, it is necessary for us to examine some of the salient themes that have emerged from washback studies. Despite the variety in perspectives from which researchers view washback, there are some shared features and overlapping themes among them. In this section, I will discuss and cite a few recurring themes relevant to my study. The findings of washback studies over the past decade have produced a number of fundamental assumptions relating to this ‘teacher factor’ which will form the foundation of this discussion.

#### *2.2.3.1 Recurring Themes in Washback*

##### *Prevalence of negative perceptions of tests.*

Of the numerous patterns or themes that have emerged from studies on washback, the most notable one is the gap that exists between teachers’ beliefs about innovation and the beliefs held by innovators. There is enough evidence indicating that teachers’ perceptions of washback seldom overlap the perceptions of test designers or policy makers (Andrews, 1994, 2004; Cheng, 1999; Qi, 2004; Shohamy, 1993). Contrary to authorities’ belief that a test is an effective tool to motivate teachers to teach, a test is often perceived by teachers as an intrusion (Shohamy, 1993) and its impact is viewed as negative (Alderson, 2004; Andrew et al., 2002; Clark, 1987; Oxenham, 1984). In some cases, well-intended changes to tests can even generate considerable opposition (Chapman & Snyder, 2000). For example, both Burrows (2004) and Clark (1987) note that teachers perceive the implementation of change as being of benefit neither to themselves nor to their students, and this affects the degree to which change is adopted. Similarly, Smith et al. (1994) note that policy makers’ definitions of the situation (e.g., the curriculum, assessment, and educational change, etc.) are often

misinterpreted by practitioners. Andrews (1994) sums up this phenomenon asserting that there tend to be discrepancies between the intention of any innovation or curriculum change and the understanding of teachers who are given the task of implementing that change. His assertion is based on a small-scale study he has conducted comparing the perceptions of examination designers with those of teachers. Such discrepancies are also exemplified by Cheng (1999). As was discussed earlier, Cheng reports that teacher talk under the new syllabus increased for two of the three subjects in the study. However, when asked about the reason for the increase in teacher talk, the observed teachers claimed it was the result of the new examination. Personally they felt they needed to explain and talk more. It was obvious that the messages from the examination authority were not well-communicated to them. Cheng accounted for such results by saying that teachers tended to redefine and reinterpret the messages about the policy they received.

Taken together, the general point made by these studies is that the negative perceptions of examinations held by teachers may cause them to misinterpret the messages about innovation, which may result in the distortion of the original intentions of test innovators. Then why do teachers' accounts of the washback effect tend to be negative? Cheng (1999) attributes the reason to the fact that teachers were unclear about how they should teach. She notes that the teachers tended to follow whatever they felt comfortable with in their teaching. They tried to use various teaching activities according to their understanding of the integrated approach and task-based approach as encouraged by the new HKCEE. A similar view is expressed by Chapman and Snyder (2000), as outlined above. Further support for this view is found in Chen (2002). Like Cheng, Chen (2002) has found in her study that the teachers did not have knowledge about how to implement the new curriculum even though they were aware of the innovated curriculum and testing objectives. She argued that the fact that teachers lack knowledge about how to change their teaching methods to align with the new curriculum should be attributed to a lack of in-service teacher training. Another explanation provided by Chapman and Snyder (2000) and Watanabe (2004b) is that teachers tend to place undue blame on the presence of the examination for what they are doing. According to Watanabe, such blame is ungrounded, in the sense that no linkage can be established between the test and what it is blamed for. For example, in his study, two teachers attributed their extensive stress on vocabulary, idioms and structure to the need for exam preparation.

Nevertheless, an analysis of the examinations shows that only 25% of the exam covered these skills. Hence, the researcher concludes that these teachers may have overemphasized these skills based on their perceptions (Watanabe, 2004b).

*Different teacher, different effect of washback.*

Another notable theme concerns differences among teachers with regard to the changes induced by testing to their teaching. The literature on washback shows that there is considerable agreement among researchers about the variability of the washback effects (Alderson & Hamp-Lyons, 1996). It is generally held that in different test situations, test impact takes different forms and exists at varying degrees. According to Shohamy (1993), the variation in the strength of the impact depends on the type of test and on the degree of the stakes tests carry. Substantial evidence from recent washback studies indicates that teachers not only perceive and interpret examination change differently, but also react differently to it (Alderson & Wall, 1993). Moreover, the washback effect is demonstrated in some aspects of teaching, but not in others. A common thread of these studies is the assumption that the relative degree and type of the washback effect varies not only from project to project but also from teacher to teacher. As Alderson (2004) has pinpointed, different teachers teach to the same test in very different ways or some teachers teach to very different tests in very similar ways. The example provided by Watanabe (1996) serves to illustrate this. In his study, the lessons of some teachers were characterized by the negative influence of the examinations (e.g., focus on formal linguistic knowledge, reliance on the grammar-translation method, etc.), but others were not (e.g., using exam materials for developing students' authentic language ability). It appears that the instructional strategies and activities adopted by teachers differ significantly. Similar cases are found in Alderson and Hamp-Lyons (1996), Burrows (2004), Cheng (1997, 1998), Hayes and Read (2004) and Shohamy (1993). Hayes and Read (2004) conducted a study of the impact of the IELTS test on the way international students prepare for academic study in New Zealand. The two teachers observed were found to maintain different objectives of teaching and the different objectives led them to deliver their courses in strikingly different ways. In addition to different teaching behaviors, Burrows (2004) reports that teachers' interpretations of the reasons leading to examination change also varied considerably. Some teachers reported changes caused by the increased accountability and responsibility brought by the implementation of

the assessment, while others attributed the reason to teaching new topics related to competencies. Another case in point is Cheng (1999). In her study, she spotted substantial variation in teachers' beliefs about how to teach and how students learn. This finding is crucial, for the difference in teachers' beliefs may help to account for the varying degree of the effects of washback on teaching. It seems to suggest that positive washback effects might occur with teachers who maintain a clear conception of their teaching and learning.

As teachers have been found to adopt different instructional strategies in the face of washback, both Alderson and Hamp-Lyons (1996) and Watanabe (2004b) suggest expanding the existing 15 Washback Hypotheses put forward by Alderson and Wall to include another hypothesis. That is, tests will produce different amounts and types of washback to different teachers. Furthermore, Andrews et al. (2002) urge that individual differences among both teachers and students be taken into account when we study the washback phenomenon.

Among the factors that are related to teaching variation, teacher ability, teachers' understanding of the test, teaching experience, the adequacy of their training are deemed as central in leading to the differences in their perceptions, interpretations and reactions.

One account given by Shohamy (1993) is that teaching variability may be associated with teacher training. In Shohamy's (1993) study, as novice teachers were equipped with a variety of methods for teaching oral English, the impact the test had on them was hardly identifiable. It seemed that they were immune to the washback effect. Instead of focusing on test-like activities, they tended to try out innovative teaching methods and organize communicative activities in their classes. Shohamy attributed such a practice to the teacher-training they had received. She seemed to suggest that teacher training is one of the major factors for why washback happens to some teachers, but not to others. In addition, Watanabe (2004b) suggested that the difference was related to teachers' past experience as students. Burrows (2004) has extended Watanabe's discussion with the suggestion that teachers' political viewpoint, educational background, culture, and their beliefs, assumptions, and knowledge also contribute to this teaching variability. In his study, a teacher expressed that it was difficult to employ other methods than the one by which he was taught when he was a student. One assumption from the above discussions is that the better equipped in instructional strategies the teacher is, the less powerful the washback will be.

Nevertheless, more examples are needed to confirm the assumption. While other studies also report that some teachers remain unaffected by tests (Watanabe, 2004b; Hayes & Read, 2004), no detailed information has been provided about their background. Hence, no inference could be drawn.

Considering that our purpose is not merely to reveal these differences, but also to account for them, more theoretical and empirical evidence is needed if explanations of such differences are to be achieved.

*Discrepancy between said and done.*

The following theme is related to the inconsistencies between what teachers perceive they do and what they practice. The data from washback studies reveal that teachers' interpretations of teaching are inconsistent with their classroom practices observed by researchers. In other words, it seems that there is a gap between what teachers claim they do and what is actually happening in the classroom. Such a discrepancy between what people say and what they do has initially been noted by Wall and Alderson (1993). Later, other researchers Cheng (1999), Wall (1999) and Watanabe (2004b) have also alerted us to the complexities and contradictions between what people say and what they practice. Wall and Alderson (1993) and Wall (1999) find that while teachers claimed that they had changed their way of teaching by using a 'communicative approach', they were in fact found to be using a teacher-centered approach that textbook designers had originally intended to discourage. Coincidentally, Cheng (1999) shows that although the three teachers in her study scheduled their lessons as oral, the tasks observed were short listening, reading and writing. She has interpreted such a finding as an instance indicating that the teachers did not necessarily practice what they believed they practiced. Moreover, the results of her study also reveal that the teaching pattern followed by the teachers in the two years made little difference. However, when asked about the washback effect of the HKCEE on their teaching, all three observed teachers claimed in the follow-up interviews that the exam had a strong impact on their teaching methods. One teacher even claimed to have changed the perception of her role as a teacher. In light of the teachers' definite claims, it seems that their interpretation of the methodology is different from that meant by the researcher. Watanabe (2004b) quotes Allwright as suggesting that we should not exclude the possibility that there may be a teacher "who proclaims he or she is exam-influenced, but whose teaching exhibits nothing that can



be related by observation to exam content” (p.31).

In seeking to understand such a discrepancy, both Alderson and Wall (1993) and Watanabe (2004) suggest taking account of attribution theory. Based on this theory, there are four main kinds of attributions to which people tend to refer for their perceived successes and failures in life: ability and effort (internal attribution), and luck and the perceived difficulty of the task with which they are faced (external attribution) (Williams & Burden, 1997). Williams and Burden (1997) have cautioned us that “individuals will vary in the way in which they personally view these attributions” (Williams & Burden, 1997, p. 105). A finding drawn on by them from this line of research is that most people tend to externalize reasons for failure, while internalizing reasons for success. However, they stress the possibility that what is seen as an external factor outside of someone’s control can be internalized and made more controllable. In this regard, teachers’ claims need to be treated in an objective way.

*Complex webs of factors entangled in the washback phenomenon.*

Through the review of pertinent literature, I find that the study of washback exhibits considerable complexity. Considerable research evidence suggests that simply examining one factor, or examining the phenomenon in only one context with little consideration of the multifaceted and complex process of testing, teaching and learning, can yield simplistic and invalid results. It is argued by researchers that the process of effecting assessment change is far more complex than proponents of positive washback have envisioned. Exerting positive washback, as Cheng and Curtis (2004) put it, involves much more than just the design of a “good” assessment. To elucidate its complexity, researchers describe washback as conditional (Shohamy, 1993), circuitous (Wall & Alderson, 1993), and indirect and unpredictable (Andrews, 1995, 2004; McNamara, 2000). Also, based on Shohamy’s analysis, rather than being static, it changes over time. Another characteristic summarized by Andrews et al. (2002) is that the washback effect might be delayed rather than immediate.

One direct consequence brought about by this complexity is that in some cases researchers have to draw tentative conclusions about their research findings (see in Watanabe 2004b and Andrews et al., 2002).

Sophisticated as the washback issues are, researchers realize that the factors underlying these complexities should not be overlooked. Rather, much weight should be attached to them. When reflecting on the nature of washback, Andrews et al. (2002)

remind us to be mindful of individual differences among both teachers and students, for he assumes that they are the cause of unpredictability of the washback effect. He also warns us to “acknowledge and work within the constraints imposed by the complexity of the innovation process: the time that it takes, the depths of the changes that successful implementation might entail and the concerns of the stakeholders” (Andrews, 2004, p.49).

*Contextual factor.*

Countries differ considerably in important contexts with regard to their culture, socio-economic systems, and their educational systems. It is widely accepted that innovations are particularly context-sensitive (Kenney, 1988; Markee, 1997; Nespor, 1987; White, 1987). Without exception, testing innovation is also believed to be entangled with various contextual factors (Cheng, 1999; Cheng & Curtis, 2004; Shohamy et al., 1996; Wall, 1996; Wall & Alderson, 1993; Wall, 2000; Watanabe, 2004). As indicated above, in some EFL contexts where teaching situations are more complicated, achieving positive washback seems to be an unobtainable goal. In contrast, certain ESL contexts demonstrate that washback is predictable if a testing innovation is accompanied by ongoing teacher training (Davison, 2008; Tavares & Hamp-Lyons, 2008; Turner, 2001, 2002, 2008, 2009; Urmston & Fang, 2008). Regardless of what the results are, there is almost a general consensus among these studies that to account for the conflicting results, the overall local educational context as well as the teaching situations which facilitate the construction of test impact should be probed into.

In the literature, these contextual factors are mainly discussed as local context (micro context) and societal context (macro context). McNamara (2000) highlights the ‘small picture’ saying that whether or not positive washback is achieved will depend on local conditions in classrooms, the established traditions of teaching, the immediate motivation of learners, etc. In comparison, Shohamy (1997) sensitizes us to the big picture – the general political, cultural and historical contexts within which washback exists. Despite this contrast, there has been congruence among them and other researchers as to the need to understand both pictures. In Wall’s (1999) words, different types of problems need to be solved at individual, school and societal levels.

Hence, due to the scope of contextual factors that are intertwined in the washback phenomenon in China, Chapter 3 is devoted to discussions of these issues.

### 2.2.3.2 *Limitations of the Research Results on Washback*

Over the past decade, there has been an unprecedented amount of research on washback. This domain of research seeks to answer, in one form or another, one fundamental question – how testing influences teaching and learning. All the research studies cited above have provided us with a steady accumulation of knowledge about the nature of washback. However, despite numerous positive qualities demonstrated in the above-mentioned washback studies, we have noticed that they are limited to some extent. In what follows, I will discuss some of these limitations.

One obvious limitation of such studies is that since they simply focus on a narrow set of factors associated with testing itself, the authors are still not able to explain the nature of the washback phenomenon. Due to the narrow research focus, many assertions and statements made in these studies, though differing in wording, overlap in meaning. Furthermore, although the issue of the ‘teacher factor’ has been touched upon by a few researchers (Davison, 2008; Green, 2006, 2007; Tan, 2008; Tavares & Hamp-Lyons, 2008; Urmston & Fang, 2008) and has begun to be explicitly and intensively dealt with in Turner (2008, 2009), additional data need to be collected to enable researchers to examine and address the issue more closely and extensively, and above all, to illustrate whether the findings from Canada, Hong Kong and South Africa apply to other contexts as well.

Little description is provided as to how teachers’ perceptions are produced and what negative impact they may exert on teaching. Furthermore, little is mentioned about how well teachers are informed of the change before it is carried out. This information is crucial in the sense that it is hard to expect teachers to take the initiative in changing their instructional approaches if the messages from the examination authority are not well-communicated to them. Hence, these flaws should also be addressed in future research.

Another flaw lies with the research design of certain studies. Thus, the following section will address some of the methodological issues worthy of note in washback research.

## 2.3 Overview of the Research Methodologies Employed in Washback Research

The above review of the literature also reflects that methodological issues pertaining to this domain of research seem as sophisticated as the washback phenomenon itself. Although different types of research designs have been developed

and a variety of methodological tools have been employed in various washback studies, they have not been intensively summarized and discussed in the literature. Therefore, it would be informative to take a brief look at the principal research methods utilized in this research area and outline a set of design characteristics that are assumed to be appropriate for examining the washback phenomenon.

### *2.3.1 The Evolution of Washback Research*

It should be noted that the methodologies utilized in washback studies have undergone a developmental change. There has been an evolution in this field of research from the use of a single method or monomethod (e.g., survey methods) to the use of multiple methods or mixed methods (e.g., survey methods complemented by observations). Between 1980 and 1990, little empirical research had been carried out to investigate the washback effect of examinations either in the field of general education or in the field of language education. Specifically, research design during this period was largely dominated by survey methods (usually interviews or written questionnaires), with observation being overlooked. Nevertheless, although the questionnaire data have provided a great deal of information on the relationship between teaching, learning and testing, these data alone could hardly provide a clear and accurate portrayal of what is actually happening in the classroom.

In comparison to earlier studies which simply used a single data source, later studies embraced multiple data sources. It is widely acknowledged that the most substantive contribution in this area which led to the popularization of the use of multiple methods is the Sri Lanka Impact study reported by Alderson and Wall (1993), Wall (1999) and Wall and Alderson (1993). This long-term impact study was jointly carried out by a research team over a period of two years. It differs from other studies in that it is the most comprehensive and thorough study that has ever been conducted in this research area. The entire study was composed of several sub-projects: a baseline study, questionnaires to teachers and teacher advisers, teacher interviews (group), document and material analyses (especially tests), and, most importantly, a two-year observation program. It is worth noting that the research team (7 Sri Lankan teachers) conducted six rounds of classroom observations in a total of 49 schools across the country.

Drawing on the last two sets of observations, Round 5 and Round 6, Wall and Alderson (1993) highlight the importance of incorporating a strong observational

component in data collection instrumentation of research on washback. They assure us that without observations, they would not have been aware of the inconsistencies between what teachers perceive they do and what they practice. There is little doubt that this study is unparalleled by any of this type in terms of the scope and quality of the data as well as the depth of data analysis. The value of this study lies in the fact that it has not only extended the theoretical basis and vision of washback research, but has also suggested new avenues for investigation as well as expanded the range of research tools and methods for this type of research. Most important of all, it has motivated a substantial amount of evidence-based, observational washback research (Alderson & Hamp-Lyons, 1996; Burrow, 2004; Cheng, 1997, 1998; Hayes & Read, 2004; Qi, 2004; Shohamy et al., 1996; Turner, 2002, 2008, 2009; Watanabe, 1996a, 1996b, 2004b). Table 1 provides detailed information about the different types of methods employed in these studies.

Table 1 *Summary of Major Washback Studies*

Researcher	Time of publication	Exam or test	Research context	Research method	Participants	Length of research
Alderson & Wall Wall & Alderson Wall	1993	Sri Lankan O-level Examination	Sri Lanka	A baseline investigation, questionnaires, document and materials analyses, classroom observations, and group interviews.	Teachers and students	Long-term (2 years)
	1993					
	1996					
	1997					
	1999					
Cheng	2000	The Hong Kong Certificate of Education Examination in English	Hong Kong	Questionnaires, interviews, and classroom observations	Test developers, textbook writers, teachers, and students	Long-term (2 years)
	1997					
	1998					
	1999					
	2001					
Turner	2003	Quebec Secondary Five ESL speaking exam	Quebec, Canada	Questionnaire, interview, and classroom observation	Teachers and students	Long-term (6 months)
	2004					
	2001					
	2002					
	2005					
Alderson & Hamp-Lyons Watanabe	2008	TOEFL	United States	Interviews and classroom observations	Teachers and students	Short-term
	2009					
	1996					
	1996a					
	1996b					
Shohamy Shohamy et al.	2004a	University Entrance Exams	Japan	Interviews and classroom observations	teachers	Short-term
	2004b					
	1993					
	1996					
		A test of Arabic as a second language; an EFL oral test; Hebrew L1 reading comprehension test	Israel	questionnaires, interviews, document analyses, and classroom observations	teachers, students, and administrators	Long-term

It is also encouraging to note that more and more researchers have expanded to

look at issues of context in order to capture the complexity of the washback phenomenon (Cheng, 2001; Cheng, 2004; Davison, 2008; Qi, 2005; Shohamy, 1993; Tavares & Hamp-Lyons, 2008; Turner 2008, 2009; Urmston & Fang, 2008; Wall, 1999; Watanabe, 1996a, 2004b). It is obvious that the washback phenomenon has been examined much more seriously, both theoretically and empirically.

Also worth noting is that adopting the mixed-method approach is the growing trend in current washback research. Based on Wall (1999), the research questions in washback research are best answered with mixed-method research designs rather than with a sole reliance on either the quantitative or the qualitative approach. Turner (2005, 2008, 2009) attests the importance of using multiple methods of data collection (a mixed-method design) and provides a good example of how rigorous washback research which combines qualitative (QUAL) and quantitative (QUAN) methods can be designed. Her study will be discussed again below.

As an outline such as this does not inform us about how current washback studies are structured and why they are designed the way they are, in the next section, I will look closely at the design features as well as the current approaches that have strong practical appeal to researchers on washback, and will discuss how well the approaches suit the purpose of their respective studies.

### *2.3.2 Design Characteristics and Approaches to Washback Research*

As indicated above, a mixed-method orientation has been embodied in the design characteristics of recent washback research (Bailey, 1999; Burrow, 2004; Cheng, 2001, 2003; Qi, 2004; Turner, 2005, 2008, 2009; Wall, 1999; Wall & Alderson, 1993; Watanabe, 2004a). However, although most washback researchers are committed to adopting multiple measures of data collection and analyses, the theoretical orientations that underlie their data collection approaches hardly entered the picture, except in the most abstract sense. It was not until recently that the use of mixed methods research (MMR) as a research design was articulated in researchers' explanations of their methodologies (Turner, 2008; 2009). Then, in what follows, I will examine the theoretical groundings of the mixed methods (MM) design as well as some of the unique design features subsumed under it.

### 2.3.2.1 *Characteristics in Research Design of Washback Research*

#### 2.3.2.1.1 *Mixed methods design.*

In Tashakkori and Teddlie's (1998) and Teddlie and Tashakkori's (2009) definition, a mixed (or combined) methods study is one in which the researcher uses multiple methods of data collection and analysis. Ample evidence shows that the MM approach has gained broad appeal in research from different disciplines (Creswell & Plano Clark, 2007; Greene et al., 1989; Johnson & Onwuegbuzie, 2004; Tashakkori & Teddlie, 1998; Teddlie & Tashakkori, 2009; Turner, 2005, 2008, 2009).

A review of the pertinent literature (Creswell, 1994; Creswell & Plano Clark, 2007; Lincoln & Guba, 1985; Miles & Huberman, 1994; Patton, 1990) indicates that MMR emerged in the 1960s and became more common in the 1980s. Researchers of various disciplines (Sieber, 1973; Creswell, 1994; Greene et al., 1989) have offered a detailed list of the reasons for employing this approach. Creswell and Plano Clark (2007) advocate conducting research along these lines, saying that "the use of quantitative and qualitative approaches in combination provides a better understanding of research problems than either approach alone" (p.18). In Greene et al.'s (1989) terms, the reasons for advancing this approach include: triangulation; complementarity, in that overlapping and different facets of a phenomenon may emerge; sequencing, wherein the first method is used sequentially to help inform the second method; initiation, wherein contradictions and fresh perspectives emerge; and expansion, wherein the mixed methods add scope and breadth to a study. In light of these practical reasons provided by different researchers, it seems that there is a need to examine the theoretical grounding of this approach.

Tashakkori and Teddlie (1998) and Teddlie and Tashakkori (2009) argue that the mixed-method approach is underpinned by philosophies of pragmatism. It is interesting to note that their argument is reinforced by Creswell and Plano Clark (2007). They assert that investigators may view MMR strictly as a "method," thus allowing researchers to choose any method from different schools of methodology based on diverse philosophical assumptions. Pragmatism is described by both Creswell and Plano Clark (2007) and Teddlie and Tashakkori (2009) as a paradigm (or a worldview) distinct from positivism or constructivism. The positivist paradigm is thought to underlie quantitative methods, while the constructivist paradigm is held to underlie qualitative methods (Creswell, 1994; Lincoln & Guba, 1985). As an epistemological orientation, pragmatism is considered to be the best paradigm for

justifying the use of the MM approach (Creswell & Plano Clark, 2007; Teddlie & Tashakkori, 2009). As Teddlie and Tashakkori (2009) have stated, pragmatically oriented theorists and researchers postulate the compatibility between quantitative and qualitative methods. Rather than viewing them as two opposing poles, pragmatists regard qualitative and quantitative research as a continuum along which research methods fall (Patton, 1990; Tashakkori & Teddlie, 1998; Teddlie & Tashakkori, 2009).

The value of the MM approach lies in that it allows researchers to mix aspects of the qualitative and quantitative paradigms at all or many methodological steps in the design (Creswell, 1994; Creswell & Plano Clark, 2007). Patton (1990) conceptualized “methodological mixes” saying that different methods (QUAN and QUAL) could be combined across three stages: design (naturalistic inquiry or experimental), measurement (QUAL data or QUAN data), and analysis (content or statistical). The first stage includes the formulation of the research question, or hypothesis. The second stage involves data collection and the third stage concerns data analysis. Allwright and Bailey (1991) extend Patton’s (1990) conceptualization saying that various combinations of quantitative and qualitative data collection and analysis are possible. Along this thread of comment, Creswell and Plano Clark (2007) provide us a more elaborated definition of four major types of MM design: (1) triangulation design – collecting qualitative and quantitative data simultaneously to understand a problem; (2) embedded design – using qualitative data in an experiment or correlational study; (3) explanatory design – explaining quantitative results with qualitative data; and (4) exploratory design – using qualitative data and analysis in an exploratory function toward developing a quantitative instrument. They also address multiple decisions involved in selecting a particular design such as timing (the order in which researchers use the data), weighting (the relative importance or priority of QUAL vs. QUAN), and mixing (the explicit relating of the two types of data). It is worthwhile to note that the four types of designs address different objectives. They can serve as a foundation for conceptualizing how to design and conduct feasible MMR.

Furthermore, two distinctive characteristics are found to exist in the MM designs. One is triangulation. Research in washback studies (Wall & Alderson, 1993; Cheng, 1997, 1998; Turner, 2002, 2005, 2008, 2009; Wall, 1999) demonstrates that all of the MM designs used triangulation techniques. The second characteristic is that such designs stress the importance and predominance of the research question over considerations of either method or paradigm (e.g., the worldview that is supposed to



underlie that method).

Subsumed under the MM approach is an array of methods combining both quantitative and qualitative research: observations, interviews, document reviews, questionnaires, and so on.

Owing to the above-mentioned strengths, the MM approach is becoming more and more popular with researchers in the domain of washback research. Except for Watanabe (1996) and Alderson and Hamp-Lyons (1996), the majority of washback studies have embraced this approach. In fact, the oft-cited Sri Lanka Impact study as well as the study by Turner (2002, 2006, 2008, 2009) have demonstrated a successful combination of survey research and QUAL procedures. Turner (2005, 2008) clearly states that the research design and analytic procedures of her study have been informed by the principles of the MM approach. Like Wall and Alderson (1993) and Cheng (1997, 1998), her longitudinal study relies heavily on QUAL methods such as observations, interviews, and document analysis, but incorporating a complementary QUAN component (e.g., questionnaires). From a methodological perspective, she forcefully argues that the complex washback phenomenon necessitates the use of both QUAL and QUAN research methodology. Her argument is supported by Watanabe (1996b) and Chen (2002) who also strongly believe that QUAL and QUAN methods can be profitably used together in the study of washback.

Given the growing recognition of the role of the mixed-method approach for investigating the washback phenomenon, there has emerged a need for at least a basic familiarity with the appropriate methods to which most washback research subscribes. Therefore, next we turn our discussion to the multiple methods subsumed under the mixed-method approach.

#### *2.3.2.1.2 Methods in studying washback.*

As mentioned above, basically, the MM washback researchers have drawn on include observations, interviews, document reviews, questionnaires, etc.

Although it is not difficult to obtain a general understanding of these methods, each method has a number of variations or sophistications that deserves our attention. Considering the crucial role these methods play in the actual research, each instrument merits a bit more discussion.

*Interviews and questionnaires.*

A key issue we should be aware of is that different methods may perform different functions in different studies. In general, researchers have used surveys (questionnaires and interviews) not only to gather information about participants' characteristics (to elicit biodata), but also to uncover the opinions and attitudes of the participants about washback as well as their views and perspectives on language teaching and learning (Cheng, 2004; Qi, 2005; Turner, 2005, 2008, 2009; Watanabe, 1996a). However, we should also be aware that the two forms of surveys differ in that "questionnaires are particularly efficient for gathering information on a large scale" (Brown, 1997, p. 111), while interviews are strong in enabling researchers to probe into the depth of teachers' knowledge, since they focus on gathering detailed information from a smaller number of subjects (Brown, 1997). Brown's (1997) position is consistent with that of researchers on washback. Cheng (2004) views questionnaires as being able to provide a general picture of how teachers and students react in the context of her study. Whereas the strength of interviews, based on Watanabe (1996a), lies in that they can detect and explain the reasons behind teachers' behaviors in the classroom. Similarly but more explicitly, Qi (2005) states that the goal of employing questionnaires in her study is to find out how far the interview results can be applied to a larger group of participants. Obviously her attempt is to establish and facilitate the generalizability of insights derived from the QUAL data of her study.

*Observations.*

As discussed above, observation has become one of the essential tools for the investigation of washback. According to Wall and Alderson (1993), the perceived value of classroom observation is that it allows researchers to have more direct access to the teachers' behaviors and interaction patterns in the classroom. In their words, it can help determine what teachers teach and how. Moreover, it eliminates the need to ask individuals about their behaviors or tendencies which are sometimes not reliable. One shared feature among a number of studies is that at least two rounds of classroom observations have been conducted before and after the implementation of a new exam (e.g., Alderson & Wall, 1993; Cheng, 1997; Wall & Alderson, 1993; Shohamy, 1993; Turner, 2005). Their assumption is that the two sets of data obtained would lend themselves to the kinds of comparisons that washback research required.

### *Case Studies.*

What deserves mention here is that a type of naturalistic inquiry that has been considered to be particularly important in washback research is the case study method. Nunan (1992) explains the term saying that in conducting a case study, “one selects an instance from the class of objects and phenomena one is investigating and investigates the way this instance functions in context” (p. 75). The reason why this method has been largely favored by washback researchers (Alderson & Hamp-Lyon, 1996; Cheng, 2003; Watanabe, 1996a, 1996b, 2004b; Turner, 2002) is that it is held to be able to produce a set of information-rich data (Cheng, 2003; Watanabe, 2004a).

#### *2.3.2.2 Existing Design Patterns in Washback Research*

Although the methods used vary from study to study, there are some shared features in these studies. In general, there are two types of designs adopted by washback researchers. One type of design is that of Watanabe (1996a, 1996b, 2004b); Alderson and Hamp-Lyons (1996); and Hayes and Read (2004). The methods these researchers have used mainly include interviews and observations. As a rule, each of their participants teaches two different courses (a regular course and an exam-preparation course). Based on Watanabe, washback is considered to exist if the following conditions are met.

- *Teaching, learning, and/or textbooks are different in exam-prep and in non-exam classes taught by the same teacher.*
- *Teaching, learning, and/or textbooks are similar in the courses which are taught by two different teachers.*

	Teacher A		Teacher B
Exam prep. lessons	↑ different	←same→	↑ different
Non-exam lessons	↓		↓

*Figure 1.* Washback research design by Watanabe (1996a, 1996b, 2004b). (Note: From Watanabe, Y. (2004a). Methodology in washback studies (p.28). In L.Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum Associates. Copyright 2004 by Lawrence Erlbaum Associates. Reproduced with permission)

The other type of design is that of Wall and Alderson (1993), Cheng (1997, 1998)

and Turner (2002).

Turner (2002) demonstrates a good example of such a design. This longitudinal study investigates how the effort to change the Speaking section of the provincial ESL exam would bring about fundamental changes in classroom teaching in Quebec secondary schools. Turner's data sources consist of questionnaires, semi-structured interviews, classroom observations, and post-observation "chats". The whole study involves three phases. In Phase 1 of the research, baseline data were collected through questionnaires, semi-structured interviews (to both teachers and students), classroom observation and a post-observation "chats". In Phase 2, time right after the official introduction of speaking exam procedures, another round of classroom observation and post-observation "chats" were carried out. In Phase 3, time right before the administration of the speaking exam, the third round of classroom observation and post-observation "chat" were conducted. Then after the speaking exam, final questionnaires and semi-structured interviews were administered (to teachers and students).

<b>Period 1</b>	<b>Period 2</b>	<b>Period 3</b>
Baseline data Regular teaching time	Time right after the official introduction of speaking exam procedures	Time right before administration of speaking exam
Initial Questionnaire & semi-structured interviews	Classroom observation	Classroom observation
Classroom observation	Post-Observation "Chats"	Post-Observation "Chats"
Post-Observation "Chats"		Final Questionnaire & semi-structured interviews (after exam)

*Figure 2. Washback research design by Turner (2002, 2005, 2008). (Note: From Turner, C.E. (2002). Investigating high-stakes test impact at the classroom level. Presented at 24th Annual Language Testing Colloquium, (LTRC '02). Hong Kong. Unpublished Manuscript. Reproduced with permission)*

What is worth noting is that Turner (2002), Watanabe (1996a, 1996b, 2004b) and the Sri Lanka Impact study all seemed to roughly follow the ethnographic sequence described by Watson-Gegeo (1988). Watson-Gegeo (1988) categorizes the methodology as a sequence comprised of three stages: 1) the comprehensive stage, in which the researcher collects all potentially pertinent data; 2) the topic-oriented stage, in which the research topic is narrowed via focused observations; and 3) the hypothesis-oriented stage, in which hypotheses are tested and research questions

answered, through in-depth interviews, more focused observations, discourse analysis, etc.

From the above discussion, we find that the preferred methods for researchers in this field are questionnaires, interviews and observations. But although the range of basic research methods is relatively small, they have been blended somewhat differently in each study. One possible explanation for this is that the variety of needs in different studies requires a variety of methods. Moreover, the “mixed-method” approach may be applied to different phases of the study. In Cheng (1997, 1998), Turner (2002) and Wall (1999)’s studies, the QUAN and QUAL approaches are mixed in the phases of both data collection and analysis, whereas in Watanabe (1996a) and Alderson and Hamp-Lyons (1996), the two approaches are mixed only in the phase of data analysis. Regardless of the way they are combined, the key point is that the methods they use are workable and allow them to generate interesting and valid answers to their research questions. These design characteristics constitute a very different approach to research on washback.

In addition to issues of methodologies and methods, data collection strategies also include procedures for monitoring the quality of the data, for methods and procedures do not guarantee validity. It is commonly held that to enhance the methodological rigor of a research study, considerations should be given to the validity and reliability of the research. As Newman (2000) put it, the more researchers are aware of these issues, the more likely it is that standards of good and effective research can be established. (Newman, 2000).

Hence in the following sections, I will probe into the qualities of washback studies – that is, the strengths and weaknesses underlying these studies in terms of the validity (in QUAN studies) or trustworthiness (in QUAL studies) of conclusions and/or inferences.

### *2.3.3 Positive Qualities in Current Washback Studies*

When evaluating the quality of research studies, a key issue we should know is that the criteria by which different researchers assess the value of various kinds of research may vary considerably. From a methodological perspective, the generally accepted standards for evaluating the quality of research are related to the internal and external validity of the research design (see in Cook & Campbell, 1979).

Traditionally, internal validity is conceptualized by QUAN researchers as “the

degree to which we can trust the conclusions/inferences of the researcher regarding the “causal” relationship between variables/events”. (Tashakkori & Teddlie, 1998, p. 67). External validity refers to the extent to which the results of a study are generalizable.

Although validity is also considered a key component of QUAL design (Maxwell, 1996), notions of QUAL research differ substantially from those of QUAN research. However, despite the conceptual differences, the design validity issues have received much attention in both traditions. Maxwell (1996) defines the term in a fairly straightforward, commonsense way. In this paper, I will follow his definition by referring to validity as “the correctness or credibility of a description, conclusion, explanation, interpretation, or other sort of account” (p 87). In fact, this commonsense use of the term is consistent with the way it is generally used by other QUAL researchers. Maxwell (1996) also argues that QUAL researchers generally deal with validity threats as particular events or progresses that could lead to invalid conclusions.

In QUAN research, researchers place much emphasis on the generalizability of the research results to other contexts, whereas the QUAL tradition has its “own canons of rigor” (Shavelson & Stern, 1981, p. 459). Rather than being concerned with establishing generalizable relationships between variables as in QUAN research, it is concerned with generating insight and understanding (Maykut & Morehouse, 1994; Nunan, 1997). In a similar vein, Bailey (1999) also holds the perspective that researchers are trying to promote insight and understanding, rather than causality and proof.

There are different sets of criteria (e.g., representativeness, retrievability and confirmability) and different types of validity (e.g., content, construct, etc.). A note of caution is that QUAL research does not have a precise equivalent to the notions of validity in QUAN studies. Instead, a global QUAL concept – “trustworthiness” has been introduced by Lincoln and Guba (1985). It can be taken as a substitute for many of the QUAN design and measurement quality issues. From Lincoln and Guba’s (1985) perspective, four criteria (credibility, transferability, dependability, confirmability) could be collectively combined to determine the trustworthiness of an inquiry. To establish these qualities in a study, a list of standards for QUAL research has been recommended by them and supported and extended by a number of other researchers (Creswell, 1994; Tashakkori & Teddlie, 1998; Miles & Huberman, 1994;

Maxwell, 1996). Included in the list are neutrality, prolonged engagement, persistent observation, triangulation, peer debriefing, feedback, member checking, thick description, audit trail, reflective journal writing, clarification of researcher bias, etc. This list of standards is crucial, for it not only provides a foundation for evaluating research, but also serves as a guideline for monitoring and ensuring the quality of QUAL research in a way that is similar to the criteria specified in Cook and Campbell (1979) for QUAN research. It is now generally accepted that a research design following these standards can lead to valid conclusions.

There are numerous positive qualities displayed in washback studies, however, it is beyond the scope of this paper to cover them all in detail here. Four main strengths that characterize this domain of research are “triangulation”, “persistent observation”, “thick description of the content”, and “explicit emphasis on research question(s)”. To put them briefly, first, the strategy of triangulation (‘data triangulation’ as in Shohamy et al., 1996 and Qi, 2004, 2005; ‘investigator triangulation’ as in Turner, 2005; ‘methodological triangulation’ as in Wall & Alderson, 1993; ‘time triangulation’ as in Turner, 2002 and Shohamy, 1993; ‘theory triangulation’ as in Watanabe, 1996 and Burrows, 2004) has helped strengthen the reliability as well as the internal validity of these studies. Second, carrying out relatively long-term repeated classroom observations (see in Cheng, 2003; Shohamy, 1993; Turner, 2002; Wall & Alderson, 1993) has helped enhance their external validity. Third, it is beyond doubt that a thick description of the context of the study, with sufficient detail (see in Qi, 2004, 2005; Shohamy, 1993; Turner, 2005; Watanabe, 1996a, 2004) can not only enhance the validity of the results and interpretations of the study, but can also enhance the potential transferability of the results to other contexts (Creswell, 1994; Watanabe, 2004a). Fourth, being more concerned with the research questions than the paradigms that underlie the methods used is obviously a positive quality of washback research. Nevertheless, despite the above positive qualities manifested in current washback research, there are numerous weaknesses and limitations in various studies that deserve our attention.

Alderson (2001) points out that washback has not been properly researched by testing bodies, who may well not welcome the results. His view can be interpreted to mean that although the “mixed-method” approach adopted by washback researchers can obtain fairly reliable indications of the washback phenomenon, they cannot, and should not, make any pretence to have measured and documented the impact of

testing on teacher beliefs and behaviors in the strict sense.

*Summary: Research methodologies in washback research.*

This section aims at covering all important issues and aspects of how research has been and can be conducted on the phenomenon of washback. These issues and aspects are pivotal, since the evolution of washback research over the past decade, the multiple design characteristics and positive qualities outlined, as well as the abundant examples cited in this chapter have provided us with valuable information that will help us conceptualize our own research design and make an informed choice when selecting research methods. What is valuable about reviewing these issues is that it has enhanced our understanding of how rigorous research which combines QUAN and QUAL methods has been and can be designed in studies on washback.

The above review of the literature in current washback research reflects the fact that conducting this type of research presents researchers with many theoretical and methodological challenges. An initial review of the innovation research suggests that some of these challenges are related to the potential problems present in innovation research at large. We note that facilitating testing innovation seems to be subject to essentially the same considerations as those that apply in other types of educational innovation. In this regard, it seems unwise to concentrate exclusively on the testing dimensions of educational change. Rather, we need to borrow ideas from other disciplines to highlight the essential issues relevant to this study. Hence, in the section which follows, I will draw on the growing body of research in different areas for insights to highlight the range of issues addressed in my study. Findings in the areas of educational psychology, general education, second language (L2) research and other innovation research will also be incorporated in this study.

## 2.4 Insights from Other Research Areas

This section starts with a review of the pertinent literature on educational innovation and compares the recurring themes that have emerged from this type of studies to those from the washback literature, discussing the similarities and differences that they share. It then offers a brief discussion of the definition of the core concepts related to the teacher factor under study and probes into the relationship each has with teacher instructional behaviors. Next, it draws on these concepts and the theoretical models developed based on these concepts and uses them as a general



theoretical framework for investigating teacher beliefs and behaviors. After that, it broadens the theoretical base of the study to include constructivist theory, which has an important influence on teaching and learning processes. The primary objective of this section is to demonstrate how a multidisciplinary conceptual framework has been constructed for understanding the range of issues in my research. The research areas described here may contribute to the washback research foundation described above. My study is situated within this multidisciplinary context.

#### *2.4.1 Change and Innovation*

In light of the above review of the washback literature, to induce washback means to bring about the deeper, underlying level of perceptual and behavioral change on the part of a teacher. It seems that there are some common areas that testing innovation and educational innovation share. Thus, in what follows, two crucial notions, ‘change’, and ‘innovation’, will be dealt with and an integrated discussion will be offered on how to ensure the success of various types of educational change.

It is worth noting that there has been a considerable amount of research across a variety of disciplines dwelling on educational innovation or change. In this research area, some researchers perceive innovation and change as different processes (White, 1987), whereas others view them as synonymous. Nonetheless, in the current study, I am using the terms interchangeably. Among the vast amount of studies conducted, the most notable and extensive are Fullan (1982, 1993, 1999, 2001), Hargreaves et al. (2001), Kennedy (1987, 1988), Markee (1997), and Woods (1996, 2003). Drawing on their own experiences in carrying out educational reforms, they illuminate the kinds of problems of innovation that teachers might confront and offer significant insights into how to cope with educational change. Also worth noting is that there is obvious agreement among these researchers regarding the central role of the ‘teacher factor’ in educational change. That is, teachers are those whom the success of educational change largely depends on. For example, White (1987) views them as one indispensable condition for change. From his discussion, unless teachers involved agree on the need of innovation, the innovation is unlikely to succeed. Thus, owing to the importance accorded to the ‘teacher factor’, there is little doubt that the implementation of education change essentially involves teacher change. But before we start looking in detail at what innovation research has revealed about how to effect teacher change, it would be informative to make a comparison between the patterns

that have emerged from washback studies with those from innovation studies. Such a comparison may help to enhance our understanding of the complexities underlying change in teachers' thinking processes as well as their instructional practices.

#### *2.4.2 Themes, Rules and Principles Common in Innovation*

Although other types of educational innovation and testing innovation are generally regarded and treated as separate issues, some themes, rules and principles are not unique to language testing innovation, but are representative of educational innovation in general. Some overlapping features are identified and discussed here.

One of them concerns teachers' resistance to change. Research evidence indicates that it is commonplace for innovation efforts to meet with resistance and criticisms, although the degree and intensity of resistance varies between different teachers and the causes of such resistance tend to be different, as well. As Markee (1997) indicates, many pedagogical innovations that policy makers perceive to be rational, necessary, and easy to implement may be opposed by teachers because, from their perspective, these innovations seem entirely irrational, unnecessary, and impossible to implement. This phenomenon has also been noted by researchers from different research contexts (Fullan, 1982, 1993; Markee, 1987; Tatto, 1998; Woods, 1996). The amazing similarity in findings from different contexts has led Fullan (1982, 1993) to conclude that dissatisfaction with or resistance to educational reform is a worldwide phenomenon. Woods (1996) accounts for this phenomenon saying that such reactions from teachers have much to do with their goals, intentions, assumptions and beliefs. From a similar perspective, Freeman and Richards (1993) maintain that the introduction of an approach will be in competition with well-established theories of teaching and learning which are the product of previous teaching and learning experience, prejudices, and beliefs. Also similarly, Markee (1997) states that the resistance may result from the different views teachers hold as to what change should be implemented and how it should be done. Obviously, a consensus exists that teachers' resistance is related to the beliefs they hold about how to teach. It must also be noted that these researchers also ascribe teachers' resistance to the use of power-coercive strategies, approaches that impose some form of change on teachers. In their view, such strategies can create blocks in the natural evolutionary processes. Hence, according to Kennedy (1987), there has emerged an obvious need to openly discuss and resolve any differences in the change process.

Another overlapping theme involves the complexity and unpredictability of change (Woods, 2003). It is widely acknowledged that change is a complex phenomenon. In Fullan's (1999) terms, it is non-linear, loaded with uncertainty. The challenges and difficulties entailed by such complexities are widely discussed by researchers in various fields. Markee's (1997) study reports on an innovation project in language education. It involves engaging ESL teachers in developing new materials, methodological skills, and values. Despite the conscious efforts exerted by him and his colleagues to manage the innovation, the project has failed to achieve intended results. The researcher concludes that innovation is an inherently messy, unpredictable business. This feature of unpredictability implicit in the change process is also alluded to by Hargreaves et al. (2001). Their study focuses on twenty-nine teachers within the context of curriculum reform in Canada. The findings show that the task to integrate the changes into teachers' practice is extremely demanding and difficult. However, although the researchers reinforce the idea that any change strategy should take this complexity into account, little explanation is provided as to what has led to such complexity.

The next recurring theme which has emerged relates to individual differences in terms of teachers' attitudes, beliefs and actions. That is, not all teachers experience the same type or degree of change. Kennedy (1988) and Markee (1997) have found that teachers vary in the amount of motivation, commitment or enthusiasm they show for an innovation. Meanwhile, Markee (1997) has also discovered that the time it takes for individuals to reach a decision varies considerably from person to person. "Some adopters embrace a given innovation relatively quickly, others need more time" (p. 57). In addition, Woods' (1996) study demonstrates that the teachers' interpretations of a curriculum and the beliefs, assumptions and knowledge that underlie these interpretations differ in important ways and are always evolving. Furthermore, the differences between 'expert' and novice teachers in their reactions to teaching tasks noted by Shohamy (1993) are also captured by Nespor (1987). The likely reasons researchers offer for the above-mentioned differences are previous life and schooling experiences, and education in pedagogy (Markee, 1997; Richardson, 1996; Clift, 1987; Grossman, 1990; Grossman & Richert, 1988).

The last overlapping theme identified relates to discrepancies between teachers' actual behaviors and their claimed beliefs about teaching. Like the results from washback studies, research findings from innovation studies (Markee, 1997; Woods,

1996) reflect that what teachers perceive they do is not necessarily congruent with their classroom practice. Woods (1996) discovers that a belief articulated in the context of an interview is not likely to be grounded in actual behavior. In a review of the ELT research, Karavas-Doukas (1996) finds that similar results have been reported by several different studies (e.g., Burns, 1990; Long and Sato, 1983; Mitchell, 1988; Nunan, 1987). That is, while most teachers profess to be following a communicative approach, in practice communicative classrooms are rarely found. These findings have been confirmed by studies concerning reforms in the teaching of mathematics. In Ernest's review of mathematics teacher beliefs research, he notes that teachers' enacted models of teaching and learning differ from their espoused or stated models (1989). Several case-studies in this field have shown such a mismatch. Why does this disparity between their expressed attitudes, beliefs or theory and their actual classroom practice exist? The question will be addressed later in this chapter.

Apart from the above overlapping themes, there are some other features implicit in innovation that testing researchers ignore. The most salient one is related to the evolving and dynamic nature of change. As this is a missing area in washback studies, in what follows it will be extensively discussed. In light of the above studies, we realize that the majority of problems encountered in other types of innovation are no different from those that washback studies have confronted.

Given the striking similarity in findings from different areas of research, these recurring features should receive proper attention during the design and implementation of an innovation project. As was summarized earlier, the implementation of the type of testing innovation under study essentially involves teacher change, and fostering teacher change involves engaging teachers in the more abstract tasks of developing their methodological skills and adopting new approaches to teaching. Underlying this methodological change should be a shift in teachers' beliefs and practice. Nevertheless, the problem is that both washback and innovation research shows that many attempts at educational change have failed to achieve the intended results. This reflects the fact that there are some superficial changes with the core of the problem – teacher beliefs and teacher practice – left untouched. Such a result raises the question of how we can effect change in teacher beliefs and teacher practice. How do teachers adopt and incorporate change into their pedagogical practice? In order for us to address these questions and interpret the above themes and patterns as well as elucidate the lingering problems such as the complex features of

innovation, we need to build up a theoretical framework as a basis for further discussion. Therefore, in the following section I will lay out this theoretical framework by gleaning and synthesizing the current research on teacher change, especially issues concerning teachers' beliefs and their relationship to practices.

### 2.4.3 Basic Concepts

With the developments of cognitive psychology, research on teaching has shifted from a focus on teacher behaviors to an emphasis on teacher thought processes (or teacher cognition) and their link to teacher practices. Over the past two decades, there has been increasing interest in the influence of teacher beliefs, knowledge and experiences (referred to as the 'teacher factor' in this study) on teacher practices (Borg, 2003, 2006; Crandall, 2000; Ernest, 1989; Freeman, 1996, 2002; Johnsons, 1999; Kalaja, 1995, 2003; Nespor, 1987; Richards & Lockhart, 1994; Shulman, 1986a, 1986b, 1987; Richards, 2008; Richardson, 1996; Thompson, 1992; Woods, 1996, 2003, 2006). Borg (2003, 2006), Crandall (2000), Freeman (2002), Johnsons (1999), and Richards (2008) have reviewed much of the relevant research, and outlined the research inquiries and trends in the general education and language education literature. They also elaborate on the myriad notions that are related to teacher perceptual and behavioral change. Relevant to this study are the notions of the 'teacher factor' (teacher factors), beliefs (belief systems), teacher thinking (teacher cognition), and knowledge (knowledge base).

To get a clear picture of the core assumptions of these notions, we need to clarify these basic concepts.

#### 2.4.3.1 The 'Teacher Factor'

The concept, the 'teacher factor', has made its appearance in ELT in Bailey et al. (1996). However, it may mean different things to different researchers. In ELT and general education, there is a widely accepted assumption that teacher internal attributes such as beliefs, assumptions, knowledge and experience make up the 'teacher factor' and this 'teacher factor' plays a powerful role both in determining teachers' perceptions of teaching and shaping their practices or actions in teaching (Borg, 2006; Richards, 2008; Woods, 1996).

As teacher beliefs, knowledge, and experiences are held to be the principal components of the 'teacher factor' which constitute relevant notions for understanding

classroom practices, they are discussed in more detail below.

#### 2.4.3.2 *Beliefs*

The term ‘beliefs’, as used in Richardson (1996), refers to “a subset of a group of constructs that name, define, and describe the structure and content of mental states that are thought to drive a person’s actions” (p. 102). What deserves mention is that also defined by Richardson as belonging to this set of constructs are terms such as attitudes, conceptions, perspectives, perceptions, orientations, theories, and stances. According to him, these terms can be used synonymously. In addition, “beliefs” is also termed “belief systems” (see Abelson, 1979). Nevertheless, as pointed out by Woods (2003), the prevailing concepts “provide an overly simplistic representation of what beliefs are” (p. 202). Rather than viewing beliefs as “a set of discrete static entities” within the individual (Woods, 2003, p. 201), modern L2 researchers (Barcelos, 2003; Kalyja, 1995; Woods, 2003) take more of a constructivist perspective on the study of beliefs. From such a perspective, beliefs are characterized as variable and evolving rather than stable (Kalyja, 1995; Woods, 2003), interconnected rather than discrete (Woods, 2003), complex rather than linear (Barcelos, 2003). Furthermore, these researchers underscore the social dimension and contextual nature of the concept in addition to its cognitive dimension. For instance, they delineate this concept as socially constructed (through specific instances of social interaction), dynamically interrelated (with other aspects of an individual’s cognitive processes such as knowledge and motivation, etc.), and embedded within sets of beliefs forming a multilayered web of relationships. It should be noted that here context is regarded as a crucial notion, as well. As illustrated by Barcelos (2003), beliefs are not characterized as internal traits but as related to teachers’ contexts and goals.

#### 2.4.3.3 *Teacher Thinking*

Another concept concerns teacher thinking, or teacher cognition. As Borg (2006) has noted, in the past two decades, “there has been a surge of research interest in language teacher cognition – what language teachers think, know and believe – and of its relationship to teachers’ classroom practices” (Borg, 2006, p.1).

Richards (2008) analyzes the role of teacher cognition and argues:

From the perspective of teacher cognition, teaching is not simply the

application of knowledge and of learned skills. It is viewed as a much more complex cognitively-driven process affected by the classroom context, the teachers' general and specific instructional goals, the learners' motivations and reactions to the lesson, the teacher's management of critical moments during a lesson. At the same time teaching reflects the teacher's personal response to such issues, hence teacher cognition is very much concerned with teachers' personal and 'situated' approaches to teaching. (Richards, 2008, p. 167 )

Ernest (1989) draws a distinction between two aspects of the psychology of teaching. They are teacher's thought processes (e.g., planning, interactive decision-making and reflection), and thought structures (e.g., teachers' knowledge, beliefs and attitudes). Based on his view, these two aspects are closely interrelated. In practice, the permanent but ever-changing and growing body of knowledge, beliefs and attitudes stored in teachers' minds are sources of the constructs through which the teacher's thought processes operate. For instance, teachers' knowledge of pedagogical skills is assumed to be able to provide the basis for their thought processes before, during and after teaching.

#### 2.4.3.4 *Knowledge or Knowledge Base*

In addition to the above three concepts, another notion that is considered to determine teacher practice in the classroom is teachers' knowledge or knowledge base (Ernest, 1989; Reagan & Osborn, 2002; Shulman, 1987; Woods, 2003). That is, a teacher needs a substantial knowledge base in the subject in order to plan for instruction and guide his or her practices. Following an examination of the literature in cognitive psychology and cognitive science, Woods (2003) illustrates that there are two types of knowledge – declarative knowledge (factual knowledge) and procedural knowledge (action-related knowledge). The latter is interpreted by him as a tacit and unconscious knowledge which guides our actions. From a constructivist perspective, the concept of knowledge is perceived by Woods (2003) as parallel to the concepts of beliefs in terms of their genesis and evolution. In his view, rather than residing in an individual, they are both seen as being actively constructed over time through interactions with a social group. With respect to the distinction between knowledge and beliefs, he posits that compared to knowledge, beliefs are relatively resistant to change. According to him, knowledge can be taken as a subset of beliefs. In contrast, the rationale Abelson (1979) provides us with for such a distinction is that beliefs serve as a means of defining goals and tasks, whereas knowledge systems come into

play where goals and the paths to their attainment are well-defined.

It should further be noted that despite their distinctive differences, the two terms, beliefs and knowledge, remain somewhat unclear or blurred in the literature. However in this paper, I will follow researchers such as Abelson (1979), Nespor (1987), Richardson (1996) and Woods (1996, 2003) and term them differently.

#### *2.4.3.5 The Relationship Between Teacher Beliefs and Teacher Practices*

As can be noted in the literature, most teacher cognition research attempts to link teacher beliefs to teacher practices.

Apart from Fullan (1982, 1993), there is almost a consensus among researchers with respect to the relationship between change in practices (or behaviors) and change in beliefs. From Fullan's perspective, people will often change their behaviors before they change their beliefs. However, his view is challenged by a number of other researchers (Kenney, 1988; Nespor, 1987) who insist that significant teacher change only occur if teachers change their deep-seated attitudes and beliefs. In Kenney's (1988) view, behavioral change is a surface phenomenon, while a change in deep-seated attitudes and beliefs is deeper and more complex change. This can be taken to mean that although the intention to change a test is targeted toward teacher behavior, to bring about such a change in behavior would necessitate a change in teacher beliefs. While Fullan's (1991) point appears to be consistent with that of others in that the relationship between beliefs and behavior is reciprocal or interactive, their interpretations of the implication of such reciprocity differ. The view Fullan (1982) held is that "trying new practices sometimes leads to questioning one's underlying beliefs; examining one's beliefs can lead to attempting new behavior (p. 287)." In contrast, the general point of other researchers is that "beliefs can drive actions; however, experiences and reflection on action may lead to changes in and/or additions to beliefs" (Richardson 1996, p. 104). Nevertheless, contrary to either of the above stances which assumed that a linear relationship exists between an individual's beliefs and his or her behaviors, Woods (2003) emphasizes that beliefs and action are related in complex and indirect ways. He illustrates that sometimes beliefs affect teacher practice in the classroom and other times they do not. Considering the intricate relationship between teacher beliefs and teacher practice, it is obvious that an understanding of teacher classroom practice rests on a clear understanding of teacher



beliefs.

In brief, the conceptualization of the above terms has evolved over the past few years. One big breakthrough in this new line of research lies in the fact that a constructivist, and process-oriented perspective has been adopted by L2 researchers. This perspective has provided fresh theoretical insights into the investigation of the complex relationships between teacher beliefs and practices.

Due to the difference in perspectives from which researchers view these relationships, there is considerable variation among their definitions of the terms (e.g., beliefs, teacher cognition, theories, knowledge, principles, practical knowledge, and maxims). Irrespective of the inconsistency in the definition of these terms, one common view shared by researchers is that there is a clear relationship between teachers' beliefs and knowledge, and their instructional behaviors. In this regard, it would still be informative for us to examine how these concepts, core assumptions and theories of beliefs inform and promote teacher pedagogical practice.

#### *2.4.4 Conceptual and Theoretical Framework Guiding this Study*

##### *2.4.4.1 Shulman's (1987) Model of Teachers' Knowledge*

Shulman (1987) has proposed a model of teachers' knowledge. As a result of extensive experimental work, Shulman (1987) outlines features of different forms of knowledge that serve to distinguish them from beliefs. In his conceptualization, four dimensions of knowledge constitute the major components of the knowledge base for the classroom teacher. They are: content knowledge, general pedagogical knowledge, curriculum knowledge and pedagogical content knowledge.

Shulman (1987) places content knowledge (which is termed as 'Knowledge about' by Richards, 2008) at the top of the teacher knowledge base. This means that a teacher must acquire subject matter knowledge in order to teach it effectively. In the context of English instruction, this can be interpreted to mean that an ESL teacher must develop a high degree of competence in the target language in terms of four language skills as well as an awareness of the political and sociocultural aspects of language and language use (Reagan & Osborn, 2002). Therefore, this content knowledge provides an essential foundation for the teacher's pedagogical knowledge and skills for teaching English. Pedagogical knowledge, as is interpreted by Fang (1996), is concerned with the dimension of teaching such as how ideas are best presented and formulated in order to make them comprehensible to others. A manifestation of this

knowledge in ESL teaching is how to represent language topics and ideas, and how to organize and manage a language class. Among the skills that can be considered to be core pedagogical knowledge are instructional planning, lesson presentation skills, questioning skills, interpersonal communication skills, classroom management skills, and knowledge of evaluation (Cooper, 1990; Reagan & Osborn, 2002). This is what Ernest (1989) and Richardson (1996) term as practical knowledge and what Richards (2008) terms as ‘knowledge how’. According to Richards (2008), this dimension of knowledge is often tacit, experiential (gained through experience) and contextual. It contributes to the formulation of teachers’ working principles that guide their teaching behaviour and functions as the source of teachers’ practices and understandings. He argues succinctly that this type of knowledge should be made explicit if teachers are to consider changes in practices. Curriculum knowledge includes knowledge of textbooks, examinations, tests and syllabuses (Ernest, 1989). Pedagogical content knowledge is basically a powerful combination of content, pedagogical, and curricular knowledge. According to Freeman (2002) and Richards (2008), this type of knowledge is central to teacher pedagogical knowledge. However, they argue that it derives from neither discipline-based content nor training-based pedagogy. Overall, such constructs as teacher’s practical knowledge, pedagogic content knowledge, and personal theories of teaching are conceptualized by Richards (2008) as components of our understanding of teacher cognition (Richards, 2008).

According to Freeman (2002), the knowledge-base is largely drawn from other disciplines, but not from the work of teaching itself.

#### 2.4.4.2 Woods’ (1996) Model (A Model of BAK Networks)

Of the number of models that have emerged, Woods (1996; 2003; 2006) is most notable. The study is significant in that it extends the discussion in the education literature related to concepts of ‘knowledge’ which implicitly include beliefs, by providing us with a theoretically and empirically grounded model of beliefs, assumptions and knowledge (BAK). In this model, a constructivist view of knowledge and beliefs is central. The model (draws on a constructivist view of knowledge and beliefs, which) adopts a process-based and dynamic orientation to research into teacher beliefs. As illustrated by Woods (2003), the three concepts are not “treated as being qualitatively different, but rather as extremes on a spectrum” (p. 205). Here, the term “assumption” is taken to mean provisional acceptance. As exemplified by Woods

(2003), beliefs are not separable from other aspects of a teacher's cognitive processes, but integrated in a larger dynamic model of thought and action. This model differs in one important way from much of the research that has been reported in the educational literature on teacher thinking. Unlike those who maintain a discrete view of teachers' beliefs, Woods (1996) argues that each teacher has an individual system of interwoven BAKs. He assumes that such interrelated networks underlie everything that teachers do and say. To be specific, they play an important role not only in teachers' interpretation of events, but also in their organization of thoughts, decisions and actions. In light of this characteristic of interconnectedness, a change in any one aspect will have an effect on other aspects.

What also distinguishes Woods (1996) from other studies is its longitudinal focus and dynamic view of beliefs. Rather than viewing teachers' BAKs as remaining unchanged or static, Woods (1996) asserts that they evolve over time. He illustrates that it is the interplay of the teachers' evolving BAKs that leads to teacher change as well as to curricular evolution. Also, it is owing to the evolving nature of change in teachers' BAKs that something that looks to be a problem may, in the long run, turn out to be to everyone's benefit. However, he suggests that there may be a "lag time" for cognitive changes to take place. In his sense, beliefs do not change easily and the characteristics of the new state brought about by the change need time to be absorbed.

Another important difference between Woods' (1996) study and those of others is that it is a participant-centered study. Instead of seeking to understand teaching from the perspective of a researcher or observer, Woods (1996) has sought to understand teaching from the point of view of the participants – eight ESL teachers. He has made an in-depth analysis of their planning and interpretative processes. Such an approach to research enables him to provide reliable and convincing results. One crucial finding is that teachers' interaction with the interviewer seems to have some effect on teachers' behavior and verbalizations as well as the data collected. Based on this result, he concludes that "the constant discussion of their behavior may have caused more evolution in teachers' thinking and behavior than would have otherwise occurred" (p. 37). Such a finding is corroborated by other studies (Karavas-Doukas, 1996; Pajares 1992; Richardson, 1996). Similar to Woods, Richardson (1996) reports that he and his colleague conducted interviews with teachers and as a result many of the teachers later acknowledged that "the interview, in combination with reading the transcription, significantly affected their approach to reflection and change" (p. 107). This finding

has led the researchers to suggest that the research methods themselves can give an opportunity to teachers to become more aware of their beliefs.

#### *2.4.4.3 Theories Concerning Teacher Conceptual and Behavioral Change*

Some of the above characteristics are also noted by other researchers (Kennedy, 1988; Markee, 1997; Nespor, 1987; Posner et al., 1982; Richards, 2008). In a similar way, Markee (1997), Posner et al. (1982), and Richards (2008) perceive change in teachers' beliefs as a process to be undertaken. According to Markee (1997), such change is incremental and time-consuming in that it involves a sequence of decisions that are made over a period of time. Similarly, Posner et al. (1982) state that "a significant change is a gradual and peaceful process of taking an initial step toward a new conception by accepting some of its claims and then gradually modifying other ideas" (223). Consistent with this view, Kennedy (1988) also insists that innovation should be regarded as a process which undergoes the stages of goal-setting, planning, implementation or completion. In his words, not only does it take time to implement, but it always takes longer to implement than originally anticipated.

Furthermore, other researchers are also aware of the crucial role of teacher experience in the evolution of teacher beliefs (Freeman 1996; Freeman & Richards 1996; Richards & Lockhart, 1994; Urmston, 2003). Sufficient evidence suggests that among the factors that shape teacher beliefs, teachers' previous life and school experiences both as students and as teachers are most powerful (Brookhart & Freeman, 1992; Markee, 1997; Nespor, 1987; Pajares, 1992; Richardson, 1996; Tatto, 1998; Thompson, 1992; Urmston, 2003). In a review of teacher beliefs research, Pajares (1992) summarizes that "beliefs are formed early and tend to self-perpetuate, persevering even against contradictions caused by reason, time, schooling, or experience" (p. 324). In this regard, Woods' (1996) conceptualization of teacher thinking seems well-grounded.

Through his research, Woods (1996) also notes that aspects of the teachers' BAK networks may take many different forms. They differ considerably from teacher to teacher. That is, each teacher's system differs from other teachers' systems not only in terms of its individual 'components', but also in terms of the interrelationships among the BAKs.

What also deserves mention is that Woods offers convincing explanations for the discrepancy between teachers' beliefs and their statements about beliefs, and between

teachers' articulated beliefs and their behaviors. One explanation is that beliefs may not always be entirely consciously accessible to teachers. In his view, although teachers have had the experiences, they may not have categorized and labeled them. He accounts for the discrepancy by suggesting that it may relate to the dynamic evolution of a teacher's BAK over time and to the developing expertise of the teacher. Another explanation is that teachers may, in responding questions about generalized beliefs, answer according to what they would like to believe, or what they would like the audience to think they believe.

Based on the analysis provided by Ernest (1989), there are three key causes for the mismatch. First, if teachers' espoused beliefs are not integrated with other knowledge and beliefs, especially pedagogical knowledge, only a limited basis for their enactment exists. Second, a mismatch may exist between teachers' awareness or judgment of his or her beliefs and his or her real beliefs. Third, the mismatch may result from the powerful influence of the social context (e.g., the expectations of colleagues and superiors). Obviously this refers to the result of the power indicated by Woods (1996).

Moreover, rather than focusing on the negative impact of such a mismatch, Woods (1996) suggests that the discrepancy can provide an impetus for subsequent decisions designed to address this discrepancy and function as an adjustment in the on-going teaching.

It is notable that Woods (1996) emphasizes the role of teachers' BAK networks in promoting educational change. According to him, there are generally two major means by which teachers' BAKs evolve. One is through experience and resolution of conflicts, namely the individual way; the other is through interaction with colleagues, curricular guidelines and directives, namely the social way. He posits that in order for a change to take place in teacher beliefs, there needs to be readiness on the part of the teacher to make a change in his or her beliefs.

One way of facilitating teachers' readiness, based on Woods (1996), is simply through developing teacher awareness about his or her BAK. Based on his account, an awareness of one's own BAK may make it easier for a teacher to accept others', to understand how they differ, and to decide that the difference can be worked through in areas of conflict. A suggestion made by him to increase teachers' awareness is that some "input" related to beliefs be offered to them. Research, or systematic inquiry, and critical reflection on experience are the principal sources for constructing a

knowledge base.

#### 2.4.4.4 *Change in Teachers' Pedagogical Beliefs*

In addition to Shulman (1987) and Woods (1996; 2003), theoretical insights can be gained from other studies addressing teacher beliefs (Ernest, 1989; Fang, 1996; Richards, 2008; Thompson, 1984; Yates & Muchisky, 2003). A similar point made by these studies is that teaching is essentially a cognitive, behavioral activity involving a set of beliefs such as teachers' perceptions of teaching, their conception of what is meant by learning or how languages are learned, and their views of language, etc. This set of beliefs is not only considered to have significant bearing on the way teachers approach teaching, but also thought to have a powerful influence on learners' perceptions as well as the extent to which educational innovations take hold. Like Woods (1996), the interrelated nature of these beliefs has received their attention. For example, as noted by Thompson (1984), Ernest (1989), Glaser and Bassok (1989) and Glaser and Silver (1994), teachers' beliefs about language are reflected in their perceptions of the teaching and learning of English, and teachers' beliefs about testing tend to follow their beliefs about teaching and learning. It is worth noting that of the interrelated beliefs, teachers' views of what is meant by learning is placed at the core of whatever the language teacher does and wherever the language teacher is situated (Ernest, 1989; Reagan & Osborn, 2002; Thompson, 1984; Yates & Muchisky, 2003). In other words, teachers' views of learning may affect not only the way English is taught (teacher pedagogy), but may also have a critical impact on the learner's experience of learning English. Given the importance ascribed to the influence of teachers' conceptions of learning on the way they do in the classroom, it is crucial to gain an insight into what is involved in the process of learning.

A trend that begs for our attention is that a social constructivist perspective, as a theoretical orientation, has been adopted not only in research areas such as psychology, cognitive science, and education, but in research on Second Language Acquisition (SLA) as well. Researchers increasingly draw on a constructivist epistemology as an important basis for their discussion of all aspects of teaching-learning process including teacher classroom practice (Keiny, 1994, Richards, 2008). One important facet of this epistemology relevant to the discussion of teacher beliefs is concerned with how teachers perceive learning and knowledge. From a constructivist point of view, learning is conceptualized as an active and constructive process, through which

knowledge is actively constructed rather than being passively received by the learner from the teacher. Such a way of learning, with its emphasis placed on learner involvement in the learning process, is also construed as experiential learning by Nunan (1999) and as being learning-centered by Samuelowicz and Bain (1992; 2001). With the emergence of such a perspective on learning, a view of learning English as the passive transmission/reception of knowledge (a transmission-based view of language learning) is seen as being incompatible with a view of learning English as the active construction of knowledge as well as a way of interpreting meaning (an interpretative or view of language learning) (Ernest, 1989; Nunan, 1999; Richards, 2008; Williams & Burden, 1997). It is interesting to note that the distinction between these two perspectives of learning is used by some researchers (Samuelowicz & Bain, 1992; 2001) as a major criterion for judging whether conceptions of teaching held by teachers are teacher-centered or learner-centered. According to them, teacher-centered teaching conceptualized as a process of imparting, or transmitting information indicates a lower level conception whereas learner-centered teaching seen as encouraging an active students' role represents a higher conceptions of teaching.

Another important facet of this epistemology concerns cognitive development. From a sociocultural perspective, cognitive development is a socially mediated activity which involves a process of internalization (Vygotsky, 1978). This process of internalization is also described as a transformative process in which a person's activity is initially mediated by other people or cultural artifacts but later comes to be controlled by the person as he or she appropriates resources to regulate his or her own activities. Also associated with this perspective is a view that knowledge is socially shared before it is recreated and appropriated by any one individual. Taking such a constructivist stance, Richards (2008) and Woods (2003) elaborate on knowledge as being "situated" in actual interactions and "distributed" through a social group rather than residing in an individual. Johnson and Golombek (2002) claim that for teachers, this process of internalization may appear in how they understand aspects of their teaching practices as well as in the actual practices they engage in during classroom instruction.

Nunan (1999) claims that the constructivist principles of learning are reflected in a great deal of thinking in L2 teaching and learning. In his view, the three interrelated and widely discussed concepts, communicative language teaching (CLT), learner-centered instruction, and task-based language teaching are grounded in an

interpretative view of language learning. Similarly, Woods (2003) reminds us that the principles of SLA are rooted in the constructivist theory of learning.

Insights from the constructivist theories of learning discussed above have led to some fundamental changes in how language and teaching are conceptualized. Nunan (1999) highlights the philosophical shift from a transmission-based view of language teaching and learning to an experiential view. Following his explanation, subsumed under this philosophical shift are not only two distinct conceptions of language, but also two contrasting ways in which teachers go about the business of language teaching. Based on his interpretation, the transmission model sees language as a system of rules, while an experiential philosophy holds that language is a system for the expression of meaning. He further posits that this new view of language has a powerful impact on teacher methodology. The assumption provided by him is that “if the aim of language teaching is to help learners develop skills for expressing different communicative meanings, then surely these ought to be reflected in classroom tasks and activities” (p.10). According to him it is this insight that has led to the task-based approaches to language teaching.

Nunan (1999, 2001) also provides us with a summary of the changes to teacher classroom practice brought about by experientialism. It is presented in Table 2.



Table 2 *Comparison Between Transmission Teaching and Experiential Teaching*

(Note: From Nunan, D. (2001). Tradition and change in the ELT curriculum. *Foreign Language Teaching and Research*, 33(4), 52. Copyright 2001 by Foreign Language Teaching and Research. Reproduced with permission.)

	<b>Transmission Teaching</b>	<b>Experiential Teaching</b>
<b>Approach to teaching (methodology)</b>	Learners are taught about language and its rules. Learning facts about language.	Learners are actively involved in using. Language Learning through doing
<b>Approach to language</b>	Grammar is taught as rules to be memorized.	Grammar and vocabulary are taught communicatively, so learners can use them to express different communicative meanings.
<b>Classroom organization</b>	Learners sit in rows facing the teacher and spend most of their time repeating what the teacher says. They don't learn how to express their own ideas.	Learners work in small groups and pairs, learning skills of co-operating with others and how to express their own opinions and feelings.
<b>Teacher's role</b>	Providing mainly frontal instruction	Facilitating learning (largely in small groups)
<b>Learner's role</b>	Relatively passive recipient of information; mainly individual work	Active participation, largely collaborative small groups

The above summary presents to us a contrast between the two different types of teaching. One significant implication of this summary is that it can profitably be taken as a tool or framework for us to examine whether the pedagogical practices carried out by teachers are fully compatible with the constructivist learning theory. To be specific, it may help us discern teacher-centered instruction from student-centered instruction.

Another notion that is worth mentioning here is “constructivist teaching”. Based on the interpretation offered by Winitzky and Kauchak (1997), “constructivist teaching” typically involves more student-centered, active learning experiences, more student-student and student-teacher interaction, and more work with concrete materials and in solving realistic problems. Nunan's (1999) presentation of experiential teaching seems to correspond to such a style of teaching. Rather than teaching being viewed as the process of transmission of knowledge, a constructivist perspective sees it as creating conditions for “the co-construction of knowledge and understanding through social participation” (Richards, 2008, p. 169).

#### *2.4.5 Implications of the Above Insights for Washback Research*

In summary, the theoretical perspectives as well as the research evidence presented above cast new light on the recurring themes that have been previously discussed. They also form a useful framework which helps us conceptualize the whole teaching process. It is beyond doubt that Woods' (1996) study as well as other studies (Ernest, 1989; Fang, 1996; Nunan, 1999; Reagan & Osborn, 2002; Richards, 2008; Samuelowicz & Bain, 1992, 2001; Shulman, 1987; Thompson, 1984; Williams & Burden, 1997; Yates & Muchisky, 2003) has a wide range of implications for our understanding of the role of the 'teacher factor' in washback. Not only do they provide us with a comprehensive understanding of the complex reality of innovation, but they also offer us a different way of thinking about notions such as teacher beliefs, knowledge, and experience (BKE) and their connection to teacher practice. Moreover, the interdisciplinary theoretical framework provides me with a broad set of conceptual tools for systematic investigations of teacher thinking and its relationship to teacher classroom practice. It also allows me to document and interpret the washback phenomenon in China.

Specifically, this theoretical framework can inform my study at least from three bases. First, washback studies may have overlooked the evolving nature of innovation as well as the dynamic aspects of teachers' BAKs that innovation researchers had underscored. Failure to adopt a dynamic perspective in viewing teacher change may consequently distort the results of the studies, for it may affect the way the data are collected and analyzed. One example drawn from washback studies to illustrate this is that a number of researchers have found it hard to make weighty claims, and thus made only tentative ones. It appears that these researchers may have failed to take into account the developmental characteristic of change. Since the research focus of the majority of studies is short-term, no conclusions can be drawn about long-term washback effects.

Second, the above studies have also offered us enlightening insights into how to look at and cope with conflicts, constraints, differences and discrepancies that have emerged from innovation. Wood's (1996) view regarding the operation of teachers' BAKs through resolution of conflicts is of practical significance here. Unlike researchers on washback who perceive conflicts, differences and discrepancies as forces that hinder the processes of change, he views them as an impetus for change and an indispensable part of teacher evolution. Extensive credit has been ascribed by

him to the role of teachers' awareness of these discrepancies. In view of the possibility that such an awareness can serve as a trigger that forces the teachers to fine-tune their perceptions and practices, this view is obviously of crucial value for my research. Meanwhile, considering the fact that teachers are not fully aware of their beliefs, there is a need for us to help them become more reflective. Underlying this need is the belief that teachers will benefit greatly if they have a chance to reflect on their own beliefs and behaviors.

Third, as the implementation of educational reforms including testing reforms calls for the conceptual change in teachers' pedagogical beliefs, teachers' perspectives about the nature of language and learning merit special consideration. In light of the interwoven characteristic of teachers' BAKs, a teacher cannot simply change one belief by itself, because each one is part of an interwoven network which includes many other beliefs. Based on Woods (1996), in order for one element to change, it means that its relationship to other elements has to change, and other elements will have to change too. This, with respect to washback research, can be interpreted to mean that in order for teachers to change their perceptions of tests, they need to change their perceptions of teaching and learning, and their perceptions of language as well, for all these beliefs are intrinsically interwoven. To be specific, teachers' beliefs of tests are likely to correspond to their beliefs of language teaching and learning. Meanwhile, their beliefs of language teaching and learning are likely to follow their conceptions of what is meant by learning as well as their beliefs what language is. Here, what we should note is that the relationship between beliefs of language teaching and beliefs of language learning is also interactive and interconnected. All these beliefs are crucial in the sense that they may not only influence how teachers conceptualize teaching but also affect the way they interpret and react to washback.

It is obvious that the multidisciplinary wisdom accumulated in different educational fields provides us with a much deeper and more coherent basis for understanding the role the 'teacher factor' plays in generating washback. Such a basis not only helps to clarify the complexity of the innovation process, but also helps to improve further innovation endeavors. Therefore, there is a need to apply these insights to washback research.

Nevertheless, although research on teacher thinking is now a popular topic of inquiry, we must be aware that it still has a lot of limitations (Cantu, 2001; Fang, 1996). A belief commonly shared among researchers is that reasoned discourses on beliefs are

scarce in educational literature (Pajares, 1982; Thompson, 1992). This is due, in large part, to what Pajares (1992) refers to as “definition problems, poor conceptualizations, and differing understandings of beliefs and belief structures” (p. 307). Such a lack of consensus on definitions in both beliefs studies and washback studies undoubtedly increases the complexity of the problem. In this regard, we need to take into account the limitations of beliefs studies and differentiate between claims or assertions on the one hand, and empirically grounded research results on the other.

This chapter started with an extensive overview of the washback research conducted both in the ESL context and in the Chinese (EFL) context. First, it examined the research on washback in language education and general education to clarify and summarize some basic concepts and theoretical perspectives related to the washback phenomenon. It then offered a discussion about the washback studies carried out in the Chinese context. Subsequent to that, it outlined patterns and themes related to the ‘teacher factor’, and covered pertinent methodological issues involved in this type of research. It then traced some of the trends and features emerging from other research fields such as language education, general education, psychology and other innovation research to see how much common ground different types of research share. After that, it provided a general conceptual framework in an attempt to highlight the overlapping patterns and themes that have emerged through the lens of this framework. Finally, it discussed the new meaning the insights outlined have brought to the research on washback.

After the brief introduction provided in Chapter 1 on the general context of the study, and the broad set of theoretical and methodological tools outlined from multiple sources in Chapter 2 for examining the research topic of this study, the next chapter presents a contextual framework of Chinese CE teaching, learning and testing.

## Chapter 3: The Educational, Sociocultural, and Historical Context of the Study

### 3.1 Introduction

This chapter focuses on the specific scene of the study and considers the overall local educational, sociocultural, and historical context in China. It starts with a brief description of the growth of the CE course, the CE curriculum, the CET as well as the reforms carried out in them over the past three decades or so. It, then, turns to look at some lingering problems confronting CE teaching and testing, and traces their roots to the Chinese educational, sociocultural and historical background. The purpose of this chapter is to set the scene historically so that a better understanding can be achieved about the overall local educational context where test impact is nurtured.

To achieve a thorough understanding of the Chinese educational context, I conducted an extensive review of the pertinent articles published in Chinese local social science journals. On the basis of such a review, in the following sections, I will sum up what I have found or identified.

### 3.2 Glimpse of Chinese CE Teaching and Testing

#### 3.2.1 “College English” Course – Stages of Development

As introduced in Chapter 1, CE is a required course for all tertiary-level non-English majors across China. As a rule, the course is offered to college students for two years.

Since the early 1980’s, the Chinese educational authorities have made consistent efforts to update ELT in China. As a result, the past two decades have witnessed a number of significant changes in CE teaching and learning. According to Cen (1999) and Yan (2002), CE teaching has undergone three stages of development. They are 1) the Restoration Stage (from late 70’s to mid 80’s when China initially carried out its opening-up policy and CE instruction was first introduced), 2) the Development Stage (from mid 80’s to early 90’s when a large number of well-trained undergraduates joined the teaching team and the notion of ‘student-centeredness’ was introduced into the Chinese ELT field), and 3) the Improvement Period since 1995, during which a higher standard was set for CE teaching and learning. While Liu’s (2002) classification is similar to that of Cen (1999) and Yan (2002), he labeled the third stage as the “Stage of Stability”.

It should be noted that coupled with CE growth are changes in its curriculum and

the CET. Both of them have seen a number of adjustments over the past few years.

### 3.2.2 Curriculum

Since the introduction of the CE course, a uniform curriculum has been prescribed by the Chinese ME as a guideline for CE instruction. Over the past three decades or so, the CE Curriculum (CEC<sup>3</sup>) has undergone three revisions, which took place in 1980, 1985 and 1999 respectively. The three versions differ from one another in their specifications of objectives and requirements for CE teaching and learning. The 1980 version places reading as the sole competence required for CE teaching and learning and clearly set students' reading speed at 17 words per minute (wpm) (Gu, 2005). Compared to the 1980 version, the 1985 version CEC is more concrete. It specifies that upon completion of the CE course, students are expected to demonstrate three levels of competence, specifically a relatively high level of competence in reading (with an intensive reading speed at 50 wpm; a fast reading speed at 90 wpm), an intermediate level of competence in listening and translation for students of science and engineering (with a listening speed at 120 wpm) and a basic level of competence in writing and speaking (with a writing speed at 100-120 words within 30 minutes) (CEC, 1985). In addition, the 1985 CEC also provides an inventory of vocabulary (composed of 4000 words) for college students, from which comes the source of vocabulary to be tested on the CET, and a description of the criteria by which each component of the language skills is to be assessed. It can be seen that the 1999 version (CEC, 1999) merely differs from that of 1985 in that it has shifted from an emphasis on three levels of competence to one on two levels – a relatively high level of competence in reading and an intermediate level of competence in listening, speaking, writing and translation. Apart from the objectives, the 1999 CEC has also updated its vocabulary inventory (from 4000 words to 4200 words) and criteria for assessing language skills.

The following table gives a comparison of the three versions of CEC.

---

<sup>3</sup> The CEC and the CECR refer to different versions of the national CE curriculum. Since 2004, the national curriculum has changed its name from the CEC to the CECR.

Table 3 *Three Versions of CEC'*

Items	1980	1985	1999
<b>Vocabulary</b>		4000 words	4200 words
	Reading (17 wpm)	Reading	Reading
		Intensive (50 wpm)	Intensive (70 wpm)
		Fast (90 wpm)	Fast (100 wpm)
<b>Criteria for the Skills</b>		Listening (120 wpm)	Listening (130-150 wpm),
		Writing (100-120 w),	Speaking,
		Speaking	Writing (120-150 w), Translation

w: word

wpm: word per minute

It is notable that since the introduction of the 1985 Curriculum, the CE course has been divided into two stages. They are, as described by Gu (2005), the “Foundation Stage” and the “Specialized Reading Stage”. Upon completion of the “Foundation Stage” (normally before their second academic year has ended), the students are expected to sit for the CET 4. If they pass the test, they will be considered as having reached the basic compulsory requirements set by the CEC. Those who pass the CET 6 are regarded as having met the higher optional requirements (Gu, 2005).

### 3.2.3 *The CET*

Following the introduction of the curriculum, the CET made its first appearance in 1987. Up until now, it has been in operation for a little over 20 years. Designed by the NCETC, it is known as a national, large-scale standardized test for undergraduates of non-English majors in China. According to Yang and Jin<sup>4</sup>, the CET has been administered on behalf of the Chinese ME and the purpose of the test is to promote the implementation of the CEC, to assess college students' English proficiency, and to provide feedback information for teachers to improve their classroom teaching (Yang & Jin, 2001; Yang, 2003).

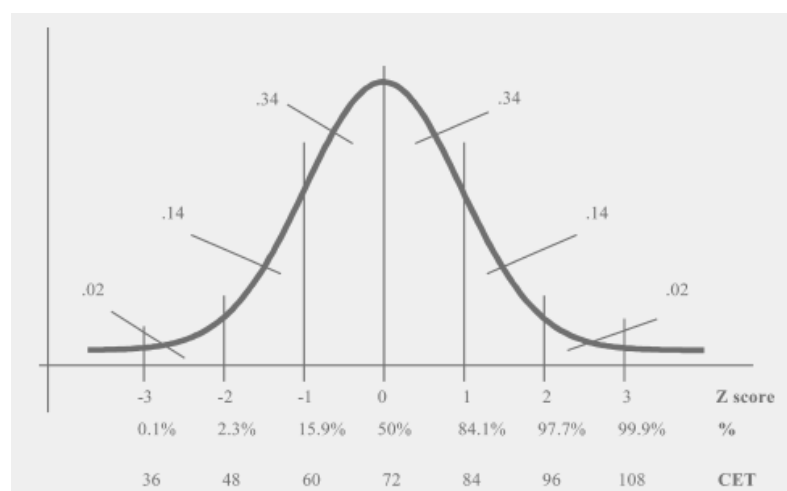
The CET (before 2006) took the form of the traditional one-hundred-point scale. As was put by Gu (2005), the test does not follow any particular set of textbooks used in colleges or universities. Based on Yang and Jin (2001), and Jin and Yang (2006), the CET has been designed as a “criterion-related norm-referenced test” in accordance with the CEC. The reason for its being called a “criterion-related norm-referenced test”, from Jin and Yang's (2006) explanation, is that:

<sup>4</sup> Yang and Jin have been directors of the NCETC.

Both the CET-4 and CET-6 adopt a normal score scale in the form of the traditional one-hundred-point scale with the mean being 72 and the standard deviation 12.

The reported score of the CET-4 conveys two items of information. First, it indicates whether a candidate has met the requirements of the NCETS<sup>5</sup> for Band 4 students. Secondly, it indicates the percentile position of a candidate in the norm group, which consists of about 10,000 college students from the top six universities in China (Beijing University, Tsinghua University, Shanghai Jiaotong University, Fudan University, University of Science and Technology of China, and Xi'an Jiaotong University). (Jin & Yang, 2006, pp. 23-24)

The normal curve of the distribution is shown in Figure 3.



*Figure 3. The CET score scale (Note: From 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. Copyright 2006 by Routledge. Reproduced with permission)*

Unlike other tests which normally adopt either the norm-referenced measurement or the criterion-referenced measurement in interpreting test scores, the CET combines the two. From Jin and Yang's (2006) interpretation, the CET bears the characteristics of a criterion-referenced test in the sense that it is intended to provide a description of a testee's performance with respect to the level of their language abilities of the pre-specified domain of Band 4. Meanwhile, the test is norm-referenced in the sense that the score of an examinee is interpreted in relation to the scores of other testees, for it is determined after it has been compared to the scores of those in the norm-reference group.

<sup>5</sup> The NCETS refers to the National College English Teaching Syllabus, the same as the College English Curriculum (CEC) I refer to in my study.



The test is administered twice a year (in January and then in June). According to Jin and Yang (2006), the number of testees has been on the rise over the years. In 2003, more than 9.15 million students sat for the test. As a rule, they have more than one opportunity to take the test.

The test is allegedly designed by incorporating features of both the structural approach, and the communicative approach (Gu, 2005). Approximately 75-85% of the test items are in Multiple Choice (MC) format (Gu, 2005).

The high-stakes nature of the CET is widely recognized by CE professionals. What is well-known is that students' lives are powerfully and directly affected by their performance on the test. To a large extent, their test scores will not only determine whether they will be granted a bachelor's degree, but will determine what kind of jobs they can secure upon graduation.

### *3.2.4 Relationship between the CEC and the CET*

When addressing the relationship between the CEC and the CET, Yang and Weir (1998) clearly state that the CET is intended to facilitate the implementation of the CEC. Gu (2005) buttresses their statement by saying that its content conforms to the CEC (Gu, (2005). Later, Yang (2003) further assures us that the purpose of the CET is to judge whether examinees have attained the standard required by the CEC.

### *3.2.5 Earlier Reforms in the CET*

Over the past decade or so, a series of attempts have been made by the ME to improve the CET in its format and content. The first major change took place in January, 1996 when some new forms of testing tasks (e.g., constructed response items) such as translation (from English to Chinese), spot and compound dictation, short-answer questions, cloze, and writing were added to its original format. Such a modification, from the test designers' perspective, was made with an attempt to mitigate the negative washback of the MC items of the test on CE teaching and testing. These items, nonetheless, only constitute a small portion of the entire test.

The second major change occurred in 1999 when the Spoken English Test (SET) was introduced (Gu, 2005). The test developers believed that the inclusion of the oral aspect of the language may call more attention from Chinese EFL teachers and students to the practical use of English. The chief goal for introducing the SET, according to the ME, was to increase an emphasis on students' oral competence (Gu,

2005). However, what is controversial is that not every student has the opportunity to sit for the SET, for only those who get scores above 80 on the CET written test are given this opportunity. Consequently, as Guo (2003) has pinpointed, neither teachers nor students are motivated to enhance the development of this skill.

### 3.3 Opposing Views on the CEC and the CET

It should be noted that there have been two opposing views as to the appropriateness of the CEC and the CET. The first position holds that the CEC and the CET are scientific, well-developed, advanced and practical (Li, 2001; Dong, 2003; Gu, 2005), while the second position maintains that, with their continuous focus on reading skills, they fall behind the development of the times (Han, 2002; Liu & Dai, 2003). A large number of Chinese EFL researchers and practitioners (Cai, 2003; Feng, 2000; Niu, 2001; Qian, 2003; Luo & Xiao, 2002; Shu, 2004) subscribe to the second view.

A similar thread of argument is made by Feng (2000) and Han et al. (2004). Based on Feng (2000), the CEC was developed under the principles of “structuralism” and it constrained the reform of teaching methodologies. This claim is supported and reinforced by Han et al. (2004) and Liu and Dai (2003). They elaborate on this saying that a special stress on mastering grammar and vocabulary is a good indicator of a structural curriculum. The vocabulary inventory provided in the CEC is held by them as giving a wrong message to the students. That is, as long as they have memorized the vocabulary list, they can pass the test.

Apart from the controversy over the CEC, there has also been a long-standing debate about the impact of the CET on CE teaching and learning (Chen, 1999; Donnelly & Yu, 2002; Gao, 2003; Han et al., 2004; Jing, 1999; Ma, 2000; Ma & Jin, 2000; Niu, 2001; Qian, 2003; Wu, 2003; Luo & Xiao, 2002;). One provocative article entitled “How can the pot of English be brought to a boil?”<sup>6</sup> is worth citing here (Chen, 1999). It was this article that set off the hot debate among EFL professionals about whether the CET constitutes a major constraint on ELT innovations in China. At the heart of this passionate debate are two fundamentally different convictions of the CET.

---

<sup>6</sup> The article was written in response to Jing (1999), an article entitled “Why can’t the pot of English teaching be brought to a boil?” published in Youth Daily.

On the one hand, in the eyes of its constructors (Jin, 2000; Yang & Jin, 2001; Jin & Yang, 2006; Yang, 1999, 2000, 2003) and leading defenders (Chen, 1999; Chen, 2004; Gu, 2003; Gu, 2005; Ye, 1998a, 1998b; Zhang, 2004), the CET is well-conceived in the sense of being scientific, impartial, valid and reliable in terms of its design, prediction and test item analysis, and score transformation (Yang, 2003; Yang & Jin, 2001; Yang & Weir, 1998). They firmly believe that the CET, with its written part of the test, does require the use of language in real-life tasks. From their perspective, the test does not solely rely on the structural approach (or grammar-translation approach). Rather, it draws on the merits of multiple theories and principles (e.g., both the structural approach and integrative approach). They hold that linguistic knowledge is fundamental to the development of communicative competence. Without acquisition of linguistic knowledge or the “right input”, from their point of view, the enhancement of other competence or output would not be possible (Jin & Yang, 2006; Yang & Jin, 2001). Guided by the above view, they praise the CET 4 and CET 6 as meaningful indices.

On the other hand, opponents of the CET deem its constructors’ words as nothing but a “strange boast” (Donnelly & Yu, 2002). A similar criticism made by them is that the test, with its overwhelming reliance on MC and discrete-point tasks rather than performance-based items, has exerted considerably adverse influence on teaching and learning. In their view, as a test constructed on the basis of theories of structuralism, the CET has been misleading CE teaching and learning by inducing teachers’ drilling of students’ testing skills and forcing them to center their instruction on language forms (e.g., phonetics, vocabulary and grammar). Meanwhile, they blame the absence of oral aspects of English in the test. Apart from the mismatch between the CET and the CEC, some opponents (Feng, 2000; Huang & Jin, 2000; Niu, 2001; Luo & Xiao, 2002) also express their disappointment at the validity of CET, saying it is questionable. For example, based on an array of data that had been collected by them, Donnelly and Yu (2002) attack its low quality. Other scholars proceeded to question its feasibility. What deserves our attention is that such perceptions of the CET are echoed by Chinese EFL researchers and teachers at large (Cai, 2003; Donnelly & Yu, 2002; Han, 2002; Liu & Dai, 2003; Ma, 2000; Ma & Jin, 2000; Niu, 2001; Shu, 2004; Wu, 2003; Xu & Gao, 2001; Luo & Xiao, 2002; Zhang, 2003).

These two groups differ so much in their views that there seems to be little common ground for discussion. To settle the dispute, a reform seems imperative.

### 3.4 Current Reforms

To respond to the fierce criticisms leveled at the CEC as well as the CET and to raise the efficiency in ELT in China, the Chinese ME has initiated a new set of reforms in CE education. A revised curriculum, namely the College English Curriculum Requirements (CECR) was officially announced in 2004. It was developed and introduced as guidelines for CE instruction. Following the revision of the curriculum, the CET was revised from its previous focus on reading comprehension to an emphasis on assessing students' aural aspects of English, namely, listening and language use. The first revised CET based on the CECR was piloted to students in 180 universities in June, 2006 and then came into large-scale use in January, 2007. Like the past few times of CE innovation, the revision of the curriculum and the CET also led to the production of new textbooks.

Based on the documents issued by the ME and the NCETC, this series of reforms were intended for serving as an impetus for developing students' communicative competence, particularly their competence in listening and speaking. The changes were also designed with the aim of innovating the existing teacher-centered pattern of CE teaching. Specifically, teachers were expected to reduce the amount of time spent on discrete linguistic forms (the knowledge component) and increase the amount of time spent on enhancing students' communicative competence through skill-oriented activities (the use component).

Below is a detailed description of these reforms.

#### 3.4.1 *The CECR*

A look at the CECR shows that its significant difference from the preceding versions of curriculum resides in its shift from an emphasis on reading skills to one on overall comprehensive language skills. In other words, it focuses on the integration of four skills into CE teaching. Also, it differs from the old versions in that it is more competency-focused than knowledge-focused. It clearly states that the essential goal of CE instruction is to foster students' overall language proficiency, particularly their competence in listening and speaking. Furthermore, rather than providing an inventory of vocabulary and grammatical items as the old versions did, it outlines three different levels of guidelines and requirements for CE teaching and learning. They are: 1) basic requirements, 2) intermediate requirements, and 3) higher requirements. Each level of requirement consists of a set of concrete standards that

students need to meet with respect to the five skill areas (e.g., listening, speaking, reading, writing, translation). See Appendix A for details. It stipulates that the basic requirements are the minimum level that all non-English majors have to reach before graduation. One benefit of setting different levels of requirements for CE instruction is that it allows local universities and teachers to have relatively more freedom to make their own decisions as to what textbooks to use and which set of requirements to follow in light of the realities of their own schools and their students.

### 3.4.2 *Revised CET*

It must be noted that the schedule of the CET reform, far beyond my expectations, has undergone several changes. For example, originally, the news concerning the reform of the CET was officially released at the press conference held on February 25, 2005 by the Chinese ME (CET News, 2005). It was announced at the conference that the revised CET was going to be administered to students of 180 pilot schools in January, 2006.

However, on October 21, 2005 (while the current study was being carried out as scheduled), the NCETC made another announcement that the administration of the new CET was postponed till June, 2006 (see Appendix B for detailed information). With the above change, slight changes were also made in the research design of this study.

#### *Format of the new version of the CET*

Like the old versions of the CET, the revised version is mainly composed of such testing categories as listening comprehension, reading comprehension, cloze and guided writing.

Appendix C provides a sample of the revised version of the CET. Since the sample test paper covers 30 pages, it may not be practical to include it in this thesis. The complete sample test is available on the NCETC official website (CET sample, 2005).

A comparison between the format of the 1999-version of the CET and that of the current revised version shows that the two tests do not differ too much in terms of the testing methods adopted. As can be seen in Table 4, both versions of the CET comprise questions in selective response format such as MC questions (which require students to select an answer or answers from options) and constructive response format (which require students to produce answers in their own words). It can be

noted that some constructive forms of testing tasks adopted in the current version, such as dictations, spot dictation and compound dictation (CD), cloze or error correction, short-answer questions (S-A Q), and translation (T), were included in the old version of the test. In spite of the similar testing tasks utilized in the two versions, however, they differ considerably in the weighting of these test items, the amount of test items adopted, and the degree of difficulty embodied in the test tasks. For example, the proportion of listening is increased from the original 20% to 35%, and at the same time the proportion of reading comprehension is decreased from the original 40% to 35%, in which the newly emerged item, fast reading, takes up 15%. Table 4 outlines the format and structures of the two versions of the CET. Another significant change is that the section of Vocabulary and Structure (15%) has been removed from the revised CET.

Table 4 *Format and Structures of the CET*

Format of the Current Version					Format of the 1999 Version				
Components	Items		Task Type	Ratio	Components	Items		Task Type	
Part I Listening Tasks	Conversations	8 short	MC	15%	Part I Listening Comprehension	10 Short conversations		MC	10%
	3 passages	2 long	MC	20%		Passages	Comprehension	MC	10%
		Comprehension Dictation	CD				Dictation	CD	
Part II Reading Tasks	Reading in Depth	Comprehension	MC	35%	Part II Reading Comprehension	Reading Comprehension (4 Passages)		MC	40 %
	2300-word passages; 1500 word passage	Vocabulary	WC						
	Fast reading 1200 words/15M	Skimming and Scanning	7 T/F, 3 S/WC						
Part III Integrative Tasks	Cloze		MC	10%	Part III Vocabulary and Structure			MC	15%
	0r Error Correction or S-A Q			5%					
Part IV Writing Tasks	T (Ch-E)			15%	Part IV Comprehensive	Cloze, T (E-Ch) or S-A Q, one of the above each time			10%
	120 Words/30M					Part V Writing			
Scores	290 – 710 (No passing line is set.)				0 – 100 (A passing line is set.)				
	CD: Compound Dictation				M: Minute				
	MC: Multiple Choice				S: Short-answer				
	S-A-Q: Short-Answer Question				T (Ch-E): Translation from Chinese into English				
	T (E-Ch): Translation from English into Chinese				S-A Q: Short-answer Question				
	T/F: True and False Question				WC: Word Completion				

We should be mindful at this point that whether the new changes introduced in the curriculum and the CET can help to engineer instructional innovation in CE teaching

and learning is still unknown. Considering the fact that negative perceptions of the CET are so prevalent in the Chinese context, empirical evidence should be gathered to settle the dispute.

### 3.4.3 *Textbooks*

To correspond to the reforms in the CE curriculum and the CET, new textbooks were developed for CE course. The three widely used ones are: “New Era Interactive English” (published by Tsinghua University Press), “New Perspective English Learning System” (published by Shanghai Foreign Language Education Press), and “New Horizon College English” (published by Foreign Language Teaching and Research Press). These textbooks were recommended by the ME. It was said that the source materials in the course books were taken from original English publications. They were claimed to have been designed to reflect and reinforce the CECR in terms of teaching and learning objectives, focuses and approaches. These course books differ from those of the past in that more cultural -related themes were incorporated in the content of the materials.

## 3.5 Cultural Roots

Through a review of the pertinent literature, I find that although there have been big strides in Chinese ELT since the 1980's, there exist many context-bound problems in EFL classrooms. One complex problem is that although CLT, task-based teaching and autonomous learning are very much in vogue in the ESL context, they are not well-adopted in Chinese EFL classrooms, including CE classrooms. As critics often pointed out, CE instruction has been entrenched in traditional, teacher-centered practices (Qian, 2003; Shu, 2004). It is charged with being lecturer-oriented, teacher-centered and test-oriented on the one hand, and attacked for being time-consuming and low-efficiency on the other (Dai, 2001; Han, 2002; Jing, 1999; Liu & Dai, 2003; Yu, 1999). Although Chinese EFL teaching and learning methods cannot simply be dismissed as ‘primitive’, ‘old-fashioned’, or ‘misguided’, there is abundant evidence showing that English instruction in this context as a whole has been far from successful (Burnaby & Sun 1989; Gatbonton & Gu 1994; Leng, 1997; Liu 1995; Tarnopolsky et al., 2001). This outcome, however, seems to be inseparable from the Chinese educational system, the traditional Chinese culture and deeply-rooted beliefs and values. Thus, next I will briefly discuss the Chinese

educational and historical setting where the above-mentioned problems were created and sustained.

### 3.6 Chinese Educational Context

#### 3.6.1 *Exam-driven*

The Chinese educational system is characterized as being examination-driven (Cheng, 1997; Shu, 2004; Ting, 1987). One typical example is that students have to sit for numerous examinations as soon as they start schooling. Under this system, exams are of exaggerated importance. At various levels of education, be they secondary or tertiary, it is a common practice that teachers teach to the test (Campbell & Zhao 1993; Liu 1988; Liu 1995). Not only are most courses tailored toward exams, but the teachers' and students' attention is also correspondingly directed at skills which will be tested in the exams. Furthermore, test scores are viewed both as a marker of students' academic success, and as the premise to their future career.

The review of the literature reflected that the problem of over-dependence on exams stems from history. Traditionally, examinations have always played an important social and educational role in China. Intended efforts to use examinations to establish and control education programs were made as early as 201 BC in China (Cheng, 1997). The first written public examinations were introduced over 2000 years ago to select the most able citizens for positions in the civil service. This examination system lasted for at least one thousand and three hundred years (Ma, 1991). While the examination system has helped strengthen the rule of the feudal emperors, it has hindered the development of education in that it often makes "passing the examination" the only goal for schools and students. It is widely held that the same type of hindrance is still in existence. Hence, a historical perspective is valuable for understanding the test impact.

#### 3.6.2 *Centralized*

It is also well-known that Chinese education is highly centralized. While teachers are given the freedom to design their own lesson plans, to a large extent the choice of textbooks and content of teaching are determined by authorities or school leaders. Crook (1998) criticizes such a top-down decision making practice (1997). From his perspective, in centralized, "authoritative" educational systems, the central aim of education is to guarantee know-what rather than know-how. One direct consequence



is that teachers are regarded as the givers of knowledge.

### 3.6.3 *Knowledge-focused*

Another typical feature of the Chinese educational system is that it is knowledge focused. Under such a knowledge-focused system, the belief is widely held that if a student wants to learn a subject well, he or she has to accumulate or acquire the basic knowledge or theories of the subject. Guided by this belief, both teachers and parents tend to stress the importance of what they perceive as ‘knowledge’ or ‘foundation’ of a subject. Within such a social setting, learning is generally viewed as a process of accumulating knowledge, analyzing and memorizing ideas in books rather than a process of acquiring practical skills. A teacher is regarded as the provider of knowledge, whereas a student is viewed as a recipient of knowledge.

Meanwhile, Chinese students' language learning styles, as described by Chen, (1999) and Harvey (1985), are characterized by two main features: meticulous attention to linguistic and stylistic details rather than communicative skills, and over-reliance on the use of translation. According to Ting (1987) and Liu (1988), the above commonly-held belief and knowledge-based practice may trace their roots to Confucianism and traditional Chinese philosophies of learning and teaching. In addition, they may also result from the influence of the Soviet education that has focused on academic study of grammar, literature, and in-depth knowledge of literary texts (Ting, 1987; Liu, 1988), because for a certain period of time in the 1950's, Russian was the only foreign language taught in Chinese schools. Shu (2004) attributes the reason for teachers' persistence in adhering to the traditional style of teaching to the low participation in research by Chinese EFL teachers. For her, it is the lack of uptake by Chinese EFL teachers of the ELT theoretical principles that resulted in low level of teaching in CE. She bases her contention on a survey involving several hundred EFL teachers from around 300 universities.

In summary, the Chinese educational context (which involves not only the classroom and school, but also the educational system, the historical background and social environment) is typically exam-oriented, centralized and knowledge focused. It appears to be this combined influence of the context that hinders CE teaching and learning. As for how these social and contextual factors combine with personal factors to hamper the current instructional and testing innovation, further investigation needs

to be done to explore the issue. The next chapter discusses the procedures for carrying out this investigation.

Chapter 1 has provided a description of the context of the study – the research problems that generated this research study, the background of the study, purpose and objectives of the study and research questions as well as the significance of, and rationale for, the study.

Chapter 2 has reviewed the pertinent research studies by taking an in-depth look at the theoretical and methodological advances pertaining to washback research and then summarizing a general conceptual framework for insights to illuminate the common themes and patterns that will emerge from this study.

This chapter has depicted the context of Chinese CE teaching and testing. Initially, it gave a brief description of the growth of the CE course, the CE curriculum, the CET as well as their relationships and earlier reforms carried out in them over the past couple of years. Then it turned to reveal the thorny issue(s) confronted by CE professionals by giving a brief account of two fundamentally different convictions of the CET and the CEC. Following that, it provided a basic introduction about the revised CE curriculum and the revised CET. Finally, it presented some major lingering problems and examined the overall local educational, sociocultural, and historical context in China where the problems are nurtured.

Chapter 4 will present the methodological procedures of the study such as: sampling procedures, participants, instruments, the pilot study as well as procedures of data collection and data analysis.

## **Chapter 4: Methodological Procedures of the Study**

### **4.1 Introduction**

This chapter starts by describing the purpose and the research questions of the study. Then, it turns to the rationale for the methodology that was applied. After that, it describes the methodology for data collection, the research design adopted, the instruments used, the participating teachers and research sites involved, and how they are selected. Finally, the data collection procedures and the process of data analysis are explained.

### **4.2 Purpose of the study**

My study examines the role that the ‘teacher factor’ plays in generating washback. As stated in Chapter 1, the purpose of this study is threefold: (1) to investigate the mainstream attitudes Chinese EFL teachers maintain toward the CET and its impact; (2) to find out what aspects pertinent to the ‘teacher factor’ (e.g., teacher beliefs, teacher knowledge, teacher proficiency, etc.) present the major barrier to the implementation of educational change so as to conclude whether tests constitute a constraint on ELT methodology innovation in China; and (3) to examine how the major components of the ‘teacher factor’ involved in the current study (e.g., teachers’ beliefs of the CET and its impact, teachers’ beliefs of teaching and learning, teachers’ knowledge base, etc.) are interrelated.

By focusing on the ‘teacher factor’, my intent is to arrive at an accurate description and explanation of Chinese EFL teachers’ beliefs, interpretations and practices related to tests and test impact, and to achieve a better understanding of the key factors underlying the alleged ‘test impact’ as well as the reason behind Chinese EFL teachers’ practice.

### **4.3 Research Questions**

The global research question guiding the study is: What role does the ‘teacher factor’ play in washback in the Chinese university context? There are two secondary questions.

1. To what extent and in what form does washback exist in China in terms of its effect on teacher beliefs (e.g., beliefs about the CET and its impact, beliefs about teaching and learning), and classroom behaviors (e.g., content and particularly teaching methodology)?

2. How is the ‘teacher factor’ manifested in such a washback effect? What aspects of the ‘teacher factor’ (e.g., beliefs, knowledge, past experiences) contribute to the way that teachers interpret and react to washback?

### *Hypotheses of the CET Impact*

Provided a relationship exists between the test and what the teachers practice, 1) there will be an increase in the amount of time spent on activities such as listening and a decrease in the amount of time spent on reading; 2) there will be activities aimed at enhancing students’ skills in fast reading (or skimming and scanning); 3) last but not the least, the amount of time devoted to language forms will be reduced.

The above hypothesis is presented visually in Figure 4. The blue dotted arrow indicates the hypothesized relationship between the CET and its washback effect on teacher classroom practice. The dark blue arrows show an increase will occur, while the light grey arrows indicate a decrease will take place.

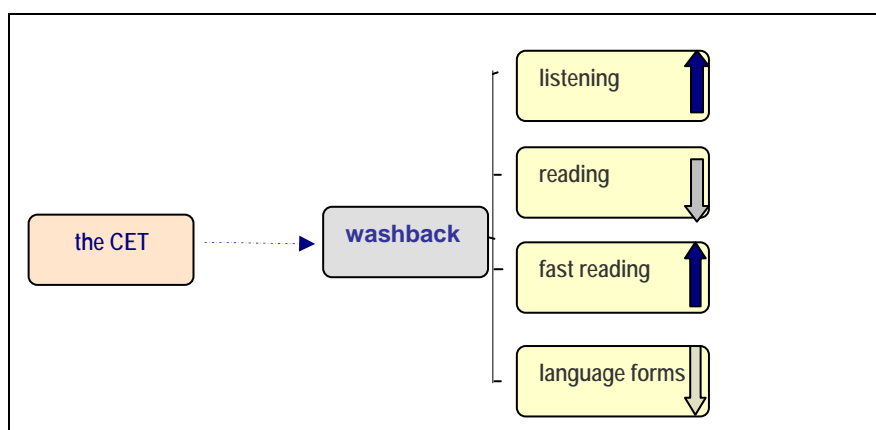


Figure 4. Hypothesized washback of the CET on teacher classroom practice.

### 4.4 Rationale for and Overview of the Methodology

Drawing on the above review of the literature in Chapter 2 and on the basis of a prior pilot study, I made the decision to adopt a MM approach as an investigative technique (see below). The choice of a MM strategy was made due to the following considerations.

The first consideration relates to the complexity and context-sensitivity of the research problem under study. By context sensitivity, as interpreted by Patten (2002), a phenomenon should be understood in all its complexity and within a particular situation and environment. As indicated above, my research objective is to examine

Chinese EFL teachers' perceptions and thinking processes which underlie their classroom actions as well as address how personal factors, institutional context and societal pressures combine to affect teachers' perceptions and behaviours throughout the process of implementing instructional change. Nevertheless, findings from washback research demonstrate that the study of washback exhibits considerable complexity (see Section 2.2.3). Moreover, this complexity is intensified in the Chinese context where the educational system and social environment is typically exam-oriented, centralized and knowledge focused (see Section 3.6). Thus, there has emerged a need to take into account not only the pedagogical but also the social and personal complexities influencing their perceptions and practices. According to Greene (2007), MMR, with its emphasis on holistic, richly detailed descriptions and analyses of teaching behaviors and the multilevel contexts in which those behaviors are nurtured, is best suited for capturing the complexity of the social phenomenon being studied (Greene, 2007). Meanwhile, as noted by Turner (2006, 2007), the MMR has the potential to "help respond to certain types of questions, especially those having to do with classroom contexts" (2009, p.108). In this regard, this approach seems to be best suited for my research purpose.

The second consideration has to do with the emphasis that MMR places on tailoring methods to research questions. As was put by Johnson and Onwuegbuzie (2004), research approaches should be mixed in ways that offer the best opportunities for answering important research questions. Based on their explanation, MMR does not dictate the choice of data collection methods. Rather it allows the procedures for conducting research to be dictated by the research question and the context of the study. One of the salient strengths of the MM approach lies is that it allows researchers to mix aspects of the qualitative and quantitative paradigms at all or many methodological steps in the design (Creswell, 1994; Creswell & Plano Clark, 2007). In view of this strength, a MM approach was deemed an appropriate avenue for addressing my research question.

A third consideration is based on a belief shared by Creswell (2009), Creswell and Plano Clark (2007), Greene (2007), Johnson and Onwuegbuzie (2004), and Teddlie and Tashakkori (2009) that MMR produces better outcomes than mono-method research.

According to these researchers, MMR has the potential to reduce some of the problems associated with single methods. From their perspective, by utilizing

quantitative and qualitative techniques within the same framework, MMR can incorporate the strengths of both methodologies. In light of the above perspective, in order to examine the Chinese context and understand the phenomenon of washback (e.g., how the network of teachers' BKEs relates to their practices and perceptions of tests), it is necessary to draw upon both types of data (QUAL and QUAN).

The fourth consideration concerns some of the distinctive characteristics of MMR. As described by Greene et al. (1989) and Johnson and Onwuegbuzie (2004), the value of MMR lies in that it possesses a number of strengths which help increase the quality of final results and provide a more comprehensive understanding of analyzed phenomena. Among the characteristics that are outlined by Greene (2007) and Greene et al. (1989) are (a) triangulation (i.e., seeking to ascertain how different methods check, validate or corroborate one another.); (b) complementarity (i.e., seeking elaboration, enhancement, and clarification of the results from one method with results from the other method); (c) initiation (i.e., discovering contradictions that lead to a re-framing of the research question); (d) development (i.e., using the findings from one method to help inform the other method); and (e) expansion (i.e., seeking to extend the breadth and range of research inquiry by using different methods for different inquiry components). Greene (2007) contends that these features would help increase the interpretability, meaningfulness, and validity of constructs and inquiry result. Considering the purposes of these features and my study, it appears they would help ensure the validity of my study, as well.

With respect to the choice of research methods, an essential first step to be taken involves an examination of all relevant and available documents related to the CET, the CE curriculum, and its teaching and learning. This step is crucial, for only when I achieve a good comprehension of the objectives and content of the test, will I be in a better position to explore teachers' accounts of and reactions to the test.

Three complementary methods (i.e., interviews, observations and questionnaires) were employed in my research design. As my research objective is to examine Chinese EFL teachers' perceptions of teaching, learning and testing and how they are shown to operate in the course of teachers' instructional practices in the classroom, semi-structured interviews and classroom observation are assumed to be best suited for this research purpose. The assumption is that observation is useful for obtaining descriptions of behavior and events, whereas interviewing is mainly useful for eliciting the perspectives of teachers (Maxwell, 1996). Meanwhile, a teacher

questionnaire will be employed to further explore teachers' various perspectives and how they are interrelated with other sets of data. Based on Maxwell (1992), although the interviews can give the researcher the required scope and depth, they cannot ensure the representativeness of the data. In addition, questionnaires also seem adequate to quantify data and to provide descriptions and comparisons of teacher beliefs (Barcelos, 2003). Thus, to facilitate the generalizability of insights derived from qualitative data, to expand the generalizability of the findings (Reichardt & Cook, 1979) and to better describe teacher beliefs by quantifying them, the data will be supplemented with the questionnaire.

Worthy of note is that when devising my study, considerations of validity (the relevance of the data) and reliability (the consistency of data analysis) also permeate every level of the research from how questions are asked and how data are gathered and analyzed; to how and to whom research results are reported. Drawing on the review of the research methodologies employed in washback research, both method triangulation and data triangulation were carried out. Method triangulation is achieved by combining aspects of quantitative and qualitative methods in the stages of data collection and data analysis. It is believed that the combination of these research methodologies would allow me to examine the perceptions and behaviors of Chinese EFL teachers from many different angles. Data triangulation is achieved by having different sets of data cross-checked.

In addition, other standards such as "persistent observation", "thick description of the content", and "explicit emphasis on research question(s)" have also been taken into account in my study. In Chapter 3, a thick description has already been provided on the context of the study and in Chapter 5, a comprehensive account of the results will be offered.

Overall, by positioning myself within the MM design, I may be able to identify specific patterns of teacher beliefs and behavior, and describe them in relevant descriptive terms, and above all, I may be able to place them in some relations to the wider social context. It shows that a match exists between what I need and what the MM approach can offer.

#### 4.5 Methodology

In Chapter 2, an introduction was provided on the theoretical underpinnings of the MM design as well as the rationale for utilizing the MM approach in washback

research. Owing to the focus of the study, the decision was made to conduct this study utilizing such an approach. Specifically, I combined aspects of quantitative and qualitative methods in the stages of data collection and data analysis. Qualitative data collection mainly involves case studies conducted through in-depth interviews and classroom observations, while quantitative data collection consists of the completion of a questionnaire.

This approach was chosen based on three aspects of the study: the type of problem being investigated, the goal of the study, and the nature of the data. The purpose of adopting this approach is to devise a solid research design that may maximize the possibility of addressing my research questions thoroughly.

Furthermore, I employed an emergent research design (Patten, 2002; Maykut & Morehouse, 1994). What this means is that research is an ongoing process. Its design evolves over time. I assumed that building flexibility into the design was crucial. As the research evolved, it was likely that I would gain different perspectives on my research topic or find more substantial ways of addressing my research questions. Thus, it was necessary to maintain some flexibility in the implementation of the research plan.

#### 4.5.1 *Research Design*

To implement the research, a MM research design was adopted. Specifically, it is a sequential exploratory triangulation design (Creswell & Plano-Clark, 2007). As both QUAN data and QUAL data were collected sequentially at different phases (periods of time), a visual diagram helps to better illustrate the process and progression of research. Figure 5 depicts such a process. What I need to mention is that here only the general procedures of data collection and analyses are presented. Detailed procedures are presented in other tables below.

As shown in the diagram, the process of research is categorized into four stages. At each stage, either QUAN or QUAL data or both forms of data (QUAN + QUAL) were collected and analyzed. With respect to the overall process of data collection, the stages proceed in a specific sequence over time (as the long light blue arrow indicates). Each stage of data collection leads to (as depicted by small light

grey arrows) the subsequent stages of research (both data collection and analysis). However, within each stage of data collection, not only were the data collected concurrently, but the data analysis is also concurrent with data collection (as the



vertical green arrows manifest).

What is worthy of note is that with the adoption of an emergent research design, the data sets were examined, contrasted, analyzed and triangulated across different stages of the research process and finally the different sets of the data were triangulated and merged into one overall interpretation (as depicted by the long black arrow as well as the circulating short dark grey arrows).

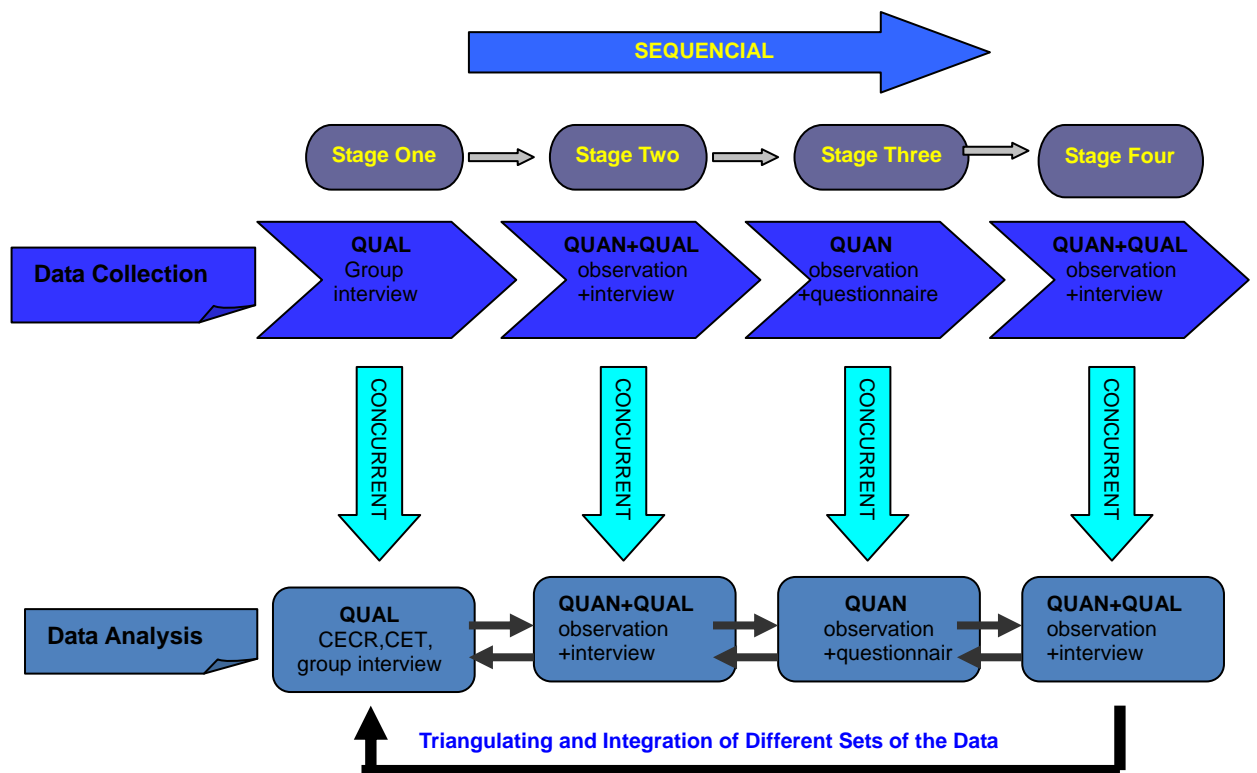


Figure 5. Visual diagram of the research design of the study.

#### *4.5.2 Participants and Research Sites*

##### *4.5.2.1 How I Selected My Participants*

The population of this study was composed of Chinese tertiary-level EFL teachers of non-English programs. All of the participating teachers were non-native speakers of English. They were recruited both for the purpose of the case study and the survey. Initially, a segment of this population, 30 teachers from case-study settings, participated in a qualitative group interview. Then six of them were selected for the case study which consisted of focused observations and in-depth interviews. After that, participants from the nationwide context were surveyed.

When selecting the participants, Patton's (1990, 2002) "maximum variation sampling" technique was followed. A purposive sample was used. Based on Patton (1990) and Maxwell (1996), purposive sampling involves selecting particular settings and individuals/groups deliberately not only based on our research questions, but also based on information available about these individuals/groups. The participating teachers were chosen on the basis of their potential for yielding data which could reveal teachers' perceptions in general. My experience as a Chinese EFL teacher gave me an insider's knowledge of the context of the research sites. This knowledge guided me in my initial sample selection.

In order to get a representative sample, the participants were chosen according to three parameters: (a) the type of students they teach and (b) the type of schools they work in (c) the geographical location of the universities. For the first parameter, only teachers of non-English majors were included in the sampling. With respect to the second parameter, I ensured that teachers selected represented schools of different levels (key school, average school, etc.). As for the third parameter, I ensured that participating teachers came from schools of various geographical locations: large/small cities; rural/urban, north/south. Since the data of this study were collected both for the purpose of the case study and the survey, there were different considerations while selecting the two groups of participants. For the former purpose, the first and the second parameters were given consideration, whereas all parameters were taken into account with regard to the latter purpose. Meanwhile, I ensured maximum variation in participants' age, gender, teaching experience, degree of education. In addition, my selection of participants was also largely based on practical considerations and teachers' willingness and interest to discuss the class and reflect on their own process of teaching.

#### *4.5.2.2 Participants of the Pilot Survey*

The pilot sample was drawn from teachers of a university in one of the largest cities in China. The sample consisted of 15 EFL teachers.

#### *4.5.2.3 Questionnaire Participants*

The questionnaire participants consisted of 195 teachers from a number of different Chinese university settings. They were recruited from two sources: teachers who participated in an EFL conference and teachers from seven different schools located in one of the largest cities in China. These teachers were believed to represent the largest population of tertiary-level English teachers in China. The survey was administered between May and June in 2006. It was mass administered both at the conference site and during regularly scheduled faculty meetings at the research sites.

Two hundred and fifty-five copies were distributed and 195 copies were returned, with a return rate of 76.5 %.

#### *4.5.2.4 Interview and Observation Participants*

The participants for the interviews and observations consisted of 30 EFL teachers working in three different universities in one of the largest cities in China. At first, all these teachers took part in a group interview. Three groups of teachers, with ten participants in each group were interviewed respectively. Then six of them were involved in the case study.

#### *4.5.2.5 How I Selected My Research Sites for the Case-study*

The research sites for this study are three tertiary-level institutions in a large city in Northern China. These sites were chosen chiefly due to practical considerations. Owing to my personal as well as professional connections with the chosen universities, I had obtained permission from these schools to administer the questionnaire, and conduct interviews and observations there. Apart from practical reasons, the research sites were also chosen on the basis of their potential for maximizing variation in teacher and student differences in terms of their academic performances and yielding data which could reveal a wide range of teachers' behaviors and perspectives. One consideration was that different types of schools (pilot/non-pilot; key/average) were included in the sample. The assumption was that these schools might exhibit differences in how they react to the CET reform, since the students from non-pilot

schools would take the revised CET about six months later than pilot schools.

#### *4.5.3 Instruments*

As indicated above, the literature relevant to washback research has informed my decision on what types of instruments to be chosen for this research.

The data for this study were collected through a pilot study, a questionnaire, interviews with teachers, observations of classes, and analyses of documents. Two different types of interviews (e.g., a group interview and in-depth individual interviews) were undertaken at different stages of the study. All the interviews were semi-structured, conducted in a systematic and consistent order which nevertheless allowed me sufficient freedom to probe far beyond the answers to the prepared questions (Bogdan & Biklen, 1998).

These instruments will now be described in detail.

##### *4.5.3.1 The Review of Documents*

In this study, I conducted an intensive review and analysis of the documents pertaining to the CECR, the CET, and the textbooks used at the research sites. The CET-related documents and reports issued by the NCETC are taken as official sources reflecting the test constructors' intentions. One purpose of the review and analysis was to find out what the revised CET sets out to measure (e.g., linguistic knowledge or language use) and whether or not the CET represents the curriculum. Another purpose was to identify the characteristics of the CET, for they would serve as the basis for a comparison to what is happening in the classroom and would help determine whether the observed classroom phenomenon is closely test-related (e.g., whether they are similar or there are gaps between the two).

##### *4.5.3.2 Group Interview*

In terms of the group interview, three groups of teachers were interviewed respectively in late October, 2005. Each interview lasted about one hour and a half. The interview data totaled around 5 hours.

The purpose of the group interview was to get a general understanding of a range of perspectives on my research topic (e.g., teachers' perceptions of the revised CET and its impact, their views of language teaching and learning, and information about what they consider to be effective ways of teaching, etc.). It was expected that this group

interview would throw light on the possible reasons why teachers claim the CET to be a constraint on their instruction. It was also expected that the insights gained from this format of interview would help me determine the precise focus of the research – what to observe, whom to interview next, and what to pursue in the research to follow.

A teacher interview protocol was developed to provide the direction for the interview. Six questions were prepared, some of which were generated based on the feedback from the pilot questionnaire. A copy of the interview protocol is included in Appendix D.

All the interview sessions were conducted in Chinese and audiorecorded with the interviewees' permission. The interviews were then transcribed. The transcripts were analyzed for patterns and commonalities.

#### 4.5.3.3 Classroom Observation

Classroom observations aimed at examining the specific activities teachers were engaged in before and after the participants read the sample revised CET. The main purpose of the observation was to find out whether the test can foster an impact on CE classroom teaching and learning. Meanwhile, it was hoped that conducting classroom observations might help determine whether teachers' accounts of their beliefs, their understanding of ELT methodologies as well as their attitudes toward washback conform to their classroom behaviors. Three rounds of observations (R1, R2, and R3) were carried out during the six-month period (between November, 2005 and May, 2006) at the three case-study settings (S1, S2, S3). In each round of observation, I observed six different classes for four consecutive hours each. A total of 70 hours of observation were conducted (R1–24 lessons; R2–24 lessons; R3–22 lessons).

Table 5 gives a description of the number of hours of observation conducted at the three research sites during the three rounds of classroom observation.

Table 5 *Number of Hours of Observation*

<b>Number of Hours of Observation</b>	<b>R1</b>	<b>R2</b>	<b>R3</b>	<b>Total</b>
<b>S1</b>	8	8	8	24
<b>S2</b>	8	8	8	24
<b>S3</b>	8	8	6	22
<b>Time of Observation</b>	Nov.28–Dec27, 2005	Mar.3–Mar.24, 2006	May 9–May 22, 2006	

It should be noted that in R3, one of the teachers at S3 had a class schedule change

and thus his lessons were only observed once during R3.

### *Observation Schedule*

An observation schedule was designed. It was modeled on the COLT Scheme (Spada & Frohlich, 1995) and Watanabe (1996b). The schedule (see Appendix E) was also designed based on the analysis of the data derived from the pilot questionnaires and group interviews. The hypotheses generated from the literature review also formed the basis of the observations. The instrument was designed to record the following aspects of information.

1) Classroom Organization Patterns: percentage of class time spent on student-centered activities (e.g., pair-work, group work, individual work, role-play); percentage of class time spent on teacher-centered activities (e.g., teacher lecturing to the whole class without interactions with students– teacher presentations, explanations of sentences, reading aloud, translations, etc.).

The purpose of exploring classroom organization patterns in teachers' instructional process is to find out who is holding the floor in the classroom.

2) Focus of Instruction: frequency of explaining language points with a focus on language forms (e.g., explanation of sentence structures, rote practice and mechanical grammar exercises; explanation of vocabulary in a decontextualized manner); frequency of involving students in meaning-based activities (e.g., discussion, role-play, comprehension exercises at the discourse-level, etc.).

This was designed to evaluate whether the lessons delivered by the participants are form-focused or meaning-focused, and to what extent teachers' instruction was communicatively-oriented.

3) Relevance to the Test: percentage of class time spent on aural/oral aspects of English (e.g., listening practice, oral practice at the discourse level encouraged by the CECR) as well as on fast reading practice (effected by the CET); frequency of giving information or advice about the CET (old/new) or test-taking strategies.

This section was devised to discern whether and to what extent teachers' instruction was related to the CET.

4) Medium of Instruction: English/Chinese/half English/half Chinese

This was designed to find out about the language used by teachers in their instruction.

5) Teaching Materials: textbooks, test-related materials (e.g., the old CET papers or

simulated test papers), audio or audio-visual materials, or other supplementary teaching materials.

By examining the materials chosen by teachers, I aimed to find out about the content of teaching.

In addition to the above activities and events listed on the observation schedule, other visible classroom events were recorded on a note-taking sheet (i.e., field notes). They were used for comparison with the characteristics of the CET to determine whether the observed classroom phenomenon was related to the test.

All the observed lessons were audio-recorded. The observation instrument includes an MP3 (for audio recording), an observation schedule, note-taking sheets, pencils and a watch. During each observation, the observation schedule was filled in. Detailed procedures for classroom observations will be discussed below.

#### 4.5.3.4 *In-depth Interview*

Two rounds of in-depth interviews were conducted during the six-month period (between November 2005 and May 2006). This was an interview on a one-to-one basis. They were a supplementary instrument used in the research. Each case-study participant was interviewed twice: first on the completion of the first round of observation; second on the completion of the third round of observation. Altogether 12 on-site post-observation interviews were conducted, with 6 interviews in each round. The time for the interviews ranged from 40 minutes to one hour. The questions covered in the interview protocol (see Appendix D) aimed at probing into the meanings the teachers attached to their classroom behaviors and they also focused on learning about teachers' thoughts, plans and decisions that are related to the observed activities and events observed. In contrast to the group interviews, in-depth interviews in this study aimed at understanding the participants' perceptions at a deeper level.

All the interviews were conducted in Chinese and audiorecorded with the interviewees' permission. All the recordings were, then, summarized in English by the researcher.

#### 4.5.3.5 *Questionnaire*

Another instrument was a 60-item questionnaire for the participating teachers. As discussed in Chapter 2, the use of questionnaires as a data collection tool appears to glean useful evidence in washback contexts. In the current study, it was designed to

further examine the way teachers interpret language teaching and learning, testing, and test impact, and react to the CET reform. Also, it was designed to find out whether the patterns and themes that had emerged from previous stages (e.g., interviews and observations) could be confirmed and applied to a larger group of teachers. Some questions in the questionnaire were adapted from Turner (2005). Other questions pertaining to the relationship between the 'teacher factor' and washback were drawn from the findings of other washback research (Cheng, 1997, 1998, 1999; Watanabe, 1996b; Wall and Alderson, 1993; Shohamy, 1993; Shohamy et al., 1996; Alderson and Hamp-Lyons, 1996; Andrews et al., 2002) and fine-tuned based on the researcher's knowledge of the Chinese EFL teachers' beliefs and experiences. Sources of the questionnaire also include: the CECR, the CET papers as well as articles about the CET washback (Yang, 1999; Jing, 1999; Niu, 2001; Liu, 2002).

The questionnaire was structured in 4 parts. Part One (from Q1 to Q8) aimed to solicit questions about the participants' background information: such as age, gender, previous language-teaching/learning experiences, etc. Part Two (from Q9 to Q15) consisting of a set of questions that are rated on a six point scale ranging from 1 (0-10) to 6 (51% or more), were related to the participants' instructional behaviors in the classroom. The last part included questions or statements which sought to find out about teachers' various beliefs and classroom behaviors. The items covered in this part were classified into five different themes: 1) teachers' beliefs about the CET and its impact on their instructional practices, 2) their beliefs of the washback effects of the revised CET on their teaching and learning, 3) their beliefs about teaching, learning, how to teach and the ways that they teach, 4) their knowledge base – curriculum knowledge (e.g., knowledge of the CECR, the textbook(s), the CET) and pedagogical knowledge (e.g., knowledge of the features of CLT, etc.), and 5) their past experiences.

The questionnaire not only combined closed- and open-ended items, but also involved questions or statements in different formats. It mainly included Likert-type questions, MC items and yes/no questions. The scale used in the Likert-type questions ranges from 1, strongly disagree, to 4, strongly agree. The MC items and yes/no questions were employed as cross-referencing questions to ensure an accurate and thorough assessment of the teachers' beliefs, knowledge and practices. In addition to the given questions, the participants were encouraged to air their personal views in the questionnaire.



The questionnaire was administered in English on two separate occasions: one at a multiday ELT conference site, and the other during regularly scheduled faculty meetings of seven universities in northern China. The survey was conducted from May to June 2006. The completion of the questionnaire took approximately 20-30 minutes.

Refer to Appendix F for the survey instrument.

#### 4.5.3.5.1 *Constructs of the ‘teacher factor’ and ‘teacher practice’.*

When the questionnaire was being developed, as demonstrated by Figure 6 and Figure 7, two constructs (the ‘teacher factor’ construct and ‘teacher practice’ construct) were hypothesized. They were established based on the insights gained from the literature and the information gathered through pilot study process. In the ‘teacher factor’ construct, there are 15 items related to the ‘teacher factor’. The clusters of the ‘teacher factor’ items were hypothesized to measure six latent factors (containing six sets of variables): ‘beliefs about teaching and learning’ variable (BTL) (measured by 3 questionnaire items), ‘beliefs of impact of the CET on teaching’ variable (BI) (measured by 3 questionnaire items), ‘beliefs of washback’ variable (BW) (measured by 3 questionnaire items), ‘pedagogical knowledge’ variable (PK) (measured by 2 questionnaire items), ‘curriculum knowledge’ variable (CK) (measured by 2 questionnaire items) and ‘teacher experience’ variable (TE) (measured by 2 questionnaire items).

Table 6 presents a taxonomy of the ‘teacher factor’ variables.

Table 6 *Taxonomy of the ‘Teacher Factor’ Questionnaire*

Components of the ‘Teacher Factor’	No. of Items Used	Question No.
<b>1. Various Beliefs</b>		
Teacher Beliefs about Teaching and Learning (BTL)	3 items	24, 26, 27
Teacher Beliefs about the Impact of the CET on Teaching (BI)	3 items	21, 23, 49
Teacher Beliefs about Washback (BW)	3 items	50, 52, 55
<b>2. Teacher Knowledge Base</b>		
Teacher Curriculum Knowledge (CK)	2 items	30, 31
Teacher Pedagogical Knowledge (PK)	2 items	40, 43
<b>3. Experience</b>		
Teacher Training (TE)	2 Items	8, 57.1
<b>Total</b>	15 items	

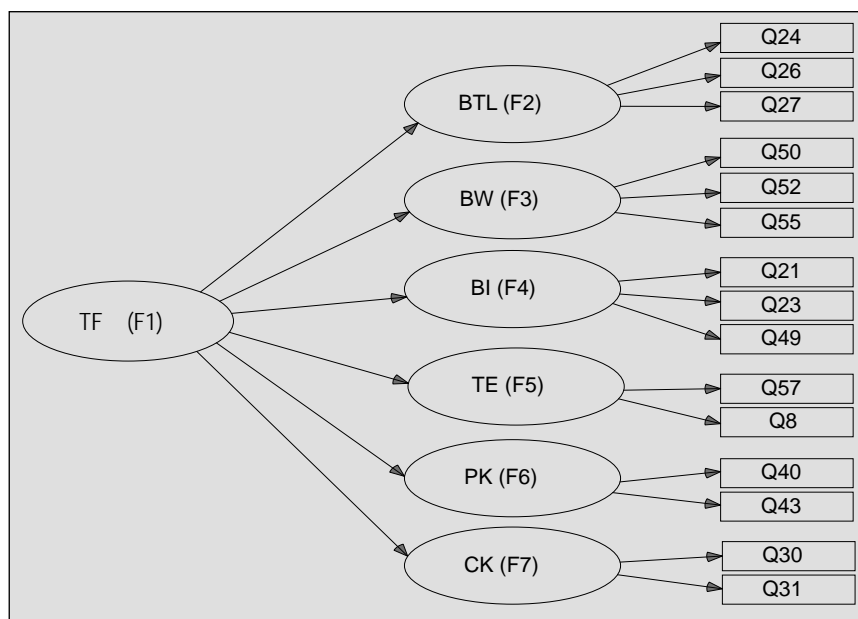


Figure 6. Hypothesized construct of the ‘teacher factor’.

The ‘teacher practice’ construct consisted of four items related to ‘teacher practice’.

To be specific, it covered the CET-related items – the percentage of time respectively spent on listening/reading/vocabulary, and the percentage of teacher-talk time in the classroom. These items were assumed to reflect, to a certain extent, the impact of the revised CET on teaching, evidence of teacher behavioral change. Table 7 presents a taxonomy of the ‘teacher practice’ items.

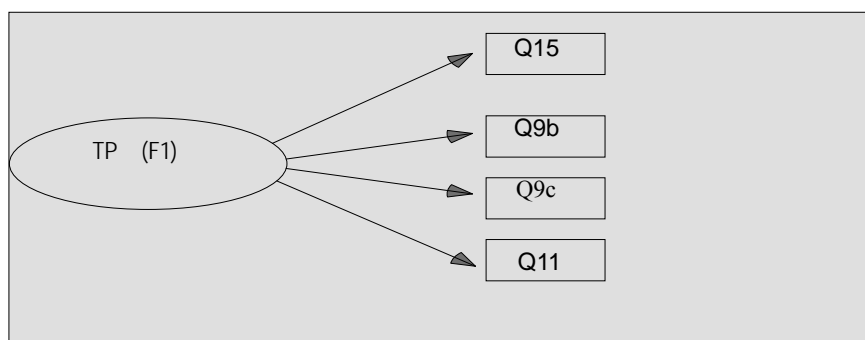


Figure 7. Hypothesized construct of ‘teacher practice’.

Table 7 Taxonomy of the ‘Teacher Practice’ Questionnaire

1. Teacher Practice	No. of Items used	Question
– CET-related Activities in the classroom	4 items	9b, 9c, 11, 15

#### 4.5.3.5.2 *Piloting.*

The initial version of the questionnaire was piloted in August, 2004 on 15 Chinese tertiary-level EFL teachers to check appropriateness of the questions. The results indicate that it was too long and time-consuming. Therefore, based on the information gained from the pilot study, it was refined, revised and shortened. Some ambiguous questions were either reframed or excluded from the final adopted version.

#### 4.5.4 *Data Collection Procedure*

The data collection involved four stages of work.

**Stage 1 :** The first stage involved the review of documents and administration of a group interviews. During this stage, I conducted a broad spectrum of observations, with mapping of the site and choosing a true representative sample. The washback effect of the CET at the macro level (e.g., current social and educational context related to washback in Chinese tertiary schools) was examined. The goal of this stage of data collection was to get a broad and holistic understanding of teachers' various perceptions.

**Stage 2:** At the second stage, a round of classroom observations were conducted before the participants read the sample test.<sup>7</sup> The washback effect at the micro level (e.g., the impact of the CET on classroom teaching and learning) was investigated. Immediately after the classroom observations, in-depth interviews were conducted with the six observed teachers. The data derived from this round of data collection were taken for the baseline data for this study and they would be compared to how teachers teach after the sample test was read. The purpose of establishing baseline data was to determine the type of teaching prior to the introduction of an examination (Wall, 1999; Wall & Horák, 2007) so that I would be able to determine whether the observed changes in the classroom could be attributed to the effect of the new exam. At this stage, the focus of the study evolved from an initially broad and holistic set of ideas to more specific questions related to teachers' reactions to the CET. This round of observations was conducted in November, 2005, one month before the implementation of the old CET.

**Stage 3:** The third stage involved the administration of a questionnaire. At the

---

<sup>7</sup> By the time my first round of observations were conducted, my case-study participants claimed that they had not read the sample test yet. The data collected then was taken as the baseline data in my study.

same time, the second round of classroom observations were carried out. It was noted that by this time, the participants had read the sample test.

**Stage 4:** The last stage of data collection consisted of another round of classroom observations and in-depth interviews. This round of qualitative data collection was conducted one month before the implementation of the revised CET. The purpose of this stage of data collection was to confirm the salient and recurring themes and patterns that had emerged from the data gathered in earlier stages and to see if the teaching of target test features accelerated right before the test. At this stage, all data sources were cross-examined to finally develop theory to explain the findings.

The data collection procedures are also presented in Table 8.

Table 8 *Data Collection Procedures*

Stage 1	Stage 2	Stage 3	Stage 4
Regular teaching time	Baseline data (before the participants read the sample test )	Time (after the participants read the sample test )	Time right before administration of the revised CET
–Document review	–Classroom	–Classroom	–Classroom
–Mapping of the site and sample selection	observation	observation	observation
–Group interview	–In-depth interview	–Questionnaire	–In-depth interview
–Washback at the macro-level	–Washback at the micro-level		

#### 4.5.5 Ethical Considerations

Ethical issues involved in collecting data, conducting research, and reporting the results were taken into careful consideration. My selection of participants was largely based on their willingness and interest to discuss the class and reflect on their own process of teaching. Early in the group interviews, I informed all the potential participants of the purpose of the research and also informed them and their respective schools that their identity would be concealed through use of pseudonyms. I also requested my case-study participants' permission for the use of digital audio recorders (MP3) during the process of data collection. Following the verbal approval to participate, the case-study participants were asked to sign an informed consent form. (See Appendix G for Informed Consent Form and Ethics Certificate.)

#### 4.5.6 Analysis of Collected Data

According to Bogdan and Biken (1998), data analysis is the process of bringing

order, structure, and meaning to the mass of collected data. This process entails uncovering patterns, themes, and categories. As indicated above, a MM approach combining both qualitative and quantitative methods of data analysis was adopted in this study. Firstly, a close examination of the pertinent documents (sample test of the CET, CECR, textbooks, etc.) was performed and reported in Chapter 3, and an intensive analysis of the characteristics of the CET were made and will be reported in the next chapter. Secondly, qualitative analyses of the case-study data as well as the group-interview data were conducted. The analyses involved the use of the constant comparative method (Bogdan & Biklen, 1998; Glasser and Strauss, 1990; Lincoln & Guba, 1985) in which the data were classified into categories. Specifically, I used inductive logic to identify and categorize emerging themes, perspectives and events from a mass of narrative data. Thirdly, quantitative analyses were performed, which involve frequency counts (and/or percentages by category), descriptive statistics and the following inferential statistical procedures: Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA) and Structural Equation Modeling (SEM). These methods were applicable to this study because they are commonly used to “analyze interrelationships among large numbers of variables and to explain these variables in terms of their common underlying dimensions” (Bentler, 1987, 1992; Kline, 2005; Purpura, 1998, 1999), also known as factors.

Finally, the different types of data sources were synthesized and integrated. To be specific, the qualitative data (through interviews and observations) were compared to the quantitative data (through the questionnaire) in search of patterns of agreement and disagreement. The purpose of the comparison was to find out whether the results from the qualitative data analysis were congruent with those from the quantitative data analysis. As a result of the comparison, the categories were combined and reorganized based on the common features found. The results of the comparison were presented with visual aids (charts, tables, etc.). The data were reviewed in a timely manner so that they could inform subsequent stages of the data collection process.

More details of how the data were analyzed are reported below.

#### *4.5.6.1 Analysis of the Data from Group Interviews and Case Studies*

In general, the data derived from the two types of interviews (group and individual) as well as the data from classroom observations were analyzed qualitatively by searching for themes and patterns. In the meantime, they were reduced and

synthesized using focused summaries pertaining to the research questions and other emerging issues.

In principle, the qualitative data analysis proceeded along the following steps.

#### *4.5.6.1.1 Organizing the data.*

First, I performed minor editing to make field notes and interview summaries manageable and retrievable. Then, I closely examined a small batch of data and jotted down the emerging themes and patterns. Having developed some preliminary categories of themes, I read through the data and grouped them according to these categories. By the time all the data had been examined, some of the initial themes and patterns developed were either modified or replaced by new ones.

Once again, I analyzed the data logically and assigned units of data into categories based on shared themes. The method that I used to analyze the data is called the constant comparative method (Strauss & Corbin, 1990). The remarks and assertions made by participating teachers during various interview sessions were constantly compared and contrasted throughout the research process.

#### *4.5.6.1.2 Developing theory.*

This step involves simplifying the codes and reducing the number of categories. Specifically, smaller categories were merged into a larger category. This procedure of combining and recombining the categories entailed data reduction. Eventually, this systematic process of induction enabled me to relate the data to theory. Drawing on the coding system developed by Strauss and Corbin (1998), I was able to build theoretical explanations, develop concepts and propositions from data. As a result, grounded theory was developed at this stage.

#### *4.5.6.1.3 Reporting the outcomes.*

This section provides a thick description of the research settings and a comprehensive account of the results. A holistic perspective was adopted when it came to presenting the participants' perspectives and views.

### *4.5. 6.2 Analysis of the Data from Classroom Observations*

The data from classroom observations were first coded according to the categories developed on the observation schedule. Then, frequency counts were applied based on

these labeled categories. The analysis involved a calculation of the duration of each classroom activity and instructional pattern in an average percentage of class time. After that, the percentages of time spent on each of the categories on the observation schedule were compared to determine the frequency of occurrence of various classroom interaction patterns and activities. After such an analysis, the observation data were compared to the data derived from the interviews to see whether they were compatible to each other. As Maxwell (1996) indicated, compatibility of interviews or observations is important.

#### *4.5.6.3 Analysis of Questionnaires*

##### *4.5.6.3.1 Computer softwares used.*

The computer program Statistical Package of the Social Sciences (SPSS, 15.0) for Windows was utilized to compute descriptive statistics and perform EFAs. In addition, Analysis of MOment Structures (AMOS 17.0) was used to perform CFAs, and analyses of covariance structure, known as SEM, to model the interrelationships among different aspects of the ‘teacher factor’ and their relationship to teacher practice.

##### *4.5.6.3.2 Statistical procedures.*

A brief introduction of the statistical procedures utilized in this study is outlined below. A flow chart of the procedures is presented in Figure 8. A detailed discussion of all these procedures is provided in Chapter 5.

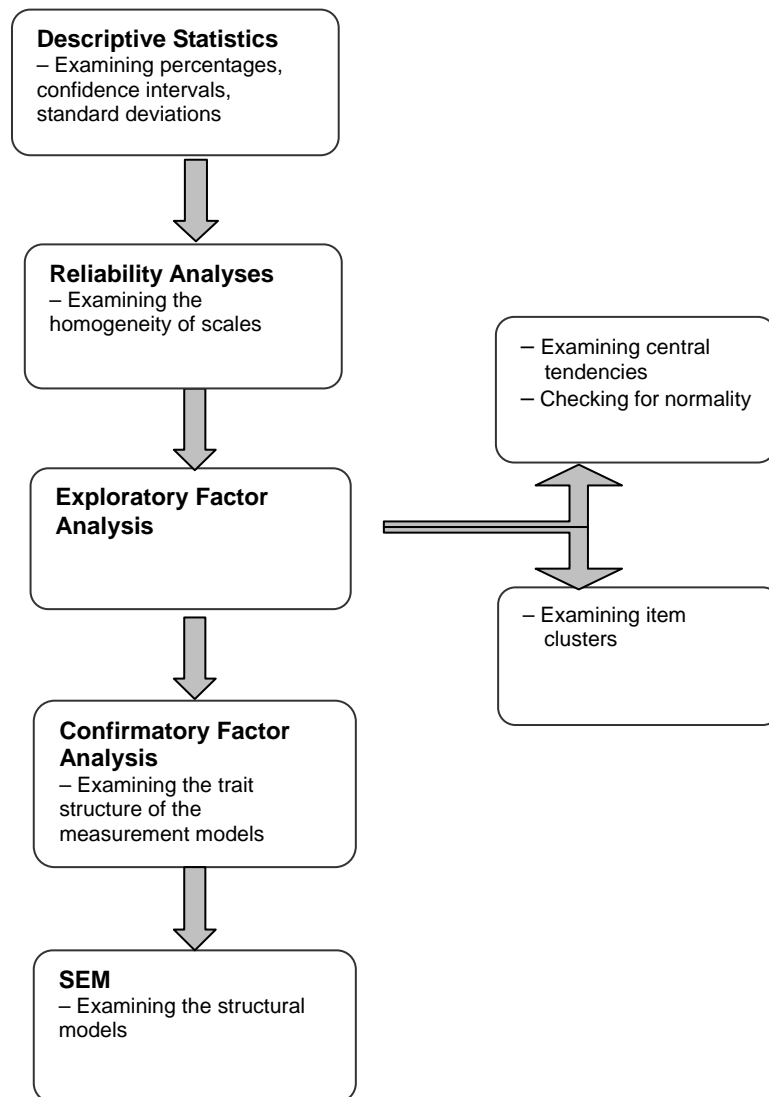


Figure 8. A flow chart of statistical procedures employed in the current study.

#### 4.5.6.3.3 Descriptive statistics.

When dealing with the questionnaire data involving various components of the ‘teacher factor’ (such as teachers’ beliefs of the CET, beliefs of test impact on teaching/learning and pedagogical knowledge, etc.), I first relied on frequency counts to find out about the frequencies and percentages of teachers’ responses by category, and I also examined the mean and standard deviation (SD) of each question. In addition, I used Confidence Interval (CI) to determine whether the mean of each question contains the true population mean.



#### 4.5.6.3.4 *Inferential statistics.*

Prior to conducting factor analyses and SEM, I first examined the means, medians, standard deviations, skewness, and kurtosis of each item and each variable to check for normality of the items as well as the variables. A normally-distributed variable is assumed to have a skewness and kurtosis near zero (Arbuckle, 2006). I then computed internal consistency reliability estimates (i.e., coefficient alpha) of the variables for all parts of the questionnaire items. Reliability for internal consistency was calculated using the Cronbach's Alpha Coefficient.

##### 4.5.6.3.4.1 *Exploratory factor analysis.*

After that, I proceeded to investigate how the various components of the 'teacher factor' are interrelated. To this end, EFAs were initially conducted to examine the patterns of correlations among the questionnaire items and group the items into categories pertaining to the 'teacher factor' (or to identify how various aspects of the 'teacher factor' clustered). The purpose of the EFA was to explore patterns in the data (e.g., how the questionnaire items cluster), while the purpose of the CFA (see Section 4.5.6.3.4.2 for detailed information) was to test the hypotheses (or models) to be discussed below. Principal Components Analysis was used to conduct the EFA. The following steps were followed in performing EFAs: creation of a correlation matrix, extraction of the initial factor solution, rotation and interpretation, and construction of a scale for further analysis (Kim & Mueller, 1978).

In general, EFAs were performed by varimax rotation. This method is reported to have the benefits of maximizing the variance of loadings. Subsequent to that, I examined how these clusterings related to those hypothesized earlier (or whether the set of questionnaire items designed measure the same underlying variable). Following the procedures outlined by Kim and Mueller (1978), first I generated a matrix of product-moment correlations among the items to be analyzed. After that, I examined the means, standard deviations, skewness and kurtosis of each item to see whether the items were normally distributed. I then examined the appropriateness of these data for EFA based on (1) Barlett's test of sphericity, (2) the Kaiser-Mayer-Olkin (KMO) test of sampling adequacy and (3) the determinant of the correlation matrix. Following that, the communality coefficients, eigenvalues, scree plot, and the component matrix coefficients were computed. These item-level analyses were performed as a preliminary step to examine the clusterings of the items. Based on these analyses,

composite variables were formed, which were then used in subsequent analyses.

Correlation analysis was also employed to check for possible significant relationships among the various aspects of the ‘teacher factor’.

#### 4.5.6.3.4.2 *Confirmatory factor analysis.*

Following EFAs, CFAs were performed to examine the trait structure of all the measurement models. As CFA is a special case of the structural equation model (SEM), it will be presented together with SEM.

#### 4.5.6.3.4.3 *Structural equation modeling.*

Subsequent to the CFAs, I utilized SEM to posit and test models related to various aspects of the ‘teacher factor’ under study. SEM was described by Purpura (1998, 1999) as a multivariate analytic procedure for representing and testing (1) hypothesized inter-relationships between observed and latent variables, and (2) hypothesized inter-relationships among latent variables, based on substantive theory or previous empirical research. In the present study, the three hypothesized structural models were intended to assess (1) how various aspects of the ‘teacher factor’ (teacher beliefs, knowledge base and experience) relate to one another, (2) how the interrelated aspects of the ‘teacher factor’ might relate to teacher practice; (2) how the interrelated aspects of the ‘teacher factor’ might relate to teacher beliefs of the impact of the CET on teaching. AMOS makes available a number of estimation methods. The present study adopted the maximum likelihood (ML) and the ML Robust methods. ML method was employed where the data meet the distribution assumption of multivariate normality, while ML Robust estimation was used where the data were found to show multivariate non-normal distribution.

To assess how well the hypothesized model fit the data, four indices of goodness of model fit were used: (1) the probability value of the chi-square statistics, (2) the Satorra-Bentler scaled chi-square statistics in cases of some multivariate non-normal distributions, (3) the Comparative Fit Index (CFI), and (4) the Root Mean Square Error of Approximation (RMSEA). Of the indices, CFI was deemed as more useful than other indices in that it is not dependent on sample size (Bentler, 1987, 1992; Kline, 2005; Purpura, 1998, 1999). According to them, values of 0.90 and above are considered acceptable indices of model fit.

The softwares used for this study included SPSS for Windows 15.0 and AMOS

17.0. A detailed description of how the data was analyzed is shown in Table 9.

Table 9 *Procedures of Data Analysis*

Analysis of Document	Case Studies	Questionnaire	Integration of Data
(e.g., CECR, textbooks, CET) –goals –contents –skills –methodology	<b>Observation</b> –coding –frequency counts  <b>Interview</b> –organizing data –categorization –developing theory (Glaser & Strauss, 1990)	<b>Closed items</b> –SPSS –descriptive stats (frequency counts, SD) –inferential stats (factor analyses, SEM)  <b>Open-ended questions</b> –Constant comparative method	<b>Questionnaire</b> + Interview + Observation  <b>QUAN<math>\longleftrightarrow</math>QUAL</b>
inferential stats: inferential statistics descriptive stats: descriptive statistics			

This chapter has presented and discussed several aspects of the research design adopted in the present study. First, an introduction was given on the application of a mixed-method and emergent design. Second, some general background information was given about the participating teachers and research sites. Third, a description was given of the instruments, along with a brief rationale for using them. Fourth, the procedures of data collection were explained. The final section provided a description of the procedures and methods of data analysis.

The next chapter presents the results of the study.

## Chapter 5: Presentation of the Findings

### 5.1 Introduction

The methods employed for gathering data have been intensively discussed in Chapter 4. This chapter presents the findings of the study (both qualitative and quantitative) in separate sections. It begins by examining the characteristics of the CET drawing on the framework of language task characteristics provided in Bachman and Palmer (1996). It then reports the qualitative findings derived mainly from interviews and classroom observations. These findings will be presented and summarized by themes. After that, the quantitative results obtained from questionnaires will be reported and summarized.

Given the substantial amount of data yielded from this study, a detailed description of all of the findings of this research is beyond the scope of a doctoral dissertation such as this. I was thus compelled to limit the presentation of results in this thesis to only the findings that specifically addressed the research questions.

The global research question guiding the study is: What role does the ‘teacher factor’ play in washback in the Chinese university context?

The major findings presented in this chapter serve to address one global research question and two secondary questions of this study: What role does the ‘teacher factor’ play in washback in the Chinese university context?

1. To what extent and in what form does washback exist in China in terms of its effect on teacher beliefs (e.g., beliefs about the CET and its impact, beliefs about teaching and learning), and classroom behaviors (e.g., content and particularly teaching methodology)?

2. How is the ‘teacher factor’ manifested in such a washback effect? What aspects of the ‘teacher factor’ (e.g., beliefs, knowledge, past experiences) contribute to the way that teachers interpret and react to washback?

### 5.2 Analysis of the Test, the CET

It is widely believed that when we design a language test or evaluate its potential usefulness, two critical measurement qualities we need to give consideration to are: reliability and validity. Validity relates to the extent to which meaningful inferences can be drawn from test scores (Bachman, 1990). In contrast, reliability concerns the consistency of measurement. Of the validity considerations for a language test,

construct validity is viewed as pivotal. It is often used to refer to the extent to which we can interpret a given test score as an indicator of a test takers' language ability(ies). The term can be interpreted to mean that if a test has good construct validity, it is a good indicator of test takers' language ability and vice versa. Then how can we determine the construct validity of a language test?

According to Bachman and Palmer (1996), when we consider the construct validity of a score interpretation, both the construct definition and the characteristics of the test tasks should be taken into account. They place special emphasis on test tasks claiming that they should be carefully selected and their characteristics should be adequately described. Construct definition is defined by Chapelle et al. (1997) as a theoretical description of the capacity that a test is supposed to measure. Through construct definition, we are supposed to clearly and specifically define what is to be measured. Bachman and Palmer (1996) seem to suggest to us that the more the test tasks reflect the construct definition, the higher the construct validity. From their perspective, construct validity is affected to some extent by the characteristics and content of the test tasks. In this regard, there is an obvious need to examine the task characteristics of the CET.

In the field of language testing research and practice, the framework proposed by Bachman and Palmer (1996) is often taken as a theoretically grounded guideline for analyzing the characteristics of a test. This conceptual framework consists of a set of principles involving five facets of tasks: setting, test rubric, input, expected response, and relationship between input and response. But here, I will discuss four features in particular which I think are crucial to my study. In the section below, I will visit the concepts of each facet and analyze the CET in light of the features of this framework.

### 5.2.1 *Task Characteristics*

#### 5.2.1.1 *Rubric*

First, the test rubric facets are portrayed by Bachman and Palmer (1996) as those characteristics of the test that provide the structure for a particular test. These characteristics include: instructions and the structure of the test (e.g., how the test is organized), time allotment (e.g., the duration of the test as a whole and of the individual tasks), and scoring procedures (e.g., how the language that is used will be evaluated, or scored).

A close look at the CET indicates that its instructions are not clearly specified with

respect to the procedures for taking the test. No examples are provided as to how to perform a task. For instance, in Part II, the section of Reading Comprehension (Skimming and Scanning), the test takers are not provided with explicit directions except being told “to answer the questions on Answer Sheet 1” (refer to Appendix C for the format of the sample test of the CET). Nor are instructions given on the Answer Sheet. Further, little information is given on score distribution, criteria for correctness, and procedures for scoring the responses (e.g., how the test will be scored as well as the steps involved in scoring the CET). Table 10 below outlines the structure of the revised CET.

Table 10 *A description of the structure of the revised CET*

Components	Task Type	Number of tasks	Time allotment (Minutes)	Score Distribution
<b>Part I Writing Tasks</b>	Writing	1	30	Not Provided
<b>Part II Reading Comprehension (Skimming and Scanning)</b>	Word/Phrase Completion	1	15	Not Provided
<b>Part III Listening Tasks</b>	MC	3	35	Not Provided
<b>Part IV Reading Comprehension</b>	CD	2	25	Not Provided
<b>Part V Integrative Tasks (Cloze)</b>	MC	1	15	Not Provided
<b>Part VI Translation</b>	Chinese – English	1	5	Not Provided
<b>Total</b>	5	9	125	

It should be noted that MC items still take up a larger percentage of the test. Even though Compound Dictation (CD) is a task included in the listening part, it makes up only a small percentage. As is shown in Table 10, test tasks in Part III, IV, and V are mainly measured through MC questions. According to Bachman and Palmer (1996), these task types are limited in that they may not be appropriate for measuring different areas of language ability.

#### 5.2.1.2 *Input*

The input facets examine the format in which input is presented and the characteristics of the language that are embodied in the input. The format includes

features such as channel, form, language, length, type, degree of ‘speededness’, and vehicle of input delivery, while the language characteristics include both organizational and pragmatic characteristics of how the language is organized. With respect to the CET, the main channel of input is visual. However, in the section of Listening Comprehension, the input is in both visual and audio format. The form of input is language, and the language by means of which input is delivered is English. On the whole, the input is presented in extended discourse. Since the majority of task types in the CET are MC questions, the input consists of items. The degree of ‘speededness’ is high, since the rate at which the test taker is expected to process the input information is high. In the section of Listening Comprehension, the vehicle in which input is presented is reproduced via audiotapes.

#### 5.2.1.3 *The Nature of Language Input of the CET*

The organizational and pragmatic characteristics of the CET can be divided into subcategories. The former can be divided into grammatical and textual aspects, while the latter can be classified into functional and sociolinguistic aspects. The grammatical aspects consist of vocabulary, morphology, syntax, phonology, and graphology. The textual aspects involve cohesion and rhetorical or conversational organization. Functional aspects can be categorized into functions such as ideational, manipulative, heuristic, and imaginative. Topical characteristics refer to the type of information provided in the input, such as personal, cultural, academic, or technical.

As far as the CET is concerned, it involves a broad range of vocabulary and grammatical structures as well as a wide range of cohesive devices and topics. An examination of the test shows that a testee’s linguistic knowledge is adequately tested, because he or she has to demonstrate that he or she has the linguistic knowledge to process the input information. However, his or her sociolinguistic competence is only tested to some extent. For instance, in Part IV, the section of Reading Comprehension, this competence is tested, for a testee needs to guess word meanings in context. But in general, the language input of the CET seems unnatural, for it seldom relates to everyday situations, as is evidenced in Appendix C. In other words, the test tasks do not often correspond to tasks in the language use (TLU) domain. Based on the view of Bachman and Palmer (1996), if we want to use the scores from a language test to make inferences about individuals’ language ability, we must be able to demonstrate how performance on that language test is related to language use in specific situations

other than the language test itself.

The format and the characteristics of the language input of the CET are presented below:

*The Format of Language Input of the CET*

Format: channel, form, language, length, type, degree of speededness, and vehicle.

Channel: both aural and visual

Form: language

Length: presented in extended discourse

Type of input: item

Degree of speededness: high

Vehicle: reproduced (via audio)

*Language Characteristics*

Organizational characteristics:

Grammatical characteristics

– Vocabulary: broad range of general vocabulary.

– Morphology and syntax: broad range of organized structures

Textual characteristics:

– Cohesion: cohesive, with a broad range of cohesive devices.

– Rhetorical organization: relatively broad range of artificial conversation

*Pragmatic characteristics*

Functional

– Ideational and manipulative

Sociolinguistic

– Standard English

– Register: formal

– Natural: unnatural

– Cultural references and figurative language: minimal

#### 5.2.1.4 *Expected Response*

The facets of expected response focus on the format in which a response is produced and the nature of the language that is employed in the response. Since the characteristics of the format and the nature of language have been introduced in the facets of the input, they will not be discussed again here. Below is a brief discussion of those characteristics relevant to the CET.

Drawing on the principles in Bachman and Palmer's (1996) framework, there are three categories of responses: limited production, selected production, and extended production. In the case of the CET, the test-takers' responses on the CET generally fall into the first two categories, for their production in the majority of test tasks, except in the Writing section, is limited /selected rather than extended. The fact is that MC and



completion questions can only elicit selected and limited production responses (e.g., a single word or phrase).

#### 5.2.1.5 *Facets of the Relationship Between Input and Response*

Facets of the relationship between input and response address issues such as reactivity, scope of relationship, and directness of relationship.

The reactivity refers to “the extent to which the input or the response directly affects subsequent input and response” (Bachman and Palmer, 1996, p.55). In terms of reactivity, there are three types of test tasks: reciprocal, non-reciprocal and adaptive. Those features that determine these test tasks include feedback and interaction between language users. Reciprocal test tasks concern both feedback and interactions and non-reciprocal tasks involve neither of the two. In comparison, the adaptive test tasks only relates to feedback.

With respect to the scope of relationship, Bachman and Palmer (1996) define it as “the amount or range of input that must be processed in order for the test taker or language user to respond as expected” (p. 55). In addition, another concept introduced by Bachman and Palmer (1996) is the directness of relationship. They explicitly interpret it as “the degree to which the expected response can be based primarily on information in the input, or whether the test taker or language user must also rely on information in the context or in his own topical knowledge” (p. 56).

A close look at the CET indicates that it contains neither reciprocal nor adaptive test tasks, for test takers are not required to engage in tasks such as giving feedback or having interactions with one another. As for the scope of relationship of the test, it ranges from broad to narrow. For instance, within the sections of reading comprehension and listening, test takers are required to process a large amount of input to make their choices, while in other sections (Part II, V, and VI, see Table 10 for detailed information) the scope of relationship is limited or narrow, for test takers do not need to process a large amount of input to complete the test tasks.

What we need to note is that the degree of the relationship is fairly direct in the case of the CET, since most of the information that can be used in the responses is provided or can be inferred in the input.

To sum up, through the above analysis, we can see that the CET tasks do not often correspond to tasks in the language use (TLU) domain. In other words, the content of

the CET does not include many items related to daily life situations. Of the small number of tasks, writing is the only one that represents a real-life task. Apart from it, other test tasks, both the tasks in MC format and those in constructed response formats (e.g., word/phrase completion in Part II, compound dictation in Part III, and translation in Part VI) can hardly relate to every day situations. In other words, they do not assess students' ability to use language through real-life situations. A case in point is that although CD as a constructive form of testing task is used in the CET, it bears little resemblance to the listening tasks that the students are engaged in in an EFL class, for real classroom listening tasks are, as Bachman and Palmer (1996) described, "interactive, conversational tasks involving a give and take".

The above analysis also reveals that the revised CET does not assess test takers' overall language skills integratively. As mentioned above, a test taker's linguistic competence is adequately assessed, for the majority of test items only test his or her discrete areas of language knowledge. Apart from writing, other test tasks of the CET cannot be interpreted as assessing test takers' sociolinguistic competence or strategic competence. In the section of Writing Task, a testee's language competencies are integratively tested, for he or she is expected to produce language in the same way as in everyday contexts. In addition to linguistic knowledge, the task involves other areas of language knowledge as well as metacognitive strategies. For instance, the language the testee produces must not only be accurate (evidence of linguistic competence) but also appropriate (evidence of sociolinguistic competence) and coherent (evidence of discourse competence).

One drawback of the CET is that the test does not include an assessment of students' ability to perform speaking tasks. Since only a small number of students (those who get scores above 80 on the CET written test) are given the opportunity to sit for the SET, a large number of testees do not have the opportunity to demonstrate strategic competence in the oral interview. In this regard, although the CET is claimed to assess students' ability to use language through real-life tasks, its results or test scores cannot truly reflect test takers' communicative competence.

Owing to the fact that little evidence can be retrieved to show that the CET score reflects the area(s) of language ability it sets out to measure, its construct validity is called into question.

### 5.3 Analysis of the Teaching Materials

It is important to realize that the teaching materials selected by teachers vary from class to class. In general, there are four major types of materials used in the observed CE classes: coursebooks, supplementary teaching materials, CET-related materials and the CET test papers.

The supplementary teaching materials selected by the participants include materials from mass media, such as English newspapers, the Internet, and TV/radio programs, etc. CET-related materials concern those materials used for fostering students' test-taking strategies. The CET test papers here refer to test papers previously used in the CET. It is worth keeping in mind that it is common practice for the Chinese EFL teachers and students to do one or two CET test papers prior to the test. The key reason is that they want to use the papers to familiarize their students with the test format.

We should note that the course-books chosen by the research sites differ from school to school (See Table 11). These books were respectively produced by Foreign Language Teaching and Research Press, Higher Education Press and Fudan University Press, and Shanghai Foreign Language Education Press.

The coursebooks employed by the research sites were recommended by the ME. They were claimed to be designed both to reflect the CECR in their objectives, emphasis and approach, and to reinforce general communication goals. One striking difference from the traditional textbooks is that they break through the traditional way of compiling course books based on grammar systems, progressing, rather, according to language complexity. Another notable difference from the traditional textbook is that more related cultural information has been incorporated in the books. Table 11 outlines the coursebooks used in the observed classes.

Table 11 *Course-books Used by the Observed Teachers in Class*

	S1	S2	S3
<b>Coursebook</b>	New Horizon College English	College English (New Edition)	21st Century College English

## 5.4 Results from Interviews and Classroom Observations

### 5.4.1 *Introduction*

As explained in Chapter 4, group interviews were conducted with 30 CE teachers at the three research sites. Six teachers were then observed and interviewed individually as in-depth case studies. Qualitative data collection lasted over a period of seven months. In this section, the qualitative results obtained from the group interviews and subsequent case studies will first be presented school by school and participant by participant, and will then be summarized in general. The reason why I do so is that I realized after an intensive analysis of the data that many of the participants' claims and assertions are context-dependent. Only when they are considered within the specific local context (e.g., group-interview context) can we make better sense of them. In the meantime, instead of initially presenting this set of data by round/period, I chose to summarize the results following some shared patterns in relation to teacher beliefs, knowledge of teaching pedagogy and behavior(s). But this is followed by a general summary of the case-study results in which the data will be presented by round as well as based on the hypothesis outlined in Chapter 4. It is believed that such a way of reporting the results may give a better picture of how teachers in different school settings perceive and react to the CET reform, on the one hand, and how teacher individual differences are manifested, on the other. In addition, instead of using the real names of the schools and participants, I will use pseudonyms. For the purposes of this study, the three schools were coded S1-S3 and the six teachers were coded T1-T6.

To set the stage for discussion, I will initially give a description of the background information of the research sites. Then I will address personal details of the participants including their age, gender, highest education level completed, number of years of teaching, previous language-teaching/learning experiences and training received in ESL, etc. Following that, I will report their teaching performances which I examined, either in their classrooms, or in the audio-taped lessons, as well as the statements, assumptions and accounts for their practices made or expressed by them either at the pre-observation group interviews at in the post-observation individual interviews. In light of the abundance of data collected, only general information will be given here. Specifically, I will focus only on the three major themes that have emerged pertaining to my research questions. The first theme describes the impact of the CET on teaching and learning. It includes: beliefs teachers articulated about the

CET and its impact. The second theme reports on teachers' curriculum knowledge (e.g., knowledge of the CET and the CECR). The third theme presents the real evidence of the effects of the CET on their teaching that has emerged during the three rounds of classroom observations. The last theme has emerged under the category of teachers' beliefs about teaching and classroom scenarios of how they teach. It describes both teachers' various conceptions of teaching and learning and their real practice in the classroom. In terms of teacher practice, the classroom scenarios portrayed involve: their interaction patterns, various activities organized, focus of instruction (e.g., focus on knowledge or competence), skills practiced, and medium of instruction, etc.

#### *5.4.2 Background Information of the Research Sites*

As introduced earlier, the three research sites of this study are all located in the same Chinese city. They differ considerably in terms of their ranking, their academic quality, the quality of their students, teaching facilities, and textbooks used.

S1 is well recognized as a key university in the country. The students studying there are considered to be high-level when compared to S2 and S3. Although S2 and S3 are not as prestigious as S1 in the country, they are also regarded as key schools at the municipal level. However, S1 and S2 are pilot schools, while S3 is not. While the conditions of these observed schools vary from site to site, all the classrooms I visited are well equipped with modern teaching equipment (e.g., a pre-installed desktop computer and overhead projector, etc.). In addition to the difference in the size of classrooms, the class size of these schools also differs a great deal. At S2 the number of students is limited to 30 per class, much smaller than that at S1 and S3 where the number ranges from 40 to 59 per class.

While a difference is shown in the textbooks chosen at different schools, participants at the same school not only use the same textbook, but also teach the same textbook unit.

#### *5.4.3 Profile of the Participants*

All the six case-study participants are currently CE teachers working at three different Chinese universities. Of the six participants, four are females and two are males. Five of them are middle-aged, between the ages of 35 and 45, and these teachers each have a teaching experience of more than 10 years. One of them is in her

late 20s and has taught English for about eight years. Among them, three have obtained a bachelor's degree, while the three others have received a master's degree. Their teaching hours range from 6-8 hours to 9-11 hours per week. At the time of my observations, they were offering courses at the same level of proficiency (to the sophomores). None of them have the experience of studying or working abroad. Two participants reported having received teacher training in ESL and one teacher claimed to have been exposed to task-based activities. Table 12 presents the general characteristics of the participants.

Table 12 *General Characteristics of the Case-study Participants*

	<b>T1</b>	<b>T2</b>	<b>T3</b>	<b>T4</b>	<b>T5</b>	<b>T6</b>
<b>Sex</b>	<b>Male</b>	<b>Female</b>	<b>Female</b>	<b>Female</b>	<b>Female</b>	<b>Male</b>
<b>Age</b>	30-39	40-49	40-49	20-29	30-39	30-39
<b>No. of years of teaching</b>	15-19	15-19	15-19	5-9	10-14	15-19
<b>No. of teaching hrs per week</b>	6-8	6-8	Missing	6-8	9-11	9-11
<b>Class size (No. of students in class)</b>	40-49	50-59	20-29	20-29	50-59	40-49
<b>Degree</b>	M.A.	M.A.	B.A.	B.A.	M.A.	B.A.
<b>Experience of being in an English-speaking country</b>	No	No	No	No	No	No
<b>Training in teaching methodology</b>	Yes	No	Yes	No	No	No
<b>Exposure to task-oriented activities</b>	No	No	Yes	No	No	No

#### 5.4.4 Washback of the CET in the Case-Study Setting

##### 5.4.4.1 How Teachers Perceive and React to the CET Reform at School One (S1)

###### 5.4.4.1.1 Teachers' articulated beliefs about the CET and its impact.

The group interview data revealed that the teachers at S1 maintained a negative attitude toward the CET reform. Some teachers overtly stated having grown tired of this type of reforms, asserting that these reforms came far too frequently. Other teachers added that they disliked changing back and forth according to the reforms. During the group interview, the chair of the CE program at S1 expressed her frustration over the current CET reform, saying:

<sup>8</sup> 改革过于频繁，上面好像也没有一个统一的想法。在这种情况下，我们也很迷茫，我们也不应该为了改革而改革。

[Translation: The reforms are too frequent. It seems that the authorities have not achieved a uniform view of which way to take. In such a case, we're at a loss what to do and find it hard to adapt to the changes. In fact, we should not reform for the sake of reform]

According to this chair, highly frequent reforms led to teachers' indifference or numbness to the current reform. One interesting finding is that the two participants at S1 were also skeptical about the revised CET. They wondered whether there would be fundamental changes in the revised CET. Meanwhile, they criticized the MC format that the CET often relied on. The following excerpt is taken from T1's comments.

尽管改革多了，但也没有质的变化，主要是考试形式没变，其害处就是学生只做出选择判断，无需动脑，其后果就是学生依然不能用英语表达思想。

[Translation: There might be some quantitative rather than qualitative changes, for MC format will continue to be utilized on the revised CET. The harm this type of question does is that rather than racking their own brains, students only need to make judgments or choices. Consequently, they are unable to convey their ideas in English.]

Despite their negative attitude toward the CET, the teachers at S1 reported being unaffected by the test. Based on their chair's accounts, there was no preparation whatsoever for the test at S1. She said,

我们学校从不备考，我们只考虑学生的需要。

[Translation: Our university has never made any preparations for the test. We simply teach according to the needs of our students.]

The view T1 held about the CET is consistent to that of the chair. In his view, the CET can hardly produce any effect on their instruction. He was dissatisfied with the MC test tasks employed by the CET designers, which led to his skepticism about the validity of the test. Thus, he complained,

新的四级题型虽然做了点调整，但变化不大，依然以多项选择为主，因此还是无法检测学生的交际能力。

[Translation: The new CET, although with a couple of changes in the test tasks, still cannot assess students' ability to communicate in English, for the majority of test tasks are in the form of MC.]

During the interview sessions, he reiterated that the CET can exert influence on

---

<sup>8</sup> All translations from the original Mandarin are those of the writer.

average (or second-class) schools rather than on key schools (such as S1). According to him, the lower ranking the school has, the more impact the CET produces on it. The reason he provided was that average schools need to use the CET scores to upgrade the status as well as reinforce the reputation of their schools. In contrast, the status and reputation of the key schools are so well-established that they do not need to use the CET scores to reinforce them. In his opinion, students of key universities encounter little difficulty in passing the CET. In comparison, T2's perception concerning the impact of the CET differs in part from that taken by other interviewees at S1. Although she shared their view that the CET has little effect on teaching (which partially overlaps her responses in the survey, see Table 13), she presented a different view that the CET can affect learning.

However, what confused me was that T1's responses in the survey contradicted what he expressed at the interviews. When asked whether the CET could produce washback effects on teaching (e.g., whether it would help to improve the way he taught, shift his instructional focus from linguistic knowledge to language use, and give direction as to what aspects of language needed to be taught ), all his answers were positive. Why are there inconsistencies between his responses made in public and those made in the survey? The question will be pursued in the following chapter.

Table 13 *Washback Effects of the Revised CET Based on T1 and T2's Accounts*

(Y=Yes, N=No)

Participants	Improve the way they teach	Shift instructional focus from linguistic knowledge to language use	Give direction as to what aspects of language needed to be taught
T1	Y	Y	Y
T2	N	N	Y

#### 5.4.4.1.2 Teachers' curriculum knowledge.

It was noteworthy that at the group interview when asked whether they knew the objectives of the revised CET and the CECR, why the reform was carried out and whether there were differences between the old CET and the revised CET, the teachers at S1 said that they had little idea of them. At that time, I thought that owing to their skepticism about the revised CET, they deliberately chose to say they did not when they did. However, it turned out that what the two observed teachers articulated in the post-observation interviews did show that the teachers' knowledge of the CET



and the CECR was quite limited. They both claimed that they had not read them yet.

#### 5.4.4.1.3 *Observational evidence of the CET impact.*

By observation, most of the activities the two participants conducted were focused on the content of the text rather than the CET. Except in R3, the CET was seldom referred to and few test-related activities were found in their classrooms. Nonetheless, during R3 when the time for the administration of the CET was approaching, the course was totally test-focused. Both participants stopped teaching textbooks and replaced them with test-like worksheets. Such a practice seems to be incompatible with the assertion made by the chair that a university like S1 never carries out any test preparation. When explaining why regular teaching was interrupted, the two participants claimed that such a practice was arranged by the school and also in accordance to the students' needs. They explained that they did it to familiarize the students with the test format.

#### 5.4.4.1.4 *Teachers' beliefs about teaching and classroom scenarios of how they teach*

##### **T1**

T1 initially impressed me as an advocate of student-directed instruction. He explicitly mentioned that the way he taught was in line with his teaching beliefs and the needs of his students. Throughout the group interview, I marveled at his ideas and beliefs about his role as a CE teacher. And I had expected that he would incorporate some learner-centered activities in class. But contrary to my expectations, no such activities were observed throughout the observation process. In fact, his class turned out to be typically teacher-centered. In class, he was observed spending a lot of time lecturing on linguistic knowledge. Pair work/group work activities were never implemented. Meanwhile, although I observed him occasionally talking vigorously to his students by asking them comprehension questions related to the texts being taught, his instruction was conducted in Chinese. Moreover, he supplied them with answers most of the time. The results reveal a big mismatch between what he said and what he practiced. One possible reason for such a mismatch as well as his extensive use of Chinese in the classroom was that his own language proficiency level was not very high, which made it impossible for him to adequately express his ideas in English. I made such an inference based on the statements he made at one individual interview.

When addressing the role of language proficiency in CE instruction at one interview session, he stated that he occasionally found it hard to convey his ideas in English. Another possible reason was that although he previously expressed his interest in CLT and task-based teaching approaches, he might not be well-equipped with the hands-on knowledge which would allow him to manipulate the approaches as he wished. Evidence for this inference can be found in the conflicting remarks made by him on different interview occasions. During the group interview session when he was talking before his colleagues, he criticized the structural approach (or grammar-translation approach) saying:

这种方法太陈旧过时了。我觉得大学英语课应围绕学生的输出能力，因为我们教学目标是提高学生的交际能力。

[Translation: This approach is too stereotyped and backward. I believe that a CE class should revolve around productive skills, since the goal of our teaching is to enhance students' communicative competence.]

In contrast, at one individual interview, he articulated some interesting beliefs about teaching communicatively:

尽管我们的目标是帮助学生提高交际能力，但却不容易实现。所以说，纯粹的交际是骗人的。最主要的是学生要打好语法基础。

[Translation: Although helping students acquire communicative competence is our goal, it seems unattainable. Therefore, pure communication is deceiving. Above all, students need to lay a solid foundation in grammar.]

His conflicting remarks revealed that his stated beliefs at the group interview are not necessarily his real beliefs.

## T2

Through direct observations, it was apparent that T2 taught her lessons using the structural approach. The instructional pattern she followed was a monotonous one, starting with reading the text aloud herself, then presenting and explaining language points (e.g., vocabulary and grammar), and ending up analyzing the text paragraph by paragraph and translating them into Chinese.

One strong impression from the observations was that the teacher's lecture dominated the class. I observed the students being ignored most of the time and rarely called upon in class. Interestingly, T2 herself admitted that her classes were very much teacher-centered. When asked why no group-work activities were organized,

she said,

我曾试过小组活动，但发现这种形式的活动效果不好，是浪费时间，小组活动时学生根本不用英语，而是用汉语聊天。

[Translation: I have tried using group work, but I find that such kinds of activities are ineffective. It is a waste of time to conduct them. The reason is that instead of using English, my students tended to talk to one another in Chinese.]

She also expressed her dilemma as to what to teach in class by saying:

过去，我很少给学生讲课文的语言点，而是讲内容意义，但他们却抱怨说课上什么也没学。结果，我只好花很多时间讲语法和词汇了。我有时真不知该怎么教，教什么了。

[Translation: I used to teach by focusing on the meaning of the text rather than on language points, but my students complained about the meaning-based instruction saying that they had not learned anything. As a result I had to concentrate more of my class time on teaching vocabulary and grammar. I sometimes feel I do not know what to teach and how to teach.]

Her statements seemed to imply that whether or not teachers could organize communicative activities depends on the motivation as well as the English level of their students. They also seemed to suggest that student beliefs or student role in the classroom constitute a barrier to the implementation of communicative activities in their instruction. This view was corroborated by T1 who also held that students preferred to be taught more vocabulary.

In addition, like T1, she reflected on the challenge she confronted when it came to organizing communicative activities. From her accounts, her own low oral proficiency also poses a constraint on her instruction. At one interview session, she echoed T1's claim saying that she sometimes found it hard to convey her ideas freely in English as well. When accounting for the reason that led to teachers' low oral proficiency, she articulated,

这与国家及学校的政策导向有关。一般情况下，没人注意你教好教坏。只要你发表的文章够数，别人就认为你圆满完成了工作。

[Translation: This has a lot to do with the policies and orientations of the schools as well as the government. As a rule, little attention has been given to whether you teach well or not. As long as you have published a certain amount of journal articles and done well in research, you are considered having accomplished your job.]

What she articulated above aided in interpreting her views regarding the relevant policies of the school and government. She further reminded us that similar

challenges were also faced by many other CE teachers.

#### 5.4.4.2 *How Teachers Perceive and React to the CET Reform at School Two (S2)*

##### 5.4.4.2.1 *Teachers' articulated beliefs about the CET and its impact.*

Like the teachers at S1, during the first stage of research (the group-interview session and the first round of observation), the teachers at S2 declared that they had not read the sample test yet. Although the teachers of S2 are not as critical and resistant to the CET reform as those of S1, their attitudes toward the CET, on the whole, still sound negative. While they believed in the long-term effect of the CET reform, the interviewees seemed to panic about the change(s). The following excerpt from Teacher 3 illustrates the point.

改革本身没有过错，从长远看会有效果，但目前我们却难以适应这种变化，因为没有过渡。既然变化对于学生是一个大的挑战，对老师也是一种压力。

[Translation: There is nothing wrong with the goal of the reform. Its long-term effect might be good, but right now we find it hard to adjust to the change, because there is no transition. Now that such a change poses a big challenge to our students, correspondingly we teachers also have pressures.]

According to these teachers, the change should be gradual, but not too abrupt. T3's view of the CET is similar to that of others. She further explained why the students were so nervous about the CET.

学生感到畏惧，因为他们不仅缺少语言技能；也缺少考试技能。如果语言能力强，也就不需要害怕了。正因为学生不了解新题型，所以问过我好多次新题型是什么样的。

[Translation: The students are panicked, because they are deficient in both their language skills and testing skills. If they have acquired high language skills, they do not need to panic. Since my students are not familiar with the format of the revised CET, they asked me a couple of times about what the new test is like.]

In addition to the students' anxiety, the teachers themselves also seemed to be confused about how to teach. The following remarks made by T3 illustrated such confusion.

课时这么有限，我们真不知该如何加强听力训练。所以，我们也搞不清什么该教、什么不该教。

[Translation: We do not know how we can place more emphasis on listening within the limited class hours. So, we are vague about what kind of language should be taught and what should not be taught.]

She firmly believed that the most effective way to emphasize listening is to increase

the class hours in listening. Moreover, the way she interpreted the revised CET was also different from the way the test developers did. She said,

我不认为新题型减少了语法的比重，因为增加了翻译，这无疑是检测学生语法知识的。  
[Translation: I don't think that the percentage of grammar has been cut down in the new CET, since the inclusion of "translation" as a test task is undoubtedly to assess students' command of grammatical structures.]

However, contrary to the interviewees at S1 who denied the influence of the CET on teaching, the teachers of S2 acknowledged that the CET could have great impact on both teaching and learning. They reported that they had to concentrate their teaching on what was related to the CET. Below is the rationale provided by one teacher for why this had to be done.

我们的教学目标与学生的学习目标应该一致。总体上看，学生的目标可划分为近期目标和远期目标。他们的近期目标是找一份好工作或出国深造。但是，他们很难将他们目前所学的语言与他们将来的工作联想在一起。因此，我们的教学应考虑他们的近期目标。

[Translation: Our teaching objectives should be consistent with students' goals. As a rule, students' goals can be classified into immediate ones and distant ones. Their immediate goals are to acquire testing skills and pass the test, while their distant goals are to find a decent job or go abroad to study. However, for the time being, they can hardly associate the language that they are learning with what they are going to do in the future. Thus, we need to teach in line with their immediate goals.]

Both T3 and T4 were aware of the effect of the revised CET on teaching and learning. Nevertheless, the way they commented on the CET and anticipated its effect on teaching both at the individual interviews and in the survey varies greatly. On the one hand, from T3's interpretation, its effect on teaching is negative, while its effect on learning is positive. Below are some of the comments she made.

大部分学生都被动。因为四级考试与学位挂钩，所以他们不得已为其而学。所以说，它对学生的学习起积极作用，起码让他们多记些词汇。没有考试，学生更不知怎么学了。

[Translation: Most of the students are passive learners. But since the CET is linked to their diplomas, they are forced to learn toward it. So, I believe that its impact on learning is a positive one. At least it can facilitate them to memorize some vocabulary. Without it, the students may not know what to learn.]

This quote is representative of other group interviewees' perceptions concerning the relationship between the CET and learning. Nevertheless, contrary to her attitudes

toward the impact of the CET on learning, she takes a critical view of its impact on teaching. During one post-observation interview, when asked whether she had thought of organizing some communicative activities, without any hesitation, she attributed the absence of such activities in her class to the negative influence of the CET. Her rationale is as follows.

因为不能取消四级考试，所以课上不可能组织交际活动。没有考试的话，也许会让学生多说，但由于考试的存在，我们只能应对考试。你知道，交际练习是很耗时的。这学期学生要参加四级考试，因此课时很有限，我们根本没时间去交际。

[Translation: It is impossible for me to carry out communicative activities in class, for the simple reason that we cannot get rid of the CET. Without the test, I will probably increase the student-talk time. Whereas, with the existence of the test, we can do nothing but teach to the test. You know what, communicative activities are too time-consuming. Due to time constraint, we can hardly find time to organize them especially in this semester during which the students are about to sit the test.]

As is seen in the above comments, her views concerning the “impact of the CET” have little to do with the revised CET. However, judging from T3’s words, the CET should take the blame for the lack of communicative intent in her teaching. She further justified her point by explaining:

即使是我们领导，美国来的教授都妥协了。起初他总敦促我们在课上加强口语训练，但这引起学生的很多抱怨。许多孩子宁愿我们利用这些时间领着他们做题。

[Translation: Even the head of our program, a professor from the U.S. compromised. Originally he urged us to devote more time to the practice in oral English, but this evoked a lot of complaints from our students. They would rather that we used that period of time for leading them to do a set of tests.]

When asked whether the CET could change her way of teaching, she firmly responded:

不可能。我不明白为什么听力的比重增加了这么多，似乎没有过渡。

[Translation: It cannot. I don’t know why there has been such a big increase in listening. It seems that there is no transition.]

The above comments made by T3 seemed to suggest that she was not mentally prepared for the change. It is clear from her statements that the CET is incapable of improving the way she teaches, which overlaps her response in the survey (see Table 14). The effect of the test, based on her response in the survey, only lies in that it can help her shift her instructional focus from linguistic knowledge to language use.

Table 14 *Washback Effects of the Revised CET Based on T3 and T4's Accounts*

(Y=Yes, N=No)

Participant	Improve the way you teach?	Shift instructional focus from linguistic knowledge to language use?	Give direction as to what aspects of language needed to be taught?
T3	N	Y	N
T4	Y	N	N

T4, however, saw the effect of the revised CET on teaching and learning in a more positive light. In contrast to the perceptions of her colleagues, T4 explained the impact of the CET in a unique way. At one post-observation interview, she said,

我觉得新题型无论对老师还是学生，利大于弊。过去我会花很多时间讲词汇，但是现在就不同了，更多的时间放在语言技能的培训上，比如：阅读和听力技能。考试的改革也影响到了我们的学生。过去，他们只看词汇方面的书，而现在，他们肯花更多的时间做技能练习了。

[Translation: I think that the revised CET exerts more beneficial effects than negative ones on both teachers and students. I used to spend a lot of time teaching vocabulary. But now I feel that I no longer need to do so. I realize that more of my class time can be devoted to drilling students' language skills, for instance their skills in reading and listening. The test innovation has affected our students as well. In the past, they simply read vocabulary books. Now they have to spend more time doing skill-oriented exercises.]

In contrast to T3, her response in the survey (see Table 14) revealed that the effect of the revised CET resides in the fact that the test can help her improve the way she teaches. But it is interesting to note that despite the positive beliefs she articulated about the revised CET and its washback effects on several interview occasions, she responded on the survey that the test cannot cause her to shift her instructional focus and nor can it give her direction as to what to teach. Obviously, there is incongruence between her statements on the two different types of research occasions.

#### 5.4.4.2.2 *Teachers' curriculum knowledge.*

Unlike the teachers at S1 who claimed to know little about the CET, the interviewees at S2 demonstrated a good knowledge of the revised test and the CECR in terms of their objectives and the main content. Through the interview questions, I learned that both participants demonstrated a clear understanding of them.

#### 5.4.4.2.3 *Observational evidence of the CET impact.*

Classroom observations revealed that the revised CET has affected the instructional behaviors of the two participants to a limited extent. Despite the positive view T4 expressed about the new test during the interview sessions, little evidence was found showing that her classroom behaviors were closely test-related. When comparing the ways she delivered her lessons in the first two rounds of observations, I discovered that the only change that can be attributed to the impact of the revised CET is that she has noticeably increased the amount (from 23.15% to 34%) of her class time devoted to listening (See Table 15).

In contrast to other participants, T3's lessons were more test-oriented. Not only was she observed constantly referring to the CET (both old version and new) by telling students about question types and language points that might appear in the test and giving them tips on test-taking strategies, but she was also often seen organizing activities geared toward the two versions of the CET. During my observations, this focus on test-related activities became more and more apparent. As shown in Table 15, there was an increase in the amount of her class time devoted to listening (14.17 to 20.30%). Meanwhile, the percentages (16.46, 11, 14.68) shown in Table 16 suggested that T3 made strenuous efforts to facilitate students' fast-reading skills. Although these numbers can hardly be taken to infer that they are changes induced by the revised CET, her explanation at one post-observation interview serves to convince us that they are, for she explicitly said that she led her students to do fast reading exercises simply because of the revised CET.

At S2, both participants were seen making full use of the available teaching instruments and visual aids.

Table 15 *Percentage of the class time devoted to aural aspects of language by T3 and T4 (%)*

Participants	R 1	R 2	R 3
T3	14.17	20.30	0
T4	23.15	34	0

Table 16 *Percentage of the class time devoted to fast reading practice by T3 and T4 (%)*

Participants	R 1	R 2	R 3
T3	16.46	11	14.68
T4	12.50	0	0



Like those at S1, the participants at S2 spent three weeks leading students to do simulated test. By going over the past exam papers, they aimed to enhance students' testing strategies.

#### 5.4.4.2.4 *Teachers' beliefs about teaching and classroom scenarios of how they teach.*

### T3

T3's lessons were also characterized as being knowledge-oriented and teacher-dominated. Reading is the primary skill emphasized by her. The focus of her instruction was predominantly on language knowledge. Teacher talk took up 70-80% of her class time. What struck me was that as part of the class routine, she invariably started each new lesson leading her students to go over the vocabulary lists provided at the end of each text before giving her lecture on the text. Apart from activities such as reading texts aloud and translation, rarely was she observed interacting with her students for the purposes of communication. Throughout the observation process, pair-work or group-work was never observed in her class. Furthermore, she seldom produced extended sentences in English. Much of the classroom instruction was carried out in Chinese.

Despite the test-related activities she organized, the method she used was stereotyped. For example, when she led her students to do fast-reading exercises, first she simply asked the students to quietly read a passage she passed out to them and then answer the given questions. After that she checked the answers with the whole class and provided them with the keys by highlighting the essential vocabulary and explaining why each choice was made. Between whiles, she gave the students tips on how to deal with the similar types of questions in case they appear on the CET.

When I asked T3 to comment retrospectively on the teaching strategies she utilized in her instruction, she articulated some interesting beliefs about why she taught vocabulary this way. Here is an excerpt from her remarks at the interview.

我没怎么考虑过教法这类的问题，只是按自己的方法去教。我个人认为首先要重视学生的阅读能力，我觉得学生阅读中的最大障碍是词汇量小，因此他们看不懂文章，写作时也表达不出个人想法。所以，我们要重视培养学生的语言知识。

[Translation: Seldom did I think of such issues as teaching methods. I simply taught using my own way of teaching. Personally I think that teaching priority should be given to developing students' abilities in reading, because I find the biggest barrier the students encountered while reading is that their vocabulary is

limited. Consequently, they had difficulty understanding the passages they read, and furthermore they could hardly convey their ideas when writing compositions. Therefore, linguistic knowledge should still be stressed.]

She further justified her practice by saying:

学生习惯了结构分析法, 而且不容易改变他们以往的学习方法。每篇课文的生词很多。如果我们不给他们解释的话, 他们不知道如何使用它们。

[Translation: The students are used to the method of structural analysis, and they find it hard to change their traditional way of learning. Each text consists of a large number of new words. If we do not explain them, the students do not know how to use them.]

She also defended her use of Chinese in class saying,

用英语讲课是浪费时间, 学生可能不习惯老师完全用英语上课。

[Translation: Using English is a waste of time. The students may not be used to being taught purely in English in class.]

#### T4

Unlike other participants who attached more importance to language forms, T4 stressed the development of students' ability to use English. She was so highly motivated that she spontaneously experimented with communicative activities as well as cooperative learning activities (e.g., pair work/ group work, language games, questions and answers) in her classes. Not only was she observed frequently utilizing authentic materials, but she was also found using textbooks more creatively and trying hard to encourage her students to interact in class. It was noted during the observations that her students showed higher motivation in learning English and were more active in class than those of other classes observed.

When recounting the reasons why she implemented these interactive activities rather than spending a lot of time on language forms, she articulated:

我尽可能地试着组织课堂活动, 主要是怕学生烦。我难以容忍就是学生上课低着头不搭理我。我不认同“填鸭式”教学, 如果老师连着讲两小时, 学生顶多获取 10% 的内容。在我看来, 尽管学生愿意, 单独教授词汇的话, 弊大于利。词汇解释得越多, 学生越糊涂。在没有上下文背景的情况下学词汇, 不可能真正地掌握它们。

[Translation: I tried organizing as many activities as possible, because I was afraid that my students would be bored with my lessons. What I cannot stand is that they all lower their heads and do not respond to me. I do not think the “Cramming-Duck” method works. I believe that if a teacher lectures for two hours, a student will only end up acquiring 10% of what he or she is taught.

In my opinion, teaching vocabulary as discrete points does more harm than good to the students, even though they prefer to be taught that way. The more vocabulary we explain, the more confused the students will become. It is impossible for students to have a command of it by learning it in such a decontextualized manner.]

The above comments provided insight into her professional stances on CE instruction. At one interview, she indicated that she was aware that most of her colleagues continued to devote plenty of their class time to teaching language forms, even if the section of vocabulary and structure as a test item has been removed from the revised CET. She proceeded to pinpoint the reasons why teachers at large have a preference for teaching vocabulary. The following is an excerpt from her interview discussion.

其中的原因很多。一个原因就是自己的老师就是这样做的；另一个原因就是个人水平问题。说得更具体些，他们没法用英语表达自己的想法。在这种情况下，讲词汇是最容易的方法，他们就不必去提高他们的语言水平了，也不必动脑筋想怎么教的问题了；我认为还有一个原因，那就是，这样做的话学生喜欢。

[Translation: There are a number of reasons for this. One is that they may have been taught that way. Another plausible reason is that they are constrained by their own language proficiency. To be specific, they have trouble expressing their ideas in English themselves. In such a case, teaching vocabulary is the easiest way they can do. By so doing, they no longer need to take the trouble to improve their own language abilities. Nor do they need to rack their brains on how to teach. Another reason might be that that way of teaching, in their opinion, appeals to their students.]

According to her, the majority of teachers have never thought of whether it is appropriate to teach vocabulary that way. They simply follow what other teachers do.

She added:

还有可能，有些老师不负责任，或缺少事业心。

[Translation: It is also possible that some teachers are not responsible or conscientious enough.]

In spite of her efforts, her class still seems deficient in that she was seldom seen calling on students to answer her questions on a one-to-one basis. She was also aware that she had encountered some obstacles while carrying out student-centered activities. At the second post-observation interview, she listed some of them, as we see in the following excerpt.

主要障碍是四级考试。其它还有：学生水平、时间不够用。如果时间充裕的话，我会组织更多的活动，比如：小组讨论、信息差，也会顾及听说读写等能力的培养。但问题是时间总不够用，而且，尽管对交际法教学理论略知一二，但那点儿知识远不够用。当我上英语课程班时，老师当初解释的也不太清楚。我和班里的同学都没完全掌握这些理论，更不知道在实际教学中怎么用这些理论。

[Translation: The main obstacle is the CET. Other obstacles include the proficiency level of the students as well as the time constraint. Had I had more time, I would have conducted more activities such as group discussion and information gap, and incorporated skills such as reading and writing as well as listening and speaking. But the problem is that I do not have adequate time. In addition, although I have some basic knowledge of the theories about CLT, etc., that knowledge is far from sufficient. When I was taking courses in an ESL graduate program, my teacher did not explain them clearly to us. Neither I nor my classmates have achieved a thorough understanding of them. Nor have we acquired hands-on knowledge of the theories.]

What we should note is that in the above statement, she listed the CET as the main obstacle to her teaching, which seems to be conflicting with the other statements she made concerning the revised CET.

With respect to the impact of learner beliefs on teaching, her opinions are consistent with those of others. One example she gave us corroborated T1 and T2's assertion that students' beliefs also have a part to play in the way that teachers teach.

有一次，我收到学生的一封信。信上讲他们很喜欢我的课。可在对我讲课的方式表示满意的同时，他们也表达了他们的担忧，担忧要准备我留给他们的小组讨论的作业，他们还提醒我：快要考试了。

[Translation: Once I received a letter from my students saying that they enjoyed my class very much. While they assured me that they were contented with the way I delivered my lessons, they expressed their worry about the group discussions that I had assigned them to prepare, for they reminded me that their exam was round the corner.]

This example serves to illustrate that students' beliefs also play an important part in teachers' decisions as to how to teach.

#### 5.4.4.3 *How Teachers Perceive and React to the CET Reform at School Three (S3)*

##### 5.4.4.3.1 *Teachers' articulated beliefs about the CET and its impact.*

The findings discussed above reflect the fact that at S1 and S2, there were more negative than positive voices heard about the CET while the group interviews were going on. At S3, a balance seems to have been achieved between the two. While most teachers at S3 acknowledged the effect of the revised CET on CE teaching and

learning, their views were divided as to whether it is a negative or a positive one.

On the one hand, some of the teachers were both critical and suspicious about the reform, claiming that it has to do with the authorities' personal beliefs. Below is one complaint made by an interviewee about the CET and curriculum reforms.

有关考试和大纲的改革不是建立在科学研究的基础上，而是根据有关领导的个人经验和个人观点，多少有点长官意识，要是换了其它领导，施行的有可能是另一套改革。  
[Translation: Rather than being based on scientific research, these reforms in the CET and curriculum simply represent a bureaucratic attempt, which is associated with the authorities' personal views or personal experience. Had it not been the current leader but someone else in charge, the reform would have taken a different appearance.]

The above complaint was also echoed by other participating teachers at the group interview. Another interviewee seconded the above view and said,

领导有时心血来潮，开展改革是出于政绩的考虑。  
[Translation: Authorities sometimes simply act on impulse. They carry out reforms for the sake of their political achievements.]

Like the teachers at S1, these teachers did not think that there are fundamental changes in the CECR and the CET. As a result, they did not believe that the reforms could bring about qualitative changes in teaching and learning. One teacher argued:

新题型和大纲的变化并不是革命性的变化，所以这种变革没给教学带来实质性变化，我们依然按以前的教学模式去教。  
[Translation: Without revolutionary changes in the CET and curriculum, few changes can be traced in our teaching. We would follow our traditional mode of teaching.]

It is clear that these teachers did not believe that the CET could foster changes in teaching and learning. They thought their teaching is more closely related to the textbook than to the CET.

On the other hand, other teachers held positive opinions about the CET reform and its impact. They maintained that the reform is indispensable and the revised CET can more or less shift their focus of teaching as well as greatly enhance the students' integrative skills in English. The following excerpt from one interviewee illustrates such an opinion.

目前的改革不可能一步解决大学英语教学的所有问题，但是它是社会发展的趋势，算是一种大的进步了。早在 80 年代，改革开放初期，那时搞口语训练没什么意义。然而，

随着国家发生翻天覆地的变化，强化听说训练，似乎势在必行。

[Translation: It is impossible to ensure that the current reform will tackle all problems in CE teaching and learning. But it is a big progress and it corresponds to the development and needs of the society. In the early 1980's when China first opened up to the outside world, it did not matter much whether or not we placed great emphasis on students' spoken English. However, today, with so many fundamental changes that have taken place in the country, an emphasis on aural/oral aspects of English seems imperative.]

The above statement was buttressed by several other interviewees.

It was interesting to note that T5 and T6 also held divergent views over the revised CET and its impact. Like T4, T5 maintained a positive view toward them. At the individual interviews, she indicated that the revised CET had motivated her to reflect on her instructional practices and she had occasionally experimented with new methods of teaching. In addition to its positive effect on teaching, she believed that the revised CET could exert a beneficial impact on learning, in the sense that it would allow her to guide her students to do more listening practice after class.

In comparison, the perceptions T6 articulated about the CET at the group interview not only sound negative, but also seem radical, for he argued that it was meaningless to carry out the current testing reform. When asked to make comments on the revised CET, he said:

考试既浪费时间，又浪费钱。就因为这考试，英语学习的目的都变了，学生现在学英语的目的就是通过考试。改也好，不改也好，或者不管它与教学是否一致，它的负面影响都很大。因为学生们转变不了他们根深蒂固的观念。

[Translation: Having this test is a waste of time and money. Owing to the test, the purpose of learning English has changed. Now passing the CET has become the sole purpose of learning the language. Whether the CET has changes or not or whether it conforms to teaching or not, its negative impact is still profound. The reason is that the students' deeply-seated beliefs can hardly be changed.]

According to him, the CET was designed for the students of key schools rather than those of average schools. Therefore, it is not suitable for his students. Further, he doubted that the CET scores could reflect the students' skills of language use.

At the group interview, he also raised the following question,

当学生连一篇课文都读不好的时候，为什么还让他们去考四级？

[Translation: When our students cannot even read a text well, why do we require them to take the CET?]

In his opinion, in no way should CE teaching and learning be connected to the CET.

Initially, what he articulated struck me that the test is contradictory to his teaching philosophies. However, what he said at the post-observation interviews helped me understand what he meant. In his opinion, the CET is not suitable for students of average schools.

考试会影响学生，自然也会阻碍教学。

[Translation: The CET poses a constraint on the students and thus it automatically hinders teaching.]

It must be noted that although he claimed that the test hinders both teaching and learning, he argued that the constraint it poses on learning is greater than on teaching. He added that it was his students who deemed the CET and teaching and learning to be in conflict. What he said seemed to imply that the ‘learner factor’ seems to be involved in this issue of test impact.

T6’s complaint about his students was echoed by his colleagues who also blamed the students for being so utilitarian and clinging to their deeply-rooted views.

Nevertheless, as with T3, what he charged the CET with has little relevance to the test itself.

It should also be noted that despite the negative remarks T6 made about the CET at the group interview, his tone, though still negative, sounds much less convinced at the two individual interviews. At the second post-observation interview, when commenting on the revised CET, he was quoted as saying:

新题型的优点在于它强调语言的使用，最主要的是听力的比重增加了。作为输入的渠道，听力很重要。它是最直接的交际方式。

[Translation: The revised version seems to be superior to the old one in that it attaches importance to language use and above all, the weighting of listening is increased. As a means of input, listening is crucial. It is the most direct way of communication.]

It is clear from his comments that this time he turned to look at the positive side of the test. Another interesting finding is that while T6 expressed negative feelings about the CET at the interviews, his responses in the questionnaire reflected a different picture. As evidenced in Table 17, when asked whether the CET could produce washback effects on teaching (e.g., whether it would help to improve the way he taught, shift his instructional focus from linguistic knowledge to language use, and give direction as to what aspects of language needed to be taught ), all his answers

were positive. We should note, therefore, that there is a discrepancy between his responses at the interviews and those in the survey.

Table 17 *Washback Effects of the Revised CET Based on T6's Accounts*  
(Y=Yes, N=No)

Participants	Improve the way they teach	Shift instructional focus from linguistic knowledge to language use	Give direction as to what aspects of language needed to be taught
T6	Y	Y	Y

#### 5.4.4.3.2 Teachers' curriculum knowledge.

Like the teachers at S2, the interviewees at S3 also demonstrated a good understanding of the revised CET as well as the background in which the reform was introduced.

#### 5.4.4.3.3 Observational evidence of the CET impact.

By observation, both participants spent most of their class time conducting activities based on the content of their textbook. However, both of them have increased, though at varying degrees, the amount of their class time spent on listening. With respect to T5, an obvious increase is shown between the percentages of her class time devoted to listening in R2 and R3 (30.6, 28.14) and that in R1 (7.04). For T6, the percentage of his class time spent on listening in R3 (13.26) is also relatively higher than that in R1 and R2 (6.09, 4.93). See Table 18.

Table 18 *Percentage of the class time devoted to aural aspects of language by T5 and T6 (%)*

Participants	Round 1	Round 2	Round 3
T5	7.04	30.62	28.14
T6	6.09	4.93	13.26

Not only did the two teachers give more attention to the development of students' skills in listening in class, but they were also observed urging their students to do listening practice after class in groups.

Apart from the above change, however, little evidence was captured that can be labeled as washback evidence. What distinguished T5 and T6's classes from those of other observed teachers was that no past test papers were used throughout the whole



observation period.

5.4.4.3.4 *Teachers' beliefs about teaching and classroom scenarios of how they teach.*

#### **T5**

The observation data showed that the way T5 dealt with her lessons exhibits features of both traditional methods of teaching and CLT. On some occasions (R.2.1, R.2.2, R3.2), she was observed using her textbook creatively by going beyond it to create local contexts for her students to use the language. At such times, she was carrying out activities to practice students' skill in speaking. On other occasions (R1.1, R1.2, R3.1), she was found using the textbook in a formal way, dealing with it as a means of reinforcing language forms such as vocabulary and grammatical points. During these times, a lot of translations and paraphrases were utilized. Furthermore, she was often seen using power-point slides to teach vocabulary and present background knowledge. When asked to expound on her reasons for using methods of translation and paraphrases, she said,

我经常把说的话翻译成汉语，是为了强调所教的内容。这样做有助于学生的理解。至于复述，很难说这种方法好与不好。很多时候，老师是把所复述的句子提前写在书上。一般，这些句子可从教师用书上抄。上课时读这些句子就行。

[Translation: I often put what I said from English into Chinese to highlight what was taught. In this way, I can clarify what the students may not have understood. With reference to paraphrasing, It is hard to say whether it is good or not. In many cases, teachers have prewritten the paraphrased sentences on their own textbooks. Generally the sentences are copied from teachers' books. In class, they simply need to read them.]

During one individual interview, she reflected retrospectively on her own way of teaching, saying regretfully,

尽管我喜欢交际法，也尝试着开展意义型教学。但是坦率地说，我花在备课上的时间太有限了。如果在教学上投入的时间和精力更多的话，我的课上的就不是这样了。

[Translation: Although I prefer to use CLT and attempt to conduct meaning-based instruction, the time I dedicated to preparing my lessons was, to be frank, quite limited. Had I committed more time and energy to my teaching, my lessons would have taken a dramatic turn.]

During the observations, she was also occasionally found devoting a lot of time to explanations of language points. She defended why she was doing so, saying:

语言是交际的手段。当学生说的时候，体现的是他们的语言知识，不然的话，他们就不知道谈些什么了。依我看，使用语言的条件就是要有真正的内容，内容是交际的载体。

[Translation: Language is a means of communication. When students talk, they need to demonstrate a good command of linguistic knowledge. Otherwise they will be at a loss what to talk about. In my view, language use should take place under the condition that there is some real content. Content is the carrier of communication.]

## T6

T6's teaching patterns could be characterized as combining aspects of the traditional method (e.g., with a focus on basic skills such as pronunciation and recitation) and CLT (e.g., engaging students in discussions and negotiation of meaning). During the observations, he was often seen asking students to read texts aloud and in the meantime modeling correct pronunciation. In addition, during R2.1 and R 2.2, his students were observed spending a lot of time reciting, both individually and in chorus, the passages he had assigned them to prepare. While he placed special emphasis on pronunciation, the time he spent on teaching vocabulary and grammar was much less than that other observed teachers did. Compared to other observed teachers, he used more English in class. Similar to the students of T4's class, his students were all well-disciplined and highly cooperative in the classroom, which might be related to the high expectations he set for them.

A look at the interview data demonstrated that his practices in the classroom reflected his teaching philosophies. The following comments reflected some of his teaching stances.

掌握一门语言就是为了使用它。因此，我们的教学应着眼于帮助学生掌握使用英语的技能。最重要的是帮他们打好基本功，比如：发音，恰当地用英语表达。培养这些技能，背诵是关键。不背诵一些规范的课文和文章，就不可能自如地表达。正是因为这个原因，我给学生留了42个题目，每次课每位学生要背3段与这些题目相关的文章。学生很配合，因为我把这样做的目的告诉了他们。跟学生沟通很重要，要让他们知道怎么去学。

[Translation: To acquire a language is to use it. So our teaching should aim at helping our students to acquire the competence to use English. Foremost, we must help them lay a solid foundation in basic skills such as pronunciation and talking in appropriate English. To develop these skills, memorization is pivotal. Without memorizing some standardized texts or articles, it will be impossible for them to express their opinions freely. That is why I assigned my students 42 topics and each class each student is supposed to be able to recite 3 paragraphs in relation to these topics.]

My students are cooperative, because I let them know my purpose of doing so. It is important to communicate with students and let them know how to learn.]

It can be seen that based on his notion, the ability of language use can be acquired through practice of basic skills such as pronunciation and rote memory.

An interesting finding is that like T4, T6 was negative about devoting a lot of time to teaching vocabulary, saying that it was quite time-consuming and impeded the enhancement of students' language competence. According to him, even if the students were taught a lot of vocabulary, they could hardly remember it. Like T4, he made an interesting comment on why many teachers prefer to teach vocabulary.

相当多的学生认为教授词汇的老师水平高、学问大，但事实上，教词汇是容易的方法了。

[Translation: Quite a number of students believe that the teachers who lecture on vocabulary are both high-leveled and knowledgeable. But the fact is that teaching vocabulary is the easiest approach.]

He also explained why he spoke English all the time in his classroom. His assumption is that because he teaches in English, his students would be able to be exposed to the language as much as possible. At one individual interview, he revisited the issue of teacher quality touched upon by his colleague at the group interview. Some interesting points he made on the issue are as follows.

大学英语教师课堂上对学生的口语输入也很重要，其作用甚至大于听录音、看电视，因为老师是更直接地与学生交流。如果老师的口语能力大大提高了，那么大学英语的课堂教学水平会提高一大截。然而这次改革不可能实现这一目标。

[Translation: CE teachers' input in the classroom plays a crucial role in students' exposure to the language. It's more important than the exposure they receive when listening to the recordings or watching TV, since the interactions between a teacher and his/her students are more direct. If the teachers' oral proficiency has improved, then CE teaching will be upgraded to a higher level. However, the reforms being implemented this time cannot attain such a goal.]

#### 5.4.5 Overview of the Washback Evidence from Case Studies

##### 5.4.5.1 Summary of the Results from Case Studies

The data presented above give us an idea of the participants' beliefs and some scenarios of their teaching practices in the classroom. On the whole, all the participants were interactive and cooperative. They all impressed me as committed and responsible CE teachers, although their conceptions of teaching, their levels of

language proficiency (e.g., competence in terms of four skills, awareness of the sociocultural aspects of language and language use as well as knowledge of pedagogy), the ways in which they conducted their lessons, and their devotion to work differ to varying degrees.

The findings reveal that due to school differences as well as differences among teachers and students, not only did the ways teachers perceive and react to the CET and its washback vary from school to school, but they also differ from individual to individual. On the one hand, teachers' beliefs and knowledge of the CET and the CECR were found to vary from context to context. As reported above, the opinions the teachers of the key school (S1) held about the CET and the CECR were different from those held by teachers of average universities (S2 and S3). Compared to teachers of average schools, the key school teachers were seemingly less concerned about or somewhat indifferent to the reforms in the CET and the curriculum. As evidenced in the excerpts from their chair's accounts, they were resistant to the changes newly introduced to CE education, for they overtly claimed having grown tired of these reforms. When talking about the effects of the CET on their teaching, the majority of them insisted that they would not be affected by the test. Unexpectedly, when it came to the last stage of the study, the teachers continued to claim that they had not read the CECR and the sample revised CET, which could be interpreted that their curriculum knowledge was indeed limited or insufficient.

In contrast, teachers of average schools not only exhibited a better knowledge of the CECR and the CET, but were also aware of the impact of the test on CE teaching and learning. One significant feature that has emerged from the data is that the pilot school teachers seemed to be more nervous about the revised CET than those of the non-pilot school, for quite a number of teachers there reported at the group interview that they were pressured by the CET change and were in a dilemma as to what to do.

On the other hand, the case-study results reflected that the participants, guided by their personal beliefs, were split in their perceptions of the CET, its impact, and the CECR. Worthy of note is that only two of the six participants (T4 and T5) saw the CET in a positive light. While T3 expressed negative feelings toward the CET and its impact, her feelings seemed to be mixed. Although she claimed that the test affected her teaching negatively, she asserted that it had a beneficial impact on learning in that it motivated her students to learn. Interestingly, T6's negative attitudes toward the CET differ from those of T3, for he assumed that the CET constrained learning more

than it did teaching.

#### 5.4.5.2 *Observable Washback Evidence*

While the classroom observations were conducted sequentially at selected times, they were not done continuously. Thus, it is hard to guarantee that they can capture a comprehensive picture of the teaching behaviors in the classroom. However, the data gathered are still representative in the sense that they record and reflect typical events and behaviors of the classroom.

Regardless of teachers' articulated beliefs about the CET and its impact, there is only limited observable evidence showing that their instruction is related to the revised CET. Below is a brief summary of the results based on the hypothesis described at the beginning of Chapter 4.

#### *Hypotheses of the CET Impact*

Provided a relationship exists between the test and what the teachers practice, 1) there will be an increase in the amount of time spent on activities such as listening and a decrease in the amount of time spent on reading; 2) there will be activities aimed at enhancing students' skills in fast reading (or skimming and scanning); 3) last but not least, the amount of time devoted to language forms will be reduced.

#### *Results Relevant to the Hypothesis*

##### *1) Increase in aural/audio aspects of teaching?*

Although the participants increased the amount of their class time spent on listening (irrespective of R3 for S1 and S2) over the duration of six months, the ratio between listening and reading practice was still disproportional. It has been explained earlier that in R3, both S1 and S2 replaced textbooks by the past exam papers. Thus, the data collected then has little value for reference.

Table 19 and Figure 9 show the percentage of their class time devoted to aural aspects of language in each round of observation.

Table 19 *Percentage of the class time devoted to aural aspects of language by round (%)*

Participants	Round 1	Round 2	Round 3
T1	0	12.90	0
T2	2.78	9.04	0
T3	14.17	20.30	0
T4	23.15	34	0
T5	7.04	30.62	28.14
T6	6.09	4.93	13.26

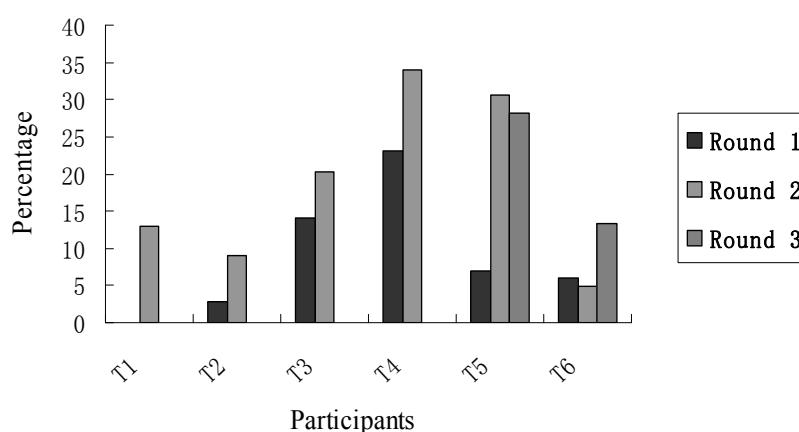


Figure 9. Bar chart of time devoted to aural aspects of language across the observation period (Round 1 – 3).

### 2) Perform fast-reading exercises?

As mentioned earlier, only one pilot class participant (T3), as shown in Table 20, performed fast reading exercises on a regular basis in class. Other participants seldom carried out such exercises.

Table 20 *Percentage of the class time devoted to fast reading practice by round (%)*

Participants	Round 1	Round 2	Round 3
T1	0	0	0
T2	0	0	0
T3	16.46	11	14.68
T4	12.50	0	0
T5	0	0	0
T6	0	0	0

### 3) Decrease in reading-related activities?

From observation, it is found that the impact from old versions of the CET and

curriculum (especially the 1985 version) continued to exist in both pilot and non-pilot classes. As is seen in Table 21 and Figure 10 below, nearly all the participants continued to emphasize the cultivation of their students' reading skills. In any round of observation, their class hours devoted to reading skills exceeded 50%.

Table 21 *Percentage of the class time devoted to reading practice by round (%)*

Participants	Round 1	Round 2	Round 3
T1	100	74.2	100
T2	80.04	78.16	96.26
T3	69.21	63.66	85.32
T4	50.74	66	99
T5	92.96	51.78	69.51
T6	67.76	75.35	60.18

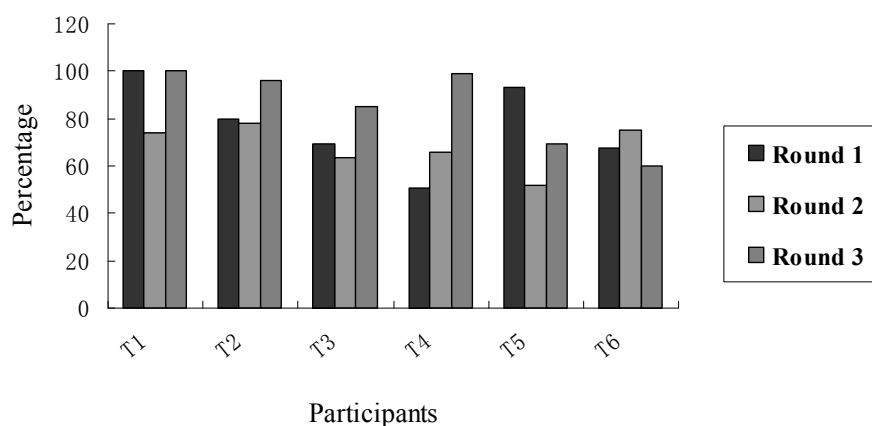


Figure 10. Bar chart of time devoted to reading practice across the observation period (Round 1 – 3).

#### 4) Instruction – form-based or meaning-based?

Except T5 whose class steadily reduced the amount of time devoted to language knowledge between R1 and R3 (from 43.79% to 21.84%), other participants continued to spend most of their class time either teaching grammatical and lexical items (e.g., in the classes of T1, T2, and T3) or drilling students' pronunciation (e.g., in T6's class). Although the following percentages do not indicate a decrease in the amount of time T4 spent on language knowledge, in comparison to other participants, she gave less weight to it. Table 22 and Figure 11 illustrate the percentage of mechanical manipulation of language forms (e.g., reference to vocabulary, grammar, and pronunciation) in each round of observation.

Table 22 *Percentage of the class time devoted to language forms by round (%)*

Participants	Round 1	Round 2	Round 3
T1	47.62	48	40
T2	43.37	55.56	51.84
T3	50.51	42.59	56.41
T4	3.62	17.26	28.88
T5	43.79	34.72	21.84
T6	64.62	18.31	64.78

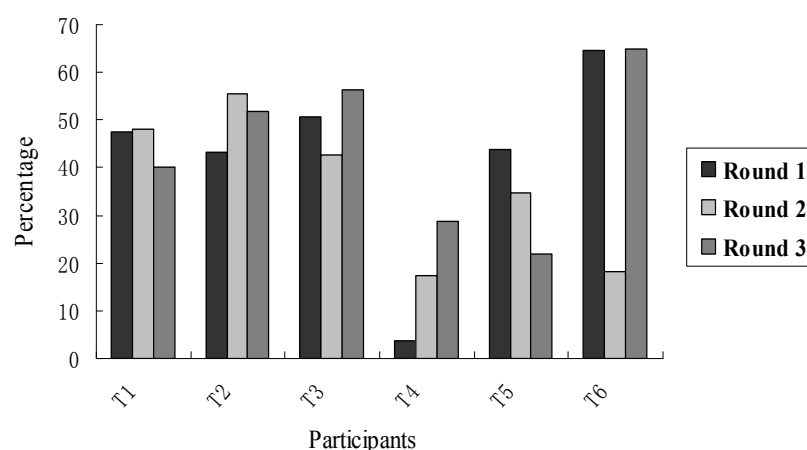


Figure 11. Bar chart of time devoted to language forms across the observation period (Round 1 – 3).

In summary, a comparison between the way the case-study participants teach (i.e., the time they spent on teaching aural aspects of language, fast reading practice, reading, and language forms) and the task characteristics of the revised CET illustrated that the CET has more or less had an impact on some aspects of the participants' teaching behavior. However, there are still big gaps between the two with respect to the degree, intensity and scope of the test impact on the activities teachers organize, the focus their instruction is placed, and the teaching approach(es) they employ. Despite the participants' shift of focus from an emphasis on reading to one on listening, such a shift is inadequate considering the disproportional ratio between the time devoted to this skill and that of reading. As you can see from Table 23 and Figure 12 below, the average percentage of class time spent on reading by each participant is still much higher than that spent on listening. Moreover, limited evidence was found showing the participants' efforts to facilitate language use in the classroom as encouraged by the revised CET and the CECR. Rather, most of the participants (except T4) continued to spend a large amount of time on the formal aspects of



language (see Table 23 and Figure 12).

Table 23 *Average Percentage of the Class Time Respectively Devoted to Language forms, Aural Aspects of Language, Reading Practice (%)*

Participants	Time on language form	Time on Reading	Time on Listening
T1	45.21	87.1	6.46
T2	50.26	79.1	5.91
T3	49.84	66.44	17.24
T4	16.59	58.37	28.58
T5	33.45	72.37	18.83
T6	49.24	71.56	5.51

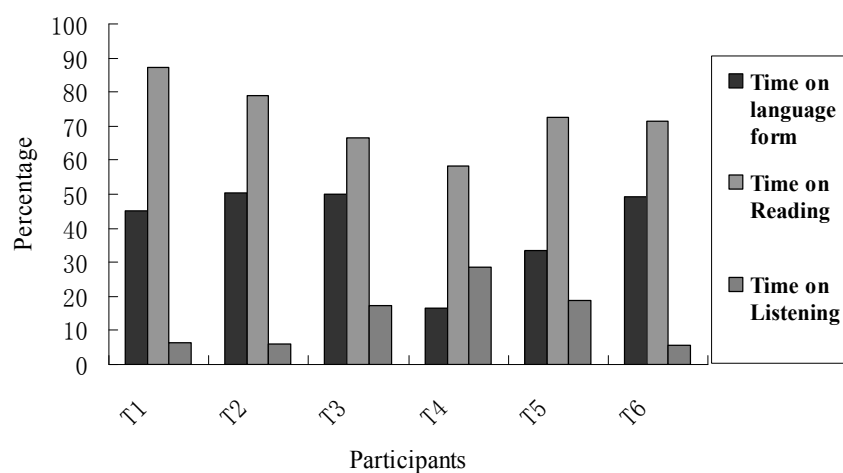


Figure 12. Time devoted to language form, aural aspects of language, reading practice.

In view of the limited amount of test-like activities observed in the first two rounds of observations, a conclusion can be drawn that a link cannot be established between what the CET encourages and what a teacher does.

The interview and observation data also indicated that in general, the teachers did not rely on the CET and the CECR as guidance for their instructional practice. Rather, they taught simply based on their own philosophies or conceptions of teaching. With respect to the conceptual issue, a more intensive discussion will be conducted in the next chapter. In contrast to the hypothesized washback effects of the CET reflected in Figure 4, the observed washback effects of the CET on teacher practice is illustrated in Figure 13.

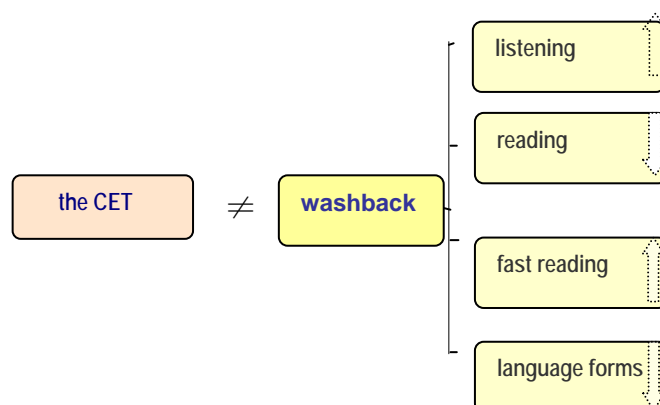


Figure 13. Observed washback effect of the CET on teacher classroom practice.

Overall, the data set presented in this section is qualitative. The next section presents the quantitative data collected through a teacher survey.

As presented in 4.4 and illustrated in Figure 5, this study adopted a mixed-methods approach for data collection and data analysis. Three complementary methods (i.e., interviews, observations and questionnaires) were utilized, with the aim of getting a deeper and more comprehensive understanding of how the role of the ‘teacher factor’ operates in the washback phenomenon.

The different types of data, qualitative and quantitative, were intended to serve different purposes. The qualitative data collected through group/in-depth interviews and focused classroom observations allowed for a more intensive and in-depth examination of how individual teachers perceive and react to the CET innovation with respect to their specific teaching contexts. The value of this set of data is twofold. On the one hand, they enhanced our understanding of the participating teachers’ interpretations of test impact, their conceptions of teaching, and the meanings they attached to their classroom practices. On the other hand, they provided a clear picture of how teachers’ thoughts and conceptions were related to their activities and practices observed.

As was previously presented in detail (see Section 4.4 and 4.5.3.5), the qualitative data were supplemented with the survey data. The survey was used, for it was assumed to be best suited for quantifying the qualitative data and providing descriptions and comparisons of patterns of teacher beliefs and behaviors. The instrument would permit the generalizability of insights derived from the qualitative data and help me determine whether the patterns and themes that had emerged from previous stages (case studies) could be confirmed and applied to a larger group of

participants (questionnaire participants).

## 5.5 Results from the Teacher Questionnaire

### 5.5.1 *Introduction*

As introduced in Chapter 4, a survey was administered to the participating teachers of this study to poll their beliefs about the revised CET and its impact, their views of language teaching and learning, and information about what they consider to be effective ways of teaching. Out of 255 copies of surveys distributed, 195 were returned, yielding a return rate of 76.5%.

This section presents the results from statistical analyses. It consists of two parts. Part 5.5.2 outlines the main responses given by Chinese EFL teachers to the survey questions and presents the results from descriptive analyses (i.e., frequencies, percentages, means, standard deviations), while Part 5.5.3 reports the results from inferential analyses of the ‘teacher factor’ items (i.e., EFAs, CFAs, SEM).

For purposes of reporting, the decimal numbers calculated were rounded off to the nearest whole number when I reported the percentages.

### 5.5.2 *Results from Descriptive Statistics*

This part begins by providing the background information about the participating teachers. After that, it examines the views and beliefs that teachers maintain about testing, teaching, and learning. It, then, proceeds to probe into the knowledge base that teachers have built as well as their prior professional experiences.

What needs to be mentioned here is that the questionnaire statements are reported as if they were questions. For instance, Q18 refers to Statement Number 18. Questions include both positive and negative statements. If a question or statement has negative wording, it was then reverse-coded.

#### 5.5.2.1 *Participants’ Background Information*

This part describes the schools and the teachers who participated in the study. Personal demographic data requested from respondents included age, years of teaching experience, education, work load, and past experiences.

Of the 195 respondents, female teachers constituted a larger proportion, 76% of all teachers. Fifty percent of them were under the age of 30. The teachers all held a

diploma, but there is a wide gap among teachers' educational backgrounds. Teachers with a master's degree or an Advanced /Assistant Teacher Training Certificate (ATTC) account for 53%. Two teachers (1%) held a doctorate. Only 12% of the teachers had studied English in an English-speaking country. Fifty percent of them claimed that they had received specific training in ESL.

More than half of the teachers (57%) taught over 12 class hours per week. 19% of them taught over 15 hours. Their working experience ranges from 1 year to more than 20 years, with an average of 5-9 years. Table 24 shows the educational background of the survey participants in percentages.

Table 24 *Educational Background of the Survey Participants in Percentages (%)*

Degree Held	Bachelors	TTC	MA	M ED	PhD	Other
No. of Teachers (%)	88 (45.1)	18 (9.2)	81 (41.5)	5 (2.6)	2 (1.0)	1 (0.5)

#### 5.5.2.2 *Major Aspects of the 'Teacher Factor' and the 'Teacher Practice'*

This section presents the teachers' attitudes towards the CET, the CECR, the impact of the CET on teaching and learning, as well as their beliefs about EFL teaching and learning, etc. Since the questions on the teacher survey are organized by themes, the statements discussed here are also presented by themes. In the study, I have also performed some correlation analyses on teachers' beliefs and knowledge. However, due to the scope of this paper, they will not be presented in this dissertation. Here, only five major themes are reported: 1) the teachers' beliefs about the CET and its impact, 2) the teachers' beliefs about the washback effects of the revised CET on their teaching and learning, 3) their beliefs about teaching, learning, how to teach and the ways that they teach, 4) their knowledge base, and 5) their past experiences. All the themes pertain to the two research questions that were posed by this study.

##### 5.5.2.2.1 *Various beliefs as reported by the teachers.*

##### 5.5.2.2.1.1 *Teachers' beliefs about the CET and its impact.*

The first theme touched upon in the teacher survey concerns teachers' beliefs about the CET and their assumptions about the washback effects of the CET on teaching and learning. Question 19 (Q19) asked whether the participants believed that the test scores on the CET were an appropriate indicator of a student's English ability. The

results show that 50% of the respondents believed (strongly agree or agree) that they were, while 49% of them believed (strongly disagree or disagree) that they were not. Q23 inquired about whether they believed the CET hindered their application of CLT, while Q20 probed into their views on the impact of the CET on learning (e.g., whether the test can motivate students, help students understand their own learning needs, etc.). The results reflect that although the majority of teachers (69%) were negative about the impact of the CET on teaching, a large proportion of them (65%) were positive about its impact on learning. Such results suggest that while the teachers deemed the relationship between teaching and testing to be a conflict, they assumed that the CET promotes learning.

When asked about whether the CET reflects the goals and objectives of the CECR, 53 % of them maintained that the two are not in line with each other.

Table 25 shows the frequencies and mean (M) values of teachers' responses to the above questions.

Table 25 *Descriptive Statistics 1 (n=195)* (SD=Strongly Disagree; SA= Strongly Agree; D= Disagree; A= Agree)

Question	Participants' Responses		(M)	(SD)	95% Confidence Interval for Mean	
	SD/D	SA/ A			Lower Bound	Upper Bound
19	95 (48.7%)	98 (50.3%)	2.46	.69	2.35	2.56
20	67 (34.3%)	127 (65.1%)	2.67	.66	2.57	2.77
23	135 (69.2%)	59 (30.2%)	2.23	.71	2.12	2.34
22	104 (53.3%)	90 (46.2%)	2.43	.69	2.33	2.54

As a cross-referencing question, Q49 asked the participants whether there is conflict between teaching and testing. To this question, 76% of teachers responded Yes. In summary, there is a similar pattern in the teachers' responses to the above questions. The results reflect that the participants did not believe that the test was valid and could reflect actual teaching and learning.

#### 5.5.2.2.1.2 *The washback effects of the revised CET on their teaching and learning.*

A set of Yes (Y) or No (N) questions (Q44, Q50, Q52, Q55) on the teacher survey

asked whether the teachers think that the CET has caused them to change their instructional practice in terms of time allotment, content and methods of teaching.

Responses to the above questions showed that 67% of them believed that the revised CET has caused them to increase the amount of time devoted to teaching aural/oral aspects of English and 55% of them thought that it has caused them to shift their instructional focus from linguistic knowledge to language use. Similarly, 61% of them assumed that the revised CET can give them the direction as to what aspects of language needed to be taught. Nevertheless, only 40% of them held that it can help them improve the way they teach.

Regardless of the respondents' accounts of the CET impact, their responses to the above questions seemed to indicate that the CET has induced a certain degree of positive washback on their teaching practices in terms of time allocation, teaching focus, and teaching content. However, to confirm the washback effects of the CET, these responses need to be triangulated with other data sources (e.g., interviews and observations).

#### 5.5.2.2.1.3 Teachers' beliefs about teaching, learning, and how to teach.

##### 5.5.2.2.1.3.1 Beliefs about teaching and learning.

The theme focused on in Q26, Q27 and Q28 is related to the respondents' views of English teaching and learning as well as their beliefs about their role as an English teacher in the language classroom.

Responses to Q26 and Q27 indicated that while quite a number of teachers (45%, 48%) no longer maintained a transmission-based view of language learning and teaching, a little more than half of them (54% and 52%) still viewed teaching and learning as a process of imparting and receiving knowledge.

Table 26 shows the frequencies and mean values of teachers' responses to the above questions.

Table 26 *Descriptive Statistics 2 (n=195) (D=Disagree; A=Agree)*

Question	Participants' Responses		(M)	(SD)	95% Confidence Interval for Mean	
	SD/D	SA/A			Lower Bound	Upper Bound
26	88 (45.1%)	106 (54.3%)	2.46	.78	2.34	2.57
27	93 (47.7%)	101 (51.8%)	2.41	.83	2.27	2.55

When asked what they value most when students talk (see Q41), 69% of the respondents answered that what they stress is that the learners get their messages across. Only 30% of them acknowledged that they emphasized language forms such as using correct grammar, correct vocabulary and native-like pronunciation and intonation. The inference that can be drawn from these responses is that the majority of teachers believed that they placed an emphasis on meaning rather than on form in their instruction.

#### *5.5.2.2.1.3.2 Views on how to teach.*

##### *Views on Teaching Approach*

For Q37 and Q38, concerning the respondents' views on the teaching approach they prefer, 54% of the respondents claimed that they favor CLT over the structural approach or the combined approach (a combination of CLT and structural approach). When asked what methodology they believed they used in their instruction (Q32), however, 69% of them reported that their mode of instruction was a combined one. Although 51% of teachers did not support the view that the structural approach is the best method to prepare students for the CET (Q24), there are still 49% of the participants who did.

##### *Ratio of Teacher Talk to Student Talk*

Q29 asked the teachers what they think is the ideal percentage of class time for teacher talk. As many as 43% of the teachers believed that over 40% of their class time should be devoted to teacher talk, while 33% of them thought over 30% of their class time should be spent on it. However, when asked about their real practice in the classroom, 46% of the teachers claimed that the teacher-talk time constituted more than half (51% or more) of their class time. It appears that the teachers spent more time on teacher talk than they thought they should.

The histogram below (see Figure 14) illustrates the percentage of reported teacher-talk time across all teachers and all classes.

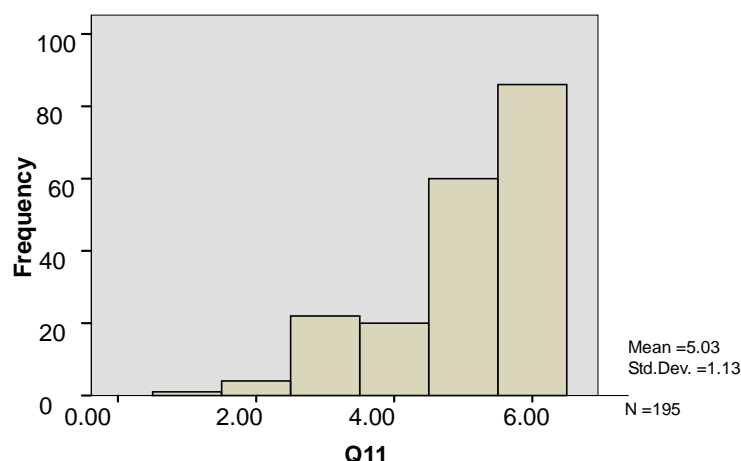


Figure 14. The percentage of teacher-talk time in the classroom.

The majority of teachers (60%) thought that the student-talk time constitutes more than 20% of their class time. Only 16% of the teachers believed that over 40% of their class time was devoted to student talk,

The histogram below (Figure 15) shows the percentage of reported student-talk time across all teachers and all classes.

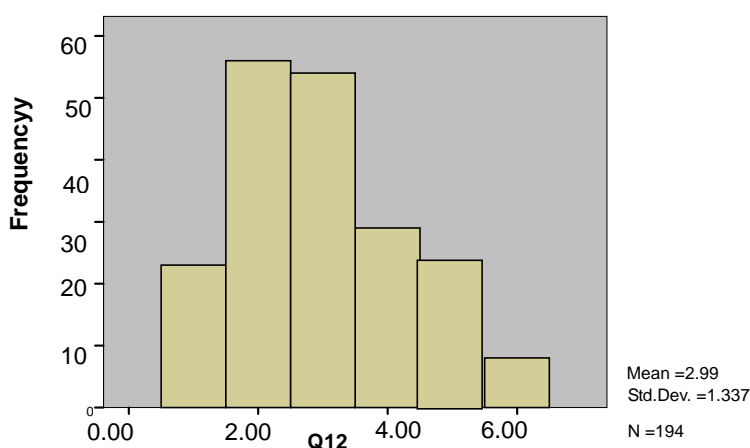


Figure 15. The percentage of student-talk time in the classroom.

The above results seem to suggest that the participants, though conceptualizing teaching and learning in different ways, have updated their views on their role as English teachers in the language classroom.

#### 5.5.2.2.1.3.3 How they think they are teaching (or: teacher practice).

Questions 9-12 consisted of a set of questions inquiring about the way that the respondents teach.



### *Organization Pattern*

Questions from 10 to 12 asked the teachers about the organization pattern of their class. Fifty one percent of them claimed that more than 40% of their class time was conducted full class (teacher centered). More than half of them (respectively 64%, 59%, 57%) claimed that their class time for group work (large or small) or pair work was less than 20%. Also, more than half of them (56%) reported that less than 30% of their class time was spent on task-based activities.

### *Language Used in Class*

Fifty five percent of the teachers claimed that more than half of their classroom instruction was conducted in English.

### *Focus of Instruction*

Q14, Q15 and Q58 asked about the percentage or frequency of the teachers' classroom instruction devoted to teaching grammar, vocabulary and translation. While 69% of the teachers claimed that less than 20% of their class time was spent on grammar, 43% of them claimed that over 30% of their class time was spent on vocabulary.

The histogram below (Figure 16) shows the amount of time spent on vocabulary across all teachers and all classes. It indicates that the amount of time the teachers claimed to spend on vocabulary still remains relatively high.

However, when teachers' responses to Q14 ( $M=2.18$ ,  $SD=1.31$ ) and Q15 ( $M=3.78$ ,  $SD=1.38$ ) are combined, the results suggest that teachers did not think that grammar and vocabulary were the focus of their instruction.

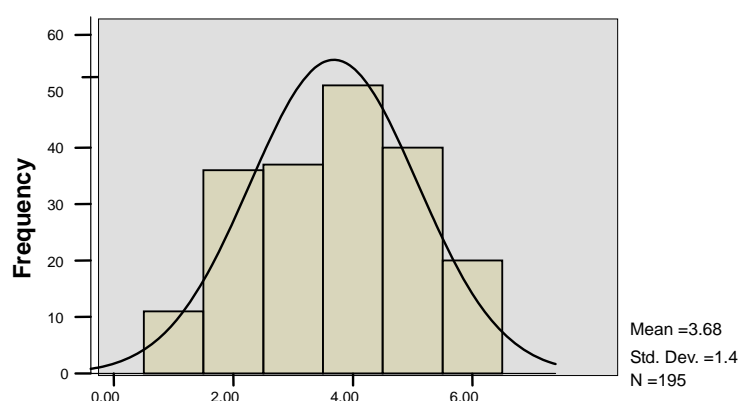


Figure 16. The percentage of time spent on vocabulary.

Q34 asks what types of activities the teachers think are effective in getting students

involved in the learning process. Based on frequency counts, 68% of the teachers seemed to favor activities such as information gap, role-play, language games, and group work or pair work.

#### *The Extent to Which They Teach to the Test*

Q9e and Q39 asked the respondents what percentage of their class time is spent on testing activities and to what extent their teaching is based on what is tested. Only 14% of the respondents claimed that they used the CET to guide and organize their lessons. Sixty-one percent of the teachers made the selection that 0-10% of their class time was spent on testing activities.

#### *5.5.2.2.2 Teachers' knowledge base.*

##### *Pedagogical Knowledge*

Q25 and Q51 ask the respondents to assess their own pedagogical knowledge (e.g., whether they know how to go about things in the course of their instruction and whether they are clear on the principles underpinning CLT). Correspondingly, Q40 and Q43 are designed to assess the respondents' actual understanding of CLT. The results (see Table 27) show that although the majority of them claimed (64%,  $M=2.75$ ; 81 %,  $M=3.11$ ) and close to half of them demonstrated a basic understanding of the meaning of CLT (47%,  $M=2.45$ ; 46%,  $M=2.42$ ), there still exist another half of the teachers (53%, 54%) who did not seem to understand it.

Such evidence reveals that quite a number of teachers still have not achieved a good understanding of CLT or they do not understand CLT in an adequate way.

Table 27 *Descriptive Statistics 3 (n=195) (D=Disagree; A=Agree; N=No; Y=Yes)*

Question	Participants' Responses		(M)	(SD)	95% Confidence Interval for Mean	
	SD/D or N	SA/A or Y			Lower Bound	Upper Bound
25	69 (35.6%)	125 (64.4%)	2.75	.70	2.65	2.85
51	37 (19.1%)	157 (81%)	3.11	.79	3.01	3.23
40	102 (52.6%)	92 (47.4%)	2.45	.93	2.31	2.59
43	105 (54.1%)	89 (45.9%)	2.42	.99	2.27	2.56

The histogram below (Figure 17) reflects pedagogical knowledge across all

teachers and all classes.

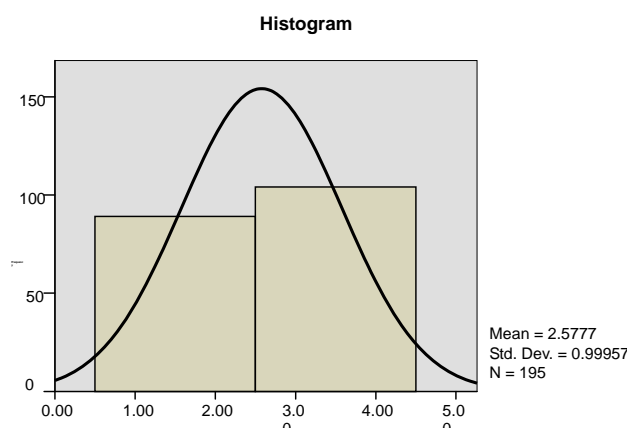


Figure 17. Histogram of pedagogical knowledge.

A group of questions (Q45, Q46, Q47, Q30, Q31) aimed at finding out about the teachers' curriculum knowledge of the revised CET and the CECR, such as whether they are familiar with the goals and objectives of the revised CET, and what skills are assessed in the test, etc.

Based on teachers' claims, 74% (for Q45) and 68% (for Q46) of them were familiar with the revised CET. Their responses to Q 30 and Q31 illustrate that the majority of them (82%, 78%) did demonstrate a basic understanding of the revised CET.

In addition, as for Q53 and Q42, concerning their knowledge of the CECR, 79% of the respondents claimed that they are familiar with the CECR. Their responses to Q42 indicate that 71% of the teachers demonstrated a basic understanding of the CECR.

#### 5.5.2.2.3 Teachers' past experience.

##### *Exposure to task-oriented activities*

Q56 seeks to find out whether the respondents have ever been exposed to task-oriented activities. The results suggest that the number of teachers (97) who have a lot of exposure to task-oriented activities almost equals the number of those (93) who have little exposure to them.

##### *Teacher Training*

For Q57 and Q8, whether they have been involved in workshops focusing on teaching methodology, and whether they have had specific training in ESL, the teachers' answers are quite similar. The number of teachers who have received

training (95, 87) almost equals the number of those who have not (98, 107). However, as for the type of training they have received, the teachers reported that they were mainly lectures offered by professors of the university.

### 5.5.3 Results from Inferential Statistics

In the previous section, the survey results concerning the principal aspects involved in the washback phenomenon—various components of the ‘teacher factor’ and teacher practice were respectively presented. In this part, the major research topic of this study—“how the ‘teacher factor’ is manifested in the washback effect in the Chinese context”—is explored more extensively. Concretely, the salient findings derived through inferential analyses of the questionnaire data are presented.

Based on the literature review, the ‘teacher factor’ is an integrated system encompassing interwoven attributes – beliefs, knowledge, and experiences (BKE). It is widely believed to be closely related to how teachers react to pedagogical change in terms of their beliefs and behaviors. The qualitative data derived in this study demonstrated that teachers’ BKE seems to play an essential role in the kind of washback effects that take place in classroom contexts (e.g., how teachers interpret testing innovations and how they change their practices). However, whether there is a direct relationship between the networked ‘teacher factor’ and the washback effect of the CET needs to be further examined empirically. To gather this empirical evidence, four steps were followed in this study.

**Step 1:** It is important to realize that although two constructs (see Figure 6 and Figure 7 in Section 4.5.3.5.1) were hypothesized when the clusters of questionnaire items were developed based on the literature, they were not tested empirically. Thus as indicated above, tests need to be conducted to explore whether there are underlying constructs of the ‘teacher factor’ and ‘teacher practice’ among the questionnaire items of this study. In this step, EFA was performed to identify how the questionnaire items clustered. Its purpose was to explore patterns in the data. At the same time, it could help to verify whether the underlying constructs correspond to those hypothesized.

**Step 2:** This step involved testing models of the ‘teacher factor’ and ‘teacher practice’. That is, if the constructs of the ‘teacher factor’ and ‘teacher practice’ are found to exist through the use of EFA, a model of the ‘teacher factor’ as well as a

model of the ‘teacher practice’ can be then hypothesized (see Model 1 and Model 2, as shown in Figure 20 and Figure 22) and tested. Model 1 and Model 2 were hypothesized, for they would serve as the basis for Model 3 and Model 4. Also, one of my research interests in this study was to confirm whether the model of the ‘teacher factor’ (Model 2) is one, as discussed in the literature, mainly made up of interwoven attributes such as BKE. In this step, the model was tested through using CFA. As indicated in Section 4.5.6.3.4.2, CFA is regarded as a special case of the SEM. It was considered to be useful in determining whether the predefined factor model fits an observed set of data. In my study, I was most interested in testing whether aspects of the ‘teacher factor’ are correlated or uncorrelated and how they are correlated. Here, we should note that CFA requires the specification of one or more hypothesized factor structures. Based on Kline (2005), in performing CFA, the first thing that should be done is to precisely define the model we wish to test. With respect to my study, the models of the ‘teacher factor’ and ‘teacher practice’ were constructed based on the results of the EFAs.

**Step 3:** Step 3 and Step 4 involved tests conducted to examine how the integrated system of the ‘teacher factor’ as well as its respective components relates to how teachers interpret and react to washback. In this step, a causal relationship was hypothesized to exist between the interrelated components of the ‘teacher factor’ and teacher test-related practices (see Model 3 as illustrated by Figure 24). Provided that Model 3 is confirmed, the networked ‘teacher factor’, rather than the CET itself, might be held responsible for what teachers do in the classroom. As Step 3 and Step 4 involved analyses of more complicated measurement models that had been hypothesized, SEM was performed. The rationale for this was that the SEM would allow complex relationships between variables to be examined. (Bentler, 1987, 1992; Kline, 2005).

**Step 4:** In the last step, another test was conducted on how the ‘teacher factor’ and its respective components relate to teachers’ beliefs about test impact on teaching. The reason for conducting this test was that drawing on the literature review, teachers’ negative views of test impact on teaching were assumed to be in line with the network of their BKEs. This assumption seems to suggest that there might be a causal relationship between teachers’ beliefs about test impact on teaching and the networked components of the ‘teacher factor’. Thus, Model 4 was hypothesized to test this relationship. The model may help to account for Chinese teachers’ accounts of the

constraint of the CET on teaching (see Section 1.2). As mentioned above, SEM was used to perform this test.

Here, it is important to explain that due to length restrictions of this dissertation, not all aspects of the ‘teacher factor’, ‘teacher practice’ and all questionnaire items were included in the following analyses. Rather, I have chosen to focus on some major aspects and questionnaire items relevant to my research questions. In addition, I need to explain that Model 3 and Model 4 are unrelated to each other, as they serve to address different parts of the second research question. Hence, the two models would not be combined into one.

To summarize the models hypothesized above, Model 1, depicted by Figure 20, posits a one-factor model of ‘teacher practice’ which involves four observed variables (items) related to the way teachers teach. Model 2, as Figure 22 demonstrates, postulates that the variables of the ‘teacher factor’ are interrelated. Model 3, reflected by Figure 24, hypothesizes that the interrelated components of the ‘teacher factor’ combine to exert an impact on instruction. Model 4, shown in Figure 26, presents the hypothesis that a link exists between the assumed washback (teacher beliefs about the impact the CET on teaching) and related aspects of the ‘teacher factor’ (BTL, BW, TE, PK).

Table 28 provides a summary of the above steps that were followed in conducting inferential analyses.

Table 28 *Procedures of Inferential Analyses*

Procedures	Objectives	Questionnaire Constructs and Tested Models	Method of Analysis
<b>Step 1</b>	To explore how the questionnaire items cluster; whether the clusterings discovered correspond to the hypothesized clusterings.	The Construct of the 'teacher factor' and the construct of 'teacher practice'	EFA
<b>Step 2</b>	To establish the base models for Model 3 and Model 4; to test the intricate relationships among various aspects of the 'teacher factor' (whether the factors are interrelated).	Model 1 + Model 2	CFA
<b>Step 3</b>	To test the causal relationship between the 'teacher factor' and the behavioral dimension of washback.	Model 3	SEM
<b>Step 4</b>	To test the causal relationship between the 'teacher factor' and the perceptual dimension of washback.	Model 4	SEM

In the following sections, the hypothesized models were tested using (1) Item-Level Exploratory Factor Analyses (EFAs), (2) Confirmatory Factor Analyses (CFAs), and (3) Structural Equation Modeling (SEM). The results obtained through these analyses are presented below.

Before modeling the multifaceted relationships among various aspects of the 'teacher factor', a series of item-level analyses were carried out on the questionnaire data to assess the clusterings of the items and examine whether they correspond to the hypothesized clusterings presented in Section 4.5.3.5.1.

I performed EFAs on both the 15 'teacher factor' items and 4 'teacher practice' items in the questionnaire. As indicated in Chapter 4, initially, I examined the means, standard deviations, skewness and kurtosis of each item and each variable to see whether the variables were normally distributed. Then, I performed a number of internal consistency reliability analyses on the data to examine the homogeneity of the items.

### 5.5.3.1 Item-Level Exploratory Factor Analysis for the Questionnaire

#### 5.5.3.1.1 Items of the ‘teacher factor’.

##### 5.5.3.1.1.1 Distribution.

The descriptive statistics for the 15 items are presented in Table 29. The means ranged from 1.95 to 3.14 and the standard deviations ranged from 0.73 to 1.00. The medians and modes ranged from 1.5 to 3.5. The values for skewness ranged from -1.68 to 1.33, and kurtosis ranged from -2.02 to 0.82. All values for skewness and kurtosis were within the accepted limits of  $\pm 3.0$ , indicating that the items appeared to be normally distributed.

Table 29 *Descriptive Statistics – Distribution of the various aspects of the ‘teacher factor’*

Items	Mean	Std. Dev	Skewness	Kurtosis	Median	Mode
Q8	2.40	1.00	.20	-1.97	1.50	1.50
Q57_1	2.49	1.00	.02	-2.02	1.50	1.50
Q55	2.62	.99	-.24	-1.96	3.50	3.50
Q52	2.73	.98	-.47	-1.79	3.50	3.50
Q 50	2.30	.98	.40	-1.86	1.50	1.50
Q 49	1.95	.84	1.33	-.243	1.50	1.50
Q 43	2.42	.99	.17	-1.99	1.50	1.50
Q 40	2.45	1.00	.10	-2.01	1.50	1.50
Q 24	2.53	.93	-.03	-.84	3.00	3.00
Q 26	2.46	.79	.18	-.38	2.00	2.00
Q 27	2.45	.91	.03	-.79	2.00	2.00
Q 23	2.23	.73	.41	.141	2.00	2.00
Q 21	2.43	.77	-.10	-.42	2.00	3.00
Q 30	3.14	.77	-1.68	.82	3.50	3.50
Q 31	3.06	.83	-1.35	.18	3.50	3.50

##### 5.5.3.1.1.2 Reliabilities.

I then computed internal consistency reliability estimates (i.e., coefficient alpha) of the ‘teacher factor’ variables. Table 30 shows the reliability estimates for internal consistency for the 15 items of the questionnaire concerning the ‘teacher factor’. The reliabilities for these items ranged from a low 0.34 for the “teacher curriculum knowledge” (CK) scale to a relatively high 0.79 for the “teacher beliefs about teaching and learning” scale. In general, the reliability estimates for all the scales are



relatively high or moderate except that for the CK scale. Further analysis of the data also showed that the CK variable not only had low reliability coefficient results, it also had low factor loadings, which suggested that it could not be used for subsequent analysis in a meaningful way. Therefore it was dropped from further analysis, leaving five scales in the analysis (See Table 30). The remaining items produced a reliability estimate of 0.79, above the desirable threshold of 0.70 (Garson, 2007).

Table 30 *Reliability estimates for the ‘teacher factor’*

	No. of Items Used	Question No.	Reliability Estimates (alpha)
<b>1. Various Beliefs</b>			
Teacher Beliefs about Teaching and Learning (BTL)	3 items	24, 26, 27	0.79
Teacher Beliefs about the Impact of the CET on Teaching (BI)	3 items	21, 23, 49	0.68
Teacher Beliefs about Washback (BW)	3 items	50, 52, 55	0.72
<b>2. Teacher Knowledge Base</b>			
Teacher Curriculum Knowledge (CK)	2 items	30, 31	0.34
Teacher Pedagogical Knowledge (PK)	2 items	40, 43	0.68
<b>3. ExperienceTeacher Training (TE)</b>	2 Items	8, 57.1	0.63
<b>Total</b>	<b>15 items</b>		<b>0.79</b>

#### 5.5.3.1.1.3 Results.

To investigate how the questionnaire items clustered with their respective ‘teacher factor’ variables, a matrix of product-moment correlations for the ‘teacher factor’ was generated using 13 items.

After the questionnaire data were tested and confirmed as appropriate for exploratory factor analyses with (1) Bartlett’s test of sphericity, (2) the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (see Table 31) , and (3) the determinant of the correlation matrix, a series of EFAs was performed to examine the factor structures of the data (or the clusterings of the items). The initial factor extraction from the principal component analysis, as Table 32 shows, yielded five eigenvalues greater than 1.0, accounting for 69.74 % of the variance. Thus, a five-factor model that reflected theoretical findings from the literature review was established. The results reflect that when five factors were extracted, the percentage of variance explained by these factors is 69.74%. The results indicate that all the ‘teacher factor’ variables proposed in Chapter 4 were accounted for in the clusterings and they could be used for further analysis. Then the varimax rotation was performed, which seemed to maximize parsimony and interpretability.

The KMO for the 'teacher factor' items was 0.74, indicating that the data could proceed with factor analysis. According to Garson (2007), the KMO ranges from 0 to 1, and a score of 0.60 is the acceptable threshold and items with high communality can play a larger role in determining the factor model.

Table 31 *KMO and Bartlett's Test*

<b>Kaiser-Meyer-Olkin Measure of Sampling Adequacy.</b>		.74
<b>Bartlett's Test of Sphericity</b>	<b>Approx. Chi-Square</b>	669.69
	df	78
	Sig.	.00

Table 32 *Eigenvalues and Percentage of Variance Explained for Each Factor*

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.76	28.94	28.94	3.76	28.94	28.94	2.15	16.55	16.55
2	1.66	12.75	41.68	1.66	12.75	41.68	1.93	14.83	31.38
3	1.36	10.48	52.17	1.36	10.48	52.17	1.87	14.38	45.77
4	1.24	9.56	61.72	1.24	9.56	61.72	1.60	12.27	58.03
5	1.04	8.01	69.74	1.02	8.01	69.74	1.52	11.70	69.74

Extraction Method: Principal Component Analysis

Figure 18 shows the screeplot for the eigenvalues presented in Table 32.

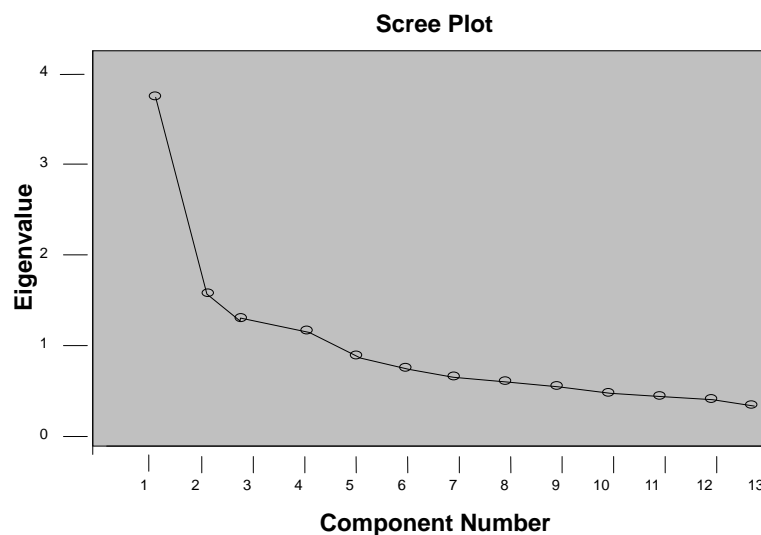


Figure 18. Scree plot of eigenvalues for each factor.

Table 33 presents the five-factor varimax rotation. Factor 1 represents BTL.

Clustered in this factor were items related to their views of teaching approach and teacher role in the classroom. This factor accounts for 28.94 % of the variance with an internal consistency estimate of 0.79. Factor 2 represents BW. The items clustering in this factor ask whether the teachers think that the CET has caused them to change their instructional practice in terms of time allotment, content and methods of teaching. This factor accounts for 12.75 % of the variance. The internal consistency estimate for this factor is 0.72. Factor 3 represents BI. The items in this factor probe into their attitudes toward the impact of the CET on teaching (whether they believe the CET hinders their teaching and their methodology). Approximately 10.48 % of the variance is explained by this factor. The internal consistency estimate for this factor is 0.68. Factor 4 represents TE and factor 5 represents PK. The items in Factor 4 ask whether they have had specific training in ESL and teaching methodology and the items in Factor 5 inquire about whether they are clear of the principles underpinning CLT. Factor 4 accounts for 9.56 % of the variance with an internal consistency estimate of 0.63, while factor 5 accounts for 8.01 % of the variance with an internal consistency estimate of 0.68.

The loading of each item on each factor was listed in Table 33. The breakdown of the variance for the EFA is shown in the pie chart in Figure 19.

Table 34 presents the reliability estimates for internal consistency for the five factors;

Table 35 shows the Factor Correlation.

Table 33 *Results of EFA for the ‘Teacher Factor’: Varimax Rotation*

Questionnaire Items	Factor Loading				
	F1	F2	F3	F4	F5
Q27	<b>.84</b>	.08	.10	.12	.01
Q26	<b>.82</b>	.11	.13	.13	.09
Q24,	<b>.81</b>	.06	.00	.03	.11
Q55	.11	<b>.79</b>	-.06	.15	.18
Q 50	.04	<b>.77</b>	.21	.23	.07
Q 52	.10	<b>.76</b>	.14	-.08	.10
Q 23	.02	.08	<b>.79</b>	.31	.11
Q 21	.13	.04	<b>.75</b>	.21	.20
Q 49	.11	.19	<b>.73</b>	-.26	.03
Q 8	.12	.06	.08	<b>.80</b>	.15
Q 57	.12	.15	.10	<b>.78</b>	.04
Q 40	.16	.12	.12	.13	<b>.83</b>
Q 43	.04	.19	.16	.07	<b>.83</b>

Extraction Method: Principal Component Analysis

Table 31 illustrates the pattern of how the questionnaire items load on each factor, which helps us to determine what each factor may represent.

Table 34 *Rotation Sums of Squared Loadings and Internal Consistency Estimates for Each Factor*

Factor	F1	F2	F3	F4	F5
% of Variance	28.94	12.75	10.48	9.56	8.01
Internal Consistency Estimate	0.79	0.72	0.68	0.63	0.68

Table 35 *Factor Correlation Matrix the 'Teacher Factor'*

Factor	1	2	3	4	5
1	1.000				
2	-.25	1.000			
3	.34	-.26	1.000		
4	-.17	.24	-.21	1.000	
5	-.38	.30	-.40	.26	1.000

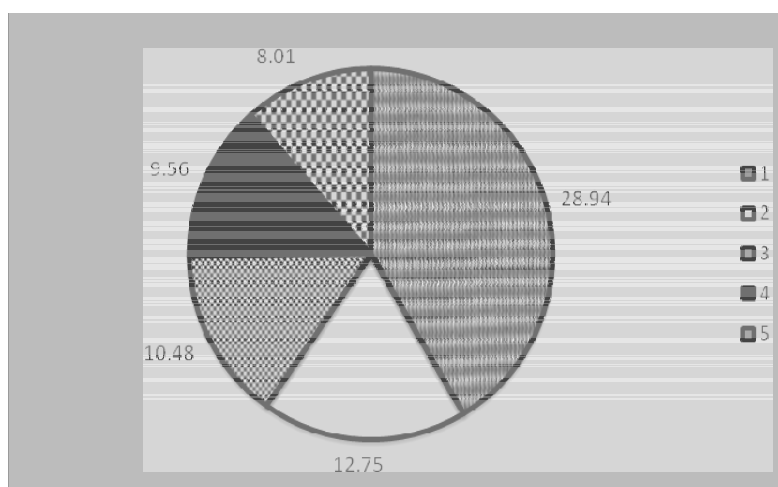


Figure 19. Breakdown of variance for factor analysis.

An examination of the factor loadings indicated that all the items have a high loading of above 0.70. In view of Factor 1, the three items have fairly high loadings of above 0.80. The items involve teacher beliefs about teaching and learning (BTL). For Factor 2 and 3, all the loadings are over 70%. The items clustered in Factor 2 are related to teacher beliefs about the washback effect(s) of the CET (BW), whereas the items nested in Factor 3 concern teacher beliefs about the impact of the CET on teaching (BI). Factor 4 has two items which also load high on it (0.80, 0.78). They

involve teacher training experience (TE). The loadings on Factor 5 also rank very high (0.83, 0.83). The two items cover teacher pedagogical knowledge (PK).

A close look at the inter-factor correlation matrix indicated that these factors were not highly correlated. The findings of these analyses reflect that the initial EFA yielded almost the same clusterings as had been originally hypothesized.

The factor solution in Table 30 was used to form five composite variables to be used in subsequent analyses. Table 36 provides a summary of the items used in the composites for each ‘teacher factor’ variable.

Table 36 *Composites for the ‘Teacher Factor’ Variables (13 Items)*

	No. of Items Used	Question No.	Reliability Estimates (alpha)
<b>1. Various Beliefs</b>			
Teacher Beliefs about Teaching and Learning (BTL)	3 items	24, 26, 27	0.79
Teacher Beliefs about the Impact of the CET on Teaching (BI)	3 items	21, 23, 49	0.68
Teacher Beliefs about Washback (BW)	3 items	50, 52, 55	0.72
<b>2. Teacher Knowledge Base</b>			
Teacher Pedagogical Knowledge (PK)	2 items	40, 43	0.68
<b>3. Experience</b>			
Teacher Training (TE)	2 Items	8, 57.1	0.63
<b>Total</b>	<b>13 items</b>		<b>0.79</b>

#### 5.5.3.1.2 *Items of teacher practice.*

##### 5.5.3.1.2.1 *Distribution.*

The descriptive statistics derived from the reliability analysis, shown in Table 37, consist of the mean, standard deviation, skewness, kurtosis, median and mode.

The means ranged from 1.93 to 3.31 and the standard deviations from 1.35 to 1.64. The medians ranged from 2 to 4. The values for skewness ranged from 0.03 to 1.10, and kurtosis ranged from -1.27 to 0.45. All values for skewness and kurtosis were within the accepted limits, indicating that the items appeared to be normally distributed.

Table 37 *Descriptive Statistics – Distribution of the ‘teacher practice’ variable*

Question	Mean	S D	Skewness	Kurtosis	Median	Mode
15	3.31	1.40	0.12	-0.89	3.00	3.00
11	1.93	1.10	1.10	0.45	2.00	1.00
9b	3.06	1.35	0.33	-0.51	3.00	3.00
9c	3.23	1.64	0.03	-1.27	4.00	4.00

#### 5.5.3.1.2.2 Reliabilities.

The reliability of the items had been estimated and it was found that the instrument yielded a Cronbach's alpha value of 0.67 .

Table 38 shows the 'teacher practice' variable along with the individual questionnaire items designed to measure it, as well as the reliability estimates for internal consistency for the variable. The reliability estimates for the scale are relatively high.

Table 38 *Reliability Estimates for 'Teacher Practice'*

<b>1. Teacher Practice</b>	<b>No. of Items used</b>	<b>Question</b>	<b>Reliability Estimates (alpha)</b>
- CET-related Activities in the classroom	4 items	9b, 9c, 11, 15	0.67

#### 5.5.3.1.2.3 Results.

To investigate how the four items clustered, a matrix of product-moment correlations for teacher practice was generated, and EFAs were performed according to the procedures detailed in both Chapter 4 and above.

The initial factor extraction from principal component analysis yielded one eigenvalue greater than 1.0, accounting for 50.65 % of the variance. Although the percentage of variance explained by this factor is not very high, it is acceptable considering that the construct only contains one factor.

The results, presented in Table 39, show that a one-factor oblimin solution seems to maximize parsimony and interpretability. As shown in this table, the loadings of the items on this factor range from 0.81 to 0.58, all of which are considered to be high or relatively high. Thus, the hypothesis was confirmed that teacher practice, is an unidimensional construct made up of different teacher activities in the classroom.

Table 39 *Results of EFA for Teacher Practice Variables: Oblimin Rotation*

<b>Variable</b>	<b>Factor</b>
Focus on reading	.81
Focus on vocabulary	.76
Focus on listening	.68
Focus on teacher-talk	.58

Extraction Method: Principal Component Analysis

Table 40 *Inter-Item Correlation Matrix for Teacher Practice*

Question	11	15	9b	9c
(11,ALL)	1.000			
(15,ALL)	.24(**)	1.000		
(9b,ALL)	.23(**)	.34(**)	1.000	
(9c,ALL)	.32(**)	.51(**)	.38(**)	1.000

These items appear to be moderately correlated and the correlations are all statistically significant (see Table 40). The significance level is set at .05.

The findings of these analyses enabled me to establish the trait structures of various aspects of the ‘teacher factor’.

In summary, this section described the analyses used in determining the factorial structure of the questionnaire items, using both reliability analyses and EFAs. From these analyses, five ‘teacher factor’ variables, and one ‘teacher practice’ variable were identified, and composites were constructed. These analyses as well as the theoretical framework summarized in Chapter 2 allowed me to posit four hypothesized models.

One hypothesis involves teacher practice. The second hypothesis concerns the interrelationships among various aspects of the ‘teacher factor’, namely the interrelationships among BTL, BI, BW, PK, and TE. Another hypothesis relates to how these interrelationships combine to affect teacher practice. The last hypothesis addresses how other aspects of the ‘teacher factor’ – BTL, BW, PK and TE relate to teacher beliefs about the impact of the CET on teaching (BI). In the following section, these hypothesized models were tested using CFA and SEM procedures.

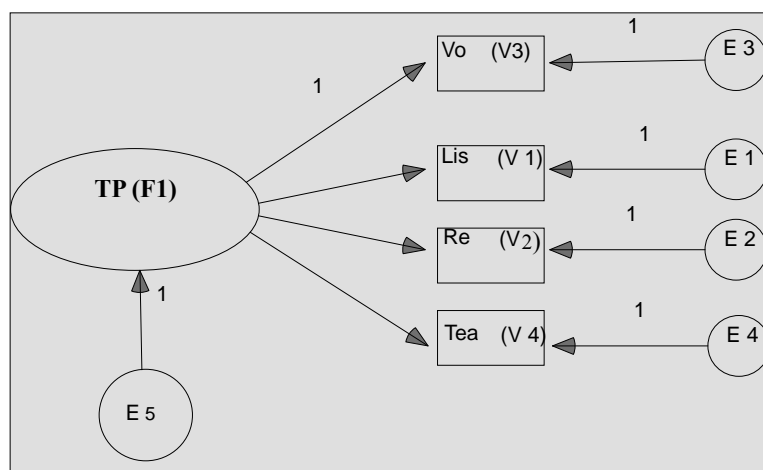
#### 5.5.3.2 *Results from Confirmatory Factor Analysis*

Subsequent to the EFAs, I used AMOS 17.0 to examine the trait structure of the individual measurement models. First I did CFAs relating to teacher practice and various aspects of the ‘teacher factor’ and then I utilized SEM procedures to test how some aspects of the ‘teacher factor’ – BTL, BW, PK and TE combine to affect teacher practice as well as how certain aspects of the ‘teacher factor’ – BTL, BW, PK and TE relates to BI, with the goal of generating the best fitting and most substantively meaningful model(s).

##### 5.5.3.2.1 *The hypothesized model: Model 1.*

Based on the results of the EFAs in the previous section, I postulated a one-factor

model of teacher practice. This model, presented in Figure 20, involved four observed variables related to the way teachers teach (e.g., the percentage of time respectively spent on listening/reading/vocabulary, and the percentage of teacher-talk time in the classroom). The figure provides a schematic representation of the model.



TP: Teacher Practice

Voc: Time on Vocabulary

Re: Time on Reading

Lis: Time on Listening

Tea : Teacher-talk Time

Figure 20. Hypothesized 1-factor model of 'teacher practice': Model 1.

#### 5.5.3.2.2 The results for Model 1.

As presented in the previous section, the skewness and kurtosis values for the 'teacher practice' variables were within acceptable limits, indicating that these variables were univariately normally distributed. With regard to multivariate kurtosis, Model 1 produced a Mardia's coefficient of 0.19, with a critical ratio of Mardia's coefficient (c.r.) that was low (0.19). This result suggests that these variables might be multivariately normal. All other statistical assumptions of the estimation procedure were examined, and no significant violations were found.

I also assessed the hypothesized model to determine to what extent the model fit the sample data. As seen in Table 41, Model 1 produced a CMIN of 0.65 with 2 degrees of freedom. This CMIN/DF ratio (0.33) was far below the recommended value of 2, suggesting a well-fitting model. It produced a CFI of 1.00, a RMR of 0.02 and an RMSEA of 0.00, indicating an excellent model-data fit. These statistics provided strong evidence for acceptance of Model 1.

What I want to explain here is that more indices are provided than needed in the



following tables reporting ‘*Goodness of Fit*’ of the models. I included them in the tables, lest they are needed for reference. As detailed in Section 4.5.6.3.4.3, the indices that I relied on most to examine the hypothesized model fit include the chi-square statistics, the Comparative Fit Index (CFI), and the Root Mean Square Error of Approximation (RMSEA).

Table 41 *Results for the Hypothesized 1-Factor Model of ‘Teacher Practice’:  
Goodness of Fit Summary for Model 1*

Indices	$\chi^2$	df	$\chi^2/df$	NFI	GFI	AGFI	CFI	IFI	RMR	RMSEA
Values	0.65	2	0.33	0.99	0.99	0.99	1	1	0.02	0

Turning to the standardized solution presented in Table 42, I found that the loadings in the standardized solution ranged from a moderate 0.40 to a high 0.76. Nonetheless, all factor loadings were found to be statistically significant at the 0.05 level. The squared multiple correlation for each variable is also included in the table, as this statistic was used as an index for determining the amount of variance accounted for in each independent variable by the predictor variables.

Table 42 *Parameter Estimates for Teacher Practice: Model 1*

Standardized Solution:			R-SQUARED
Focus on reading	=V2	= 0.76* F1+ E2	0.58
Focus on Vocabulary	=V3	= 0.66 F1+E3	0.43
Focus on listening	=V1	= 0.51* F1+ E1	0.26
Focus on teacher-talk	=V4	= 0.40* F1+ E4	0.16

Figure 21 provides a diagrammatic representation of Model 1 in which the standardized parameter estimates are indicated. An inspection of the model shows that teacher classroom practice is well represented by one factor measured by four observed variables with no correlated error terms among the observed variables. The factor was explained by the way teachers taught listening, reading, vocabulary, and the way they organized their class (in terms of the teacher-talk time in the classroom), since all these variables displayed a relatively strong (with factor loadings ranging from 0.40 to 0.76) and significant (at 0.05 level) association with the factor. They produced R squares (R<sup>2</sup>) ranging from 0.16 to 0.58, which were not strong. In spite of the strength of the relationships, all paths were significant at the 0.05 level.

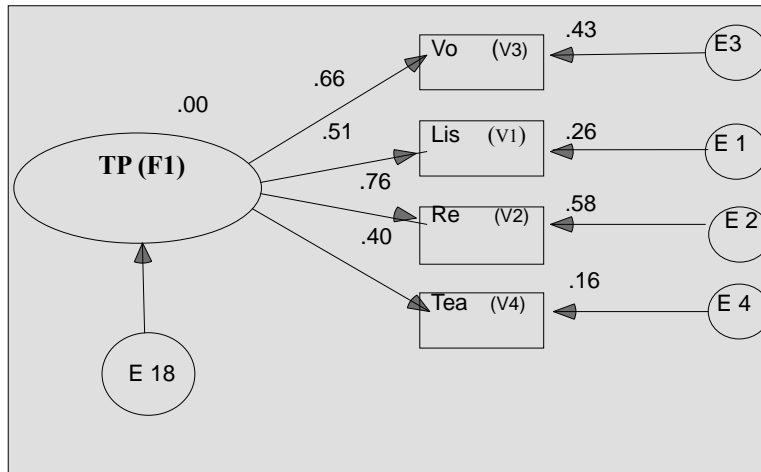


Figure 21. The model of 'teacher practice'.

In light of the above results, the model provided an excellent explanation of teacher classroom practice.

#### 5.5.3.2.3 The hypothesized model: Model 2.

Based on the EFAs and substantive literature reviewed in Chapter 2, the different interrelated aspects of the 'teacher factor' in washback were represented schematically as a five-factor model of BTL, BI, BW, PK and TE. This hypothesized model is presented in Figure 22. This model contains five intercorrelated factors with 13 observed variables (V5 through V17), with each observed variable loading on only one factor. The errors associated with each observed variable (E5 through E17) are postulated to be uncorrelated.

This model is a first-order CFA designed to test the multidimensional construct of the 'teacher factor'.

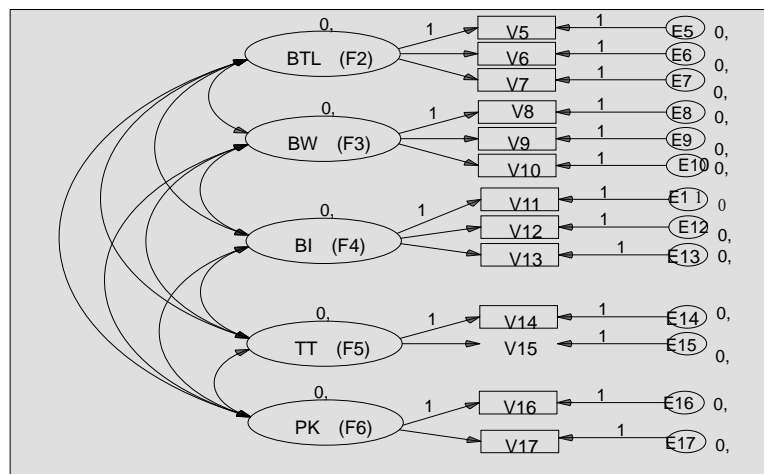


Figure 22. Hypothesized 5-factor model of the 'teacher factor' : Model 2.

#### 5.5.3.2.4 *The results for Model 2.*

Prior to examining the inter-relationships among various aspects of the ‘teacher factor’, I examined the statistical assumptions underlying the maximum likelihood estimation procedures. As indicated earlier, the skewness and kurtosis values for the ‘teacher factor’ variables were within acceptable limits, showing that these variables were univariately normally distributed. With regard to multivariate kurtosis, these sample data produced a Mardia’s coefficient of -2.46, with a low critical ratio of Mardia’s coefficient (c.r.) (-0.87), thereby suggesting multivariate normality. All other statistical assumptions of the estimation procedures were examined, and no significant violations were found.

The hypothesized model was then assessed to determine to what extent the model fit the sample data. All goodness of fit indices were high, indicating a good model-data fit.

As seen in Table 43, Model 2 produced a chi-square value (CMIN) of 64.5 with 55 degrees of freedom (DF). This CMIN/DF ( $\chi^2/\text{df}$ ) ratio (1.17) was well below the recommended value of 2, suggesting a well fitting model. The Normed Fit Index (NFI) (also known as Bentler-Bonnet normed fit index) was 0.91, and Goodness of Fit Index (GFI) and Adjusted Goodness of Fit Index (AGFI) respectively were 0.96 and 0.93, all of which were above the recommended value of  $>0.9$ . The high Comparative Fit Index (CFI=0.98,  $\approx 1$ ) and high Incremental Fit Index (IFI=0.99,  $\approx 1$ ), low Root Mean Square Residual (RMR=0.03,  $< 0.05$ ) and low Root Mean Square Error of Approximation (RMSEA=0.03, much below the accepted value of  $< 0.05$ ) offered further evidence to suggest that Model 2 fit the sample data. The overall goodness of fit suggests that Model 2 might be an appropriate representation of the interrelationships of various aspects of the ‘teacher factor’.

Table 43 *Results for the Hypothesized 5-Factor Model of the ‘Teacher Factor’:  
Goodness of Fit Summary for Model 2*

Indices	$\chi^2$	df	$\chi^2/\text{df}$	NFI	GFI	AGFI	CFI	IFI	RMR	RMSEA
Values	64.5	55	1.17	0.91	0.96	0.93	0.98	0.99	0.03	0.03

I then examined the feasibility of the individual parameter estimates. Information on the standardized parameter estimates is provided in Table 44.

Table 44 *Parameter Estimates for Teacher Factor: Model 2*

Standardized Solution:			R-SQUARED
<b>BTL</b>	=V5	= 0.67* F2+ E5	0.45
	=V6	= 0.81* F2+ E6	0.66
	=V7	= 0.77* F2+ E7	0.60
<b>BW</b>	=V8	= 0.77 F3+ E8	0.59
	=V9	= 0.57* F3+ E9	0.33
	=V10	= 0.69* F3+ E1	0.48
<b>BI</b>	=V11	= 0.79* F4+ E11	0.63
	=V12	= 0.46* F4+ E12	0.21
	=V13	= 0.73* F4+ E13	0.53
<b>TE</b>	=V14	= 0.67* F5+ E14	0.46
	=V15	= 0.68* F5+ E15	0.46
<b>PK</b>	=V16	= 0.73* F6+ E16	0.53
	=V17	= 0.71* F6+ E17	0.50

Model 2, along with the standardized parameter estimates is presented in Figure 23. An inspection of the model shows that the ‘teacher factor’ is represented by five interrelated underlying factors measured by 13 observed variables (from V5 to V17).

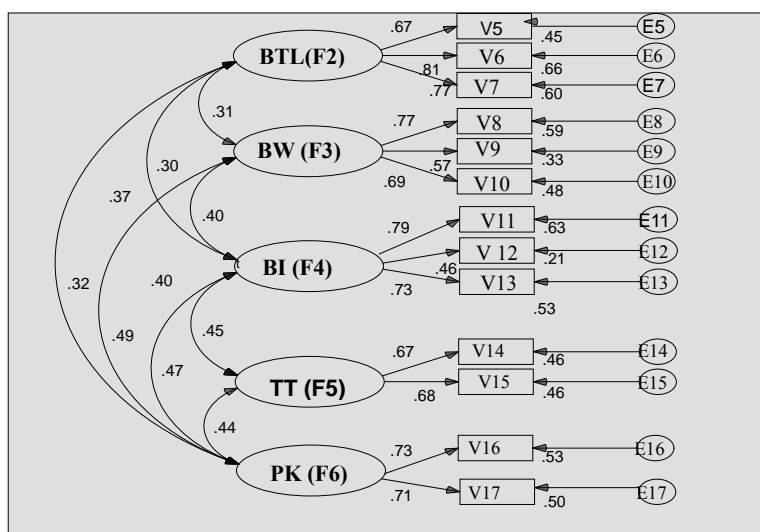


Figure 23. The model of the relationships Between various aspects of the ‘teacher factor’.

From this figure, we see that all the items only load with their respective latent factor (from F1 to F5) and all the standardized factor loadings (or regression weights) for Model 2 were, as a whole, relatively high, ranging from a moderate 0.46 to a high 0.81. Meanwhile, they were reasonably and statistically significant at the 0.05 level. In addition, the squared multiple correlation coefficients ( $R^2$ ), corresponding to 13

observed variables, ranged from 0.21 to 0.66. This means that the BTL factor explains about 66% of the variance in V6, while the BI factor accounts for 21% of the variance in V12. The results indicate that each of the five latent factors is well measured by its respective observed variables. In other words, all the items displayed a relatively strong, significant association with the respective factors that they were hypothesized to load on. Such results imply that the items selected were good indicators of various aspects of the 'teacher factor' under study.

Turning to the correlations among the factors (indicated in Figure 23 by the double-headed arrows), we find that the inter-factor correlations among these factors (ranging from 0.30 to 0.49), though not high, were positive and within the acceptable range of -1 to +1, and above all, statistically significant. The results reflect that despite the differences displayed in the strength of correlation, the components of the 'teacher factor' were significantly positively correlated with one another. An inspection of the model shows that the significant correlations appear stronger between PK and BW (0.49); PK and BI (0.47); TE and BI (0.45); TE and PK (0.44) than other relationships. In light of the significant correlations among these factors, the predicted interrelationships were confirmed, which suggests that interactive and interconnected relationships exist among these factors.

These statistics provided strong evidence for acceptance of Model 2.

Given the relatively moderate inter-factor correlations, each aspect of the 'teacher factor' cannot be seen as purely independent measures, but should be treated as complex networks that are related to one another. This is an important finding. With its five intercorrelated factors and thirteen measured variables, the model supports the hypothesis that the 'teacher factor' is a multidimensional construct composed of teachers' various beliefs, knowledge and experience (explained by teacher training), thereby confirming the hypothesis represented in Model 2.

As Model 1 and 2 showed good model-data fit, they would be used in the full-latent model for positing and testing the relationships between the 'teacher factor' and teacher practice.

### 5.5.3.3 Results from Structural Equation Modeling

#### 5.5.3.3.1 The hypothesized model: Model 3.

Based on the results of the analyses of Model 1 and 2, I formulated a model of the relationship between the 'teacher factor' and teacher practice. The 'teacher factor' is

represented by Model 2, the five-factor model with 13 observed variables, while teacher practice is characterized by Model 1, the one-factor model with four variables. The model in Figure 24 is a schematic representation of such a relationship. The ‘teacher factor’ is hypothesized to have a direct impact on teacher classroom practice. The underlying five factors (BTL, BW, BI, PK, TE) are all hypothesized to be correlated with each other.

Model 1.3 addresses one of the research questions of this study, namely,

*“How is the ‘teacher factor’ manifested in washback in the Chinese university context? What aspects of the ‘teacher factor’ (e.g., beliefs, knowledge, past experiences) contribute to the way that teachers react to washback?”.*

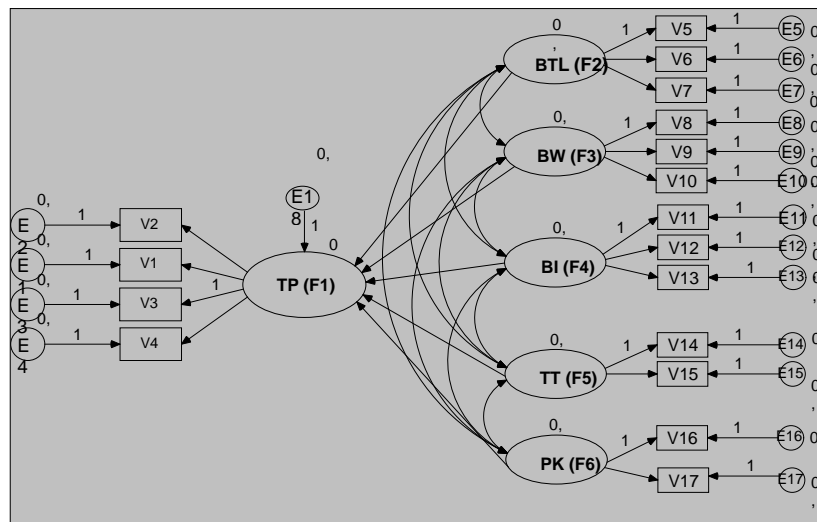


Figure 24. Hypothesized model of the relationship between the ‘teacher factor’ and teacher practice : Model 3.

#### 5.5.3.3.2 The results for Model 3.

To explore the relationship between the ‘teacher factor’ and teacher classroom practice hypothesized in Model 3, the statistical assumptions underlying the estimation procedure used in these analyses were first examined, and then model fit was assessed.

As presented earlier, the skewness and kurtosis values for the ‘teacher factor’ and teacher practice variables were within acceptable limits, showing that these variables were univariately normally distributed. With regard to multivariate kurtosis, these sample data produced a Mardia’s coefficient of 3.26, with a low critical ratio of Mardia’s coefficient (c.r.) that was below 1.96 (0.89), thereby suggesting multivariate

normality.

The hypothesized model was then assessed to determine to what extent the model fit the sample data. Except the NFI (0.86) and AGFI (0.89), all goodness of fit indices were relatively high, indicating a good model-data fit. As seen in Table 45, Model 3 produced a CMIN of 142.95 with 104 degrees of freedom. This CMIN/DF ratio (1.37) was below the recommended value of 2, suggesting a well fitting model. Although the model produced a relatively low NFI of 0.86 and AGFI of 0.89, it produced a high CFI of 0.96. The RMSEA showed a value of 0.04, which was within the recommended value of 0.05. These statistics provided evidence for acceptance of Model 3.

In light of the overall goodness of fit, the model was found to be best fitting representation of the relationship between the ‘teacher factor’ and teacher practice. Figure 25 provides a diagrammatic representation of Model 3, in which the standardized estimates are indicated.

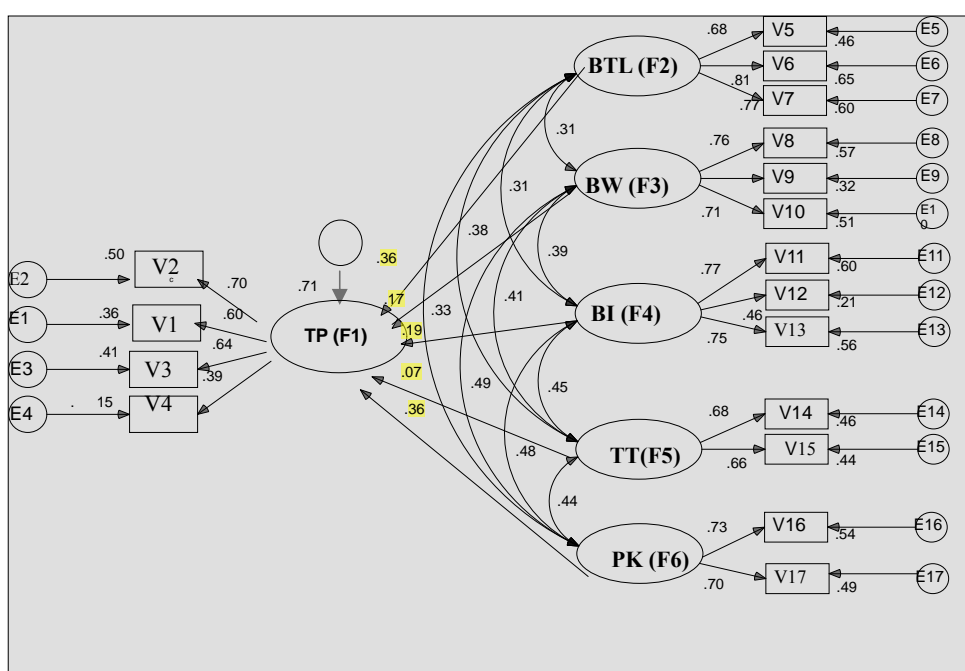


Figure 25. The model of the relationship between the ‘teacher factor’ and teacher practice: Model 3.

Table 45 Results for the Hypothesized Model of the Relationship Between the ‘Teacher Factor’ and Teacher Practice: Goodness of Fit Summary for Model 3

Indices	$\chi^2$	df	$\chi^2/df$	NFI	GFI	AGFI	CFI	IFI	RMR	RMSEA
Values	142.95	104	1.37	0.86	0.92	0.89	0.96	0.96	0.05	0.04

Table 46 *Parameter Estimates for the Relationship(s) Between ‘Teacher Factor’ and Teacher Practice: Model 3*

Standardized Solution:			R-SQUARED		
TP	=V1	= 0.60*	F6+	E1	0.36
	=V2	= 0.71*	F6+	E2	0.50
	=V3	= 0.64*	F6+	E3	0.41
	=V4	= 0.39*	F6+	E4	0.15
BTL	=V5	= 0.67*	F2+	E5	0.46
	=V6	= 0.81*	F2+	E6	0.65
	=V7	= 0.77*	F2+	E7	0.60
BW	=V8	= 0.76	F3+	E8	0.57
	=V9	= 0.57*	F3+	E9	0.32
	=V10	= 0.71*	F3+	E10	0.51
BI	=V11	= 0.77*	F4+	E11	0.60
	=V12	= 0.46*	F4+	E12	0.21
	=V13	= 0.75*	F4+	E13	0.56
TE	=V14	= 0.68*	F5+	E14	0.46
	=V15	= 0.66*	F5+	E15	0.44
PK	=V16	= 0.73*	F6+	E16	0.54
	=V17	= 0.70*	F6+	E17	0.49

As seen in Figure 25, the relationship between the ‘teacher factor’ and teacher practice is a complex one. Despite its complexity, the model appeared to represent the sample data well from both a substantive and statistical point of view.

First, the standardized regression weights associated with BI, BW and TE (respectively, 0.19, 0.17, 0.07, as indicated in Figure 25 by the single-headed arrows) were low and statistically nonsignificant, which suggests that these factors did not exhibit significant effects on teacher practice. However, as the model is a good fit to the data, these factors helped, at varying degrees, improve model-data fit. Second, the standardized regression weights associated with PK and BTL are both 0.36. Although the loadings were not high, the two relationships were significant at the 0.05 level. The results reflect that of the two relationships, teacher pedagogical knowledge and teacher beliefs about teaching and learning respectively exhibited significant, direct effects on teacher practice.

Third, an inspection of the feasibility of the individual parameter estimates revealed that the standardized regression weights for the individual observed variables range from a moderate 0.39 to a high 0.81. However, all of them were reasonable and statistically significant at the 0.05 level. Moreover, the  $R^2$  statistics corresponding to the 16 observed variables indicate that except V4 (15%) and V12 (21%), the  $R^2$  values for the respective factors explain a respectable portion of the variance (between 32% and 65%). These results imply that the underlying factors are well measured by the



observed variables. Information on the standardized parameter estimates is provided in Table 46. Fourth, further examination of the data revealed that similar to the results relating to Model 2, although the correlations among these factors are moderately positive (ranging from 0.31 to 0.49), they are statistically significant. The results further confirmed the inter-correlated relationships among these factors. They demonstrated that while factors such as teacher beliefs about the impact of the CET on teaching and learning, teacher beliefs about washback and teacher training did not, respectively, significantly affect teacher practice, they were interrelated to exert a positive impact on teacher practice.

These results are interesting in that they enabled us to identify the special effect that certain variables (PK and BTL) had on teacher practice. They demonstrate that teacher pedagogical knowledge and teacher beliefs about teaching and learning play a substantially greater role in the way teachers deliver their lessons than do any of the other factors. Additionally, the model provides some interesting insights on the relationship between the ‘teacher factor’ and teacher practice.

#### 5.5.3.3.3 *The hypothesized model: Model 4.*

Based on the results of the analyses of Model 2 and substantive review of the literature, I also formulated a model of the relationship between certain aspects of the ‘teacher factor’ (BTL, BW, TE, PK) and teacher beliefs about the impact the CET on teaching (BI). The model in Figure 26 is a schematic representation of such a relationship. By exploring this relationship, I aim to answer the following research question:

*How is the ‘teacher factor’ manifested in washback in the Chinese university context? What aspects of the ‘teacher factor’ (e.g., beliefs, knowledge, past experiences) contribute to the way that teachers interpret washback?*

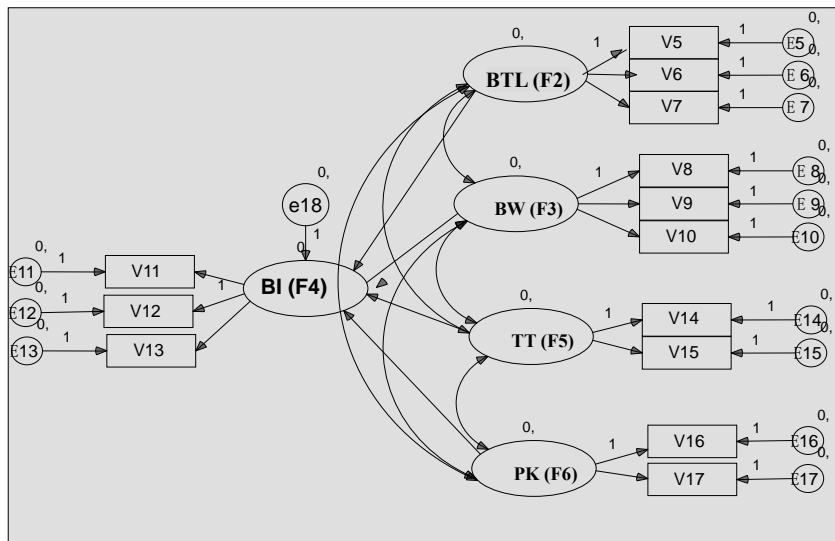


Figure 26. Hypothesized model of the relationship between the assumed washback and some aspects of the ‘teacher factor’ : Model 4.

#### 5.5.3.3.4 The results for Model 4.

The statistical assumptions underlying the estimation procedure used in these analyses were first examined, and then model fit was assessed.

As presented earlier, the univariate values for skewness and kurtosis for the ‘teacher factor’ variables were within acceptable limits, showing that these variables were univariately normally distributed. With regard to multivariate kurtosis, these sample data produced a Mardia’s coefficient of -2.46, with a low critical ratio of Mardia’s coefficient (c.r.) that was below 1.96 (-0.87), thereby suggesting multivariate normality.

I then evaluated the model for overall model-data fit. As seen in Table 47, Model 4 produced a CMIN of 64.53 with 55 degrees of freedom. This CMIN/DF ratio (1.17) was below the recommended value of 2 and the CFI was very high (0.98), suggesting a well fitting model. Except the NFI (0.91) and the AGFI (0.93), all goodness of fit indices was high, indicating a good model-data fit. Both the RMR (0.34) and the RMSEA (0.03) show a value which is below the recommended value of 0.05. These statistics provided strong evidence for acceptance of Model 4.

In light of the overall goodness of fit, the model, as predicted, was found to be the best fitting representation of the relationship between teacher beliefs about the impact the CET on teaching (asserted washback) and other aspects of the ‘teacher factor’(BTL, BW, TE, PK). Model 4, along with the standardized parameter estimates,

is presented in Figure 27.

Table 47 *Results for the Hypothesized Model of the Relationship Between the Assumed Washback and Some Aspects of the ‘Teacher Factor’: Goodness of Fit Summary for Model 4*

Indices	$\chi^2$	df	$\chi^2/df$	NFI	GFI	AGFI	CFI	IFI	RMR	RMSEA
Values	64.53	55	1.17	0.91	0.96	0.93	0.98	0.99	0.34	0.03

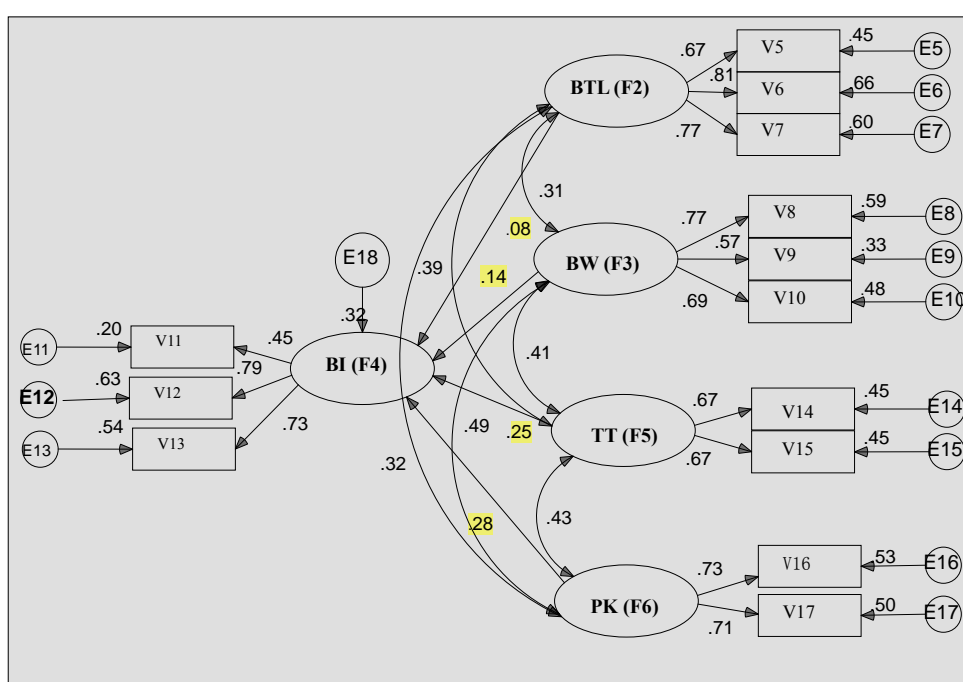


Figure 27. The model of the relationship between the assumed washback and some aspects of the ‘teacher factor’ : Model 4.

The model appeared to fit the data well both statistically and substantively. Turning to the standardized solution, we see that the factor loadings for Model 4 were all within the expected limits and ranged from a moderate 0.45 to a high 0.81. All the values appeared to be reasonable and statistically significant at the 0.05 level. The  $R^2$  statistics range from 0.20 to 0.66.

These results show that the underlying factors are well measured by the observed variables.

I then examined the feasibility of the individual parameter estimates and found only

one factor – teacher pedagogical knowledge produced a direct, significant impact on teacher beliefs about the impact of the CET on teaching with a standardized factor loading of 0.28. Although the magnitude of the impact was modest, it was still statistically significant. The finding further highlights the role of teacher pedagogical knowledge in teacher beliefs about the negative impact of the CET. However, other aspects of the ‘teacher factor’ (BTL, BW, TE) did not seem to yield a direct impact on BI. But, like Model 2 and Model 3, various components of the ‘teacher factor’ were significantly correlated (with  $R^2$ s ranging from 0.31 to 0.49 as indicated in Figure 27 by the double-headed arrows). This seemed to suggest that these interrelated components jointly affect the way teachers perceive the impact of the CET on teaching.

I had originally hypothesized that the interrelationships of the ‘teacher factor’ would exert a direct, positive influence on teacher beliefs about the impact of the CET on teaching. The findings reflect that the effects did appear in the data. Although each of the factors (BTL, BW, TE) alone did not significantly affect BI, they combined to affect the way teachers view the CET.

In sum, in this section, I performed a number of statistical procedures with the data. These procedures include:

- 1) calculating descriptive statistics, inspecting the assumptions regarding univariate and multivariate normality and computing internal consistency reliability estimates (i.e., coefficient alpha) of the variables for all parts of the questionnaire items;
- 2) examining the factorial structures of the questionnaire by carrying out EFAs and investigating the trait structure of four measurement models by conducting CFA and SEM separately.

EFAs allowed for the items sharing similar patterns to load on their respective components and as a result, they were grouped into five factors. The structural models constructed through CFA and SEM helped to describe two main relationships: (1) how various aspects of the ‘teacher factor’ relate to one another and then relate to teacher practice; (2) how various aspects of the ‘teacher factor’ relate to one another and then relate to teacher beliefs concerning the impact of the CET on teaching. As a result of the above analyses, four substantively-sound models which fit the sample data and represent the ‘teacher factor’ were found and formulated.

The results from these analyses provided empirical information regarding the relationship between the ‘teacher factor’ and teacher practice as well as the relationship between teacher beliefs concerning the impact of the CET on teaching and other aspects of the ‘teacher factor’.

From these results, we can draw several conclusions. First the ‘teacher factor’ is a multidimensional construct composed of factors of BKE and these factors are inter-correlated. Second, teacher pedagogical knowledge and teacher beliefs about teaching and learning seem to respectively produce a direct impact on the way teachers teach. Despite the insignificant effect of each of the other single factors on teacher practice, the intertwined BKE has a critical impact on how teachers perform in the classroom. However, this result merits further examination. Third, the way the teachers perceive the CET might be attributed to their pedagogical knowledge, since it produced a direct, significant impact on teacher beliefs about the impact of the CET on teaching. In addition, the data demonstrates that the interrelated BKE also has a powerful influence on teacher perceptions of the CET.

This chapter provided the results and analyses of the different sets of the data obtained from a mixed-methods approach. Sections 5.2 and 5.3 presented an analysis of the CET and the teaching materials employed by the research sites. Section 5.4 summarized the results from interviews and classroom observations, and Section 5.5 discussed the results obtained from the Teacher Questionnaire.

The next chapter will present a discussion of the research findings.

## Chapter 6: Discussion of Results

### 6.1 Introduction

A close examination of different data sets reveals some recurring themes with regard to the ‘teacher factor’ involved in the washback phenomenon. This section discusses the findings in light of the theoretical framework discussed in Chapter 2. Due to dissertation length constraints, it is not feasible to discuss all the possible patterns in this thesis. I have chosen the salient results that address the research question.

As indicated in the literature review chapter, the ‘teacher factor’ in the innovation process has been noted and elaborated on across a range of other subjects. Some of the themes, rules and principles that are related to the washback phenomenon are not unique to language testing innovation, but are representative of educational innovation in general. In this regard, the interdisciplinary framework developed in Sections 2.2 and 2.4 can be used to account for the themes that have emerged from the current study. These themes are classified into five categories. The first thematic category covers three subthemes which relate to the washback phenomenon within the context of this study. They include: (1) a disjunction between the washback rhetoric and classroom reality; (2) different teacher, different school, different effects of washback; (3) negative views on tests and innovation. The second thematic category involves the major aspects of the ‘teacher factor’ that have manifested in the process of this research. It is comprised of (1) conflicting views and contradictions related to the teacher factor in washback; (2) teacher beliefs about teaching and learning, teacher knowledge, teacher experience, and teacher proficiency. The third category pertains to the intricate interrelationships among the ‘teacher factor’ variables and their relationship to teacher practice. The fourth category relates to teachers’ perceptions of the ‘learner factor’ in washback. The last category sums up the principal findings of this study: 1) the identification of two extreme stances toward washback, 2) evidence demonstrating that washback is nested in complex webs of factors, and 3) the appearance of a contextual variable.

The themes in these categories will be discussed in relation to the two corresponding research questions stated in Chapter 5. Here, what should be noted is that the two specific research questions are closely related to one another and all of them serve to address one global question related to the ‘teacher factor’ in washback.

Thus, it might be possible that the themes and patterns to be discussed under one question pertain to other question(s) as well. It might also be possible that questions that have been addressed in one section are revisited in another section of the chapter.

6.2 Research question 1: *To what extent and in what form does washback exist in China in terms of its effect on teacher beliefs (e.g., beliefs about the CET and its impact, beliefs about teaching and learning), and classroom behaviors (e.g., content and particularly teaching methodology)?*

Of the patterns related to the first research question that have emerged from this study, the following three are most salient.

6.2.1 *Disjunction Between the Washback Rhetoric and Classroom Reality*

In the previous chapter, a comparison was made between the way the case-study participants teach and the task characteristics of the revised CET. The observation results indicate that the task types teachers adopted bear little relevance to the test. In terms of the subject matter of teaching, although the focus of emphasis has been shifted from reading to speaking and listening in the CECR and considerable weight has been given to the skill in listening in the revised CET, the corresponding teacher behavioral change fostered by these reforms seemed limited in that the amount of time teachers spent on listening is still disproportional to that they spent on reading. As you can see from Table 23 and Figure 12 in Section 5.4.5.2, the average percentage of class time spent on reading by each participant is still much higher than that spent on listening. Nearly all the participants continued to emphasize the cultivation of their students' reading skills.

This result was supplemented by the survey data, for the majority of respondents (79.4%; 65%; 43.8%) claimed that their class time devoted to listening and speaking was less than 21-30%, while their time spent on reading was more than 41-50%.

In addition, use-oriented activities facilitated by the CECR were still neglected by most teachers. For example, although teachers, as a rule, based their instruction on textbooks, nearly all activities in their books designed for the purpose of enhancing students' speaking competence (e.g., the warm-up activities) were skipped over by them (e.g., as in the case of T1, T2, T3, T6) on the grounds of their being time-consuming and impractical. Further, some teachers (e.g., T3) attributed their abandonment of such activities to the new CET. Nevertheless, the results reflect the

fact that the reason they abandoned these activities has little to do with the CET. It seems that these teachers' charges are ungrounded.

With reference to the pedagogical dimensions of teaching, hardly any observable changes were captured in the teaching methods utilized by the teachers within the whole research process. Despite the participants' (four out of six – T1, T4, T5, T6) statements in the survey that the revised CET has helped improve the way they teach, little evidence has been retrieved from observations showing a paradigm shift in their methods of teaching from knowledge-focused instruction to communication-focused instruction, and from teacher-centered instruction to the student-centered instruction. Except in the class of one participant (T5), seldom was pair-work observed in their classrooms.

The survey responses made by all the participating teachers added further evidence that the revised CET can hardly facilitate methodological change, for 59.8% of them declared that the revised CET could not help to modify their teaching approaches.

While the teachers' focus on reading skills and language forms may relate, to a certain extent, to the old version(s) of the CET (e.g., 1999 version), the excessive attention directed by the teachers toward these aspects of language, as demonstrated in Tables 21, 22 and 23, cannot be attributed to the test, for in the test, the weighting of reading (40%) is merely a little higher than that of the current version (35%), and moreover 'vocabulary and structure' tested as discrete points only constitute a small portion (15%) of the whole test. Hence, we cannot really set up a close linkage between what either version (new or old) of the test asks and what teachers teach in the classroom. Nor can we say that the participants are teaching toward it.

The results from the survey support the above finding, since 61% of teachers claimed that seldom (0-10%) is their class time spent on testing activities. These data combine to suggest that tests do not dictate language teaching practices, and the power of tests appears to have been overemphasized by washback proponents. Their negative power also appears to have been exaggerated by its opponents when it comes to the context of this study. In light of the data, it is hard to characterize the nature of washback in the Chinese CE context as positive or negative.

Although this finding seems to contradict the common portrayal of Chinese EFL teachers as being focused on test-related activities, it is consistent with the finding of Alderson and Wall (1993), for they also contended that no particular linkage can be established between tests and their impact. It also lent support to the findings of



Cheng (1999), Shohamy (1993), and Watanabe (2004b), as they also questioned the power of examinations. The view commonly held by them is that changing an examination is not a sufficient condition for changing teachers' instructional practice.

Drawing on the models established through EFA, CFA and SEM, on the one hand, teacher beliefs about teaching and learning and teacher pedagogical knowledge were respectively found to be exhibiting direct, significant effects on teacher practice. On the other hand, while some sets of beliefs (BI and BW) and teacher experience in terms of professional training significantly affect teacher performance, they were interrelated with other aspects of the 'teacher factor', and these interrelated aspects of the 'teacher factor', as a whole, produced a positive impact on teacher practice. These results reveal that rather than being guided by the CET and the CECR, the teachers teach not only according to their own philosophies or conceptions of how to teach, as well as their knowledge of teaching pedagogy, but also chiefly based on the interwoven network of BKE. They also suggest that the CET alone is insufficient enough to foster methodological change. Rather, it needs to work, in combined efforts, with teacher BKE in order for teachers to improve their practices. The results provide empirical evidence in support of the theory put forth by Woods (1996, 2003), which argued that teachers developed their own interpretation of pedagogical concepts into their own BAK, which influences how they interpret and structure the classroom events. The findings also support the point summarized in Section 2.4.5 that washback studies may have overlooked the interwoven characteristic of teachers' BAKs that innovation researchers had underscored. Owing to the interrelated nature of teachers' BKE, their integrated beliefs (BTL, BW, BI), their knowledge and experience are intrinsically interwoven and form a network to influence the way they interpret events in teaching, learning and testing, and deliver their courses. To be concrete, there are a web of factors at work. Given the intricate and interactive nature of this network, the CET itself appears not to be able to replace the role of other change agents (i.e., teacher BKE and reform teacher practices). Thus, we may conclude that the idea of using tests as a strategy to drive teaching seems to oversimplify our understanding of the process of instructional change. Furthermore, washback projects seem to have focused too much on the outcome, with the process of change overlooked. Such a conclusion corroborates Shohamy's (1993) argument that change in the test cannot reduce the complexities of teaching process into a few manageable activities. It also substantiates Cheng and Curtis' (2004) and Wall and

Alderson (1993) claim that exerting positive washback involves much more than just the design of a “good” assessment.

### 6.2.2 *Different Teacher, Different School, Different Effects of Washback*

Another notable theme concerns differences among teachers as well as schools with regard to the changes induced by testing to teaching. The literature on washback shows that there is considerable agreement among researchers (Alderson & Hamp-Lyons, 1996; Alderson & Wall, 1993; Burrows, 2004; Cheng, 1997; 1999; Hayes & Read, 2004; Shohamy et al., 1996; Tan, 2008; Turner, 2005 and Watanabe, 1996, 2004b) on the assertion that test impact or washback takes different forms and exists at varying degrees in different contexts. The results from this study not only confirm this research finding, but also add evidence that the precise nature of washback and the forms it takes vary not only from school to school, but also from teacher to teacher.

As summarized in the previous chapter, there is considerable variation among teachers’ perceptions and knowledge of the CET and the CECR as well as their willingness and efforts to innovate. The data from case studies provide us two contrasting pictures. One picture reveals the school variance in the washback phenomenon. Contrary to washback proponents as well as opponents who either argued for or argued against the use of exams as an instrument to innovate teaching and learning, the key school teachers of this study (S1) declared themselves immune to the washback effects of the CET. One salient feature exhibited in their responses at the group interview is their resistance to change. Not only were they critical of the reforms, but they also made statements showing indifference to the changes newly introduced in the curriculum and the CET (see S1’s chair’s remarks in Section 5.4.4.1.1). Furthermore, compared to teachers of other case-study schools, their curriculum knowledge was indeed limited and inadequate. Based on these teachers, the test neither acts as a motivator, nor poses a hindrance to their teaching. It should be noted that their perceptions of washback have a significant bearing on the way they approach teaching. As in the case of the participants of the school (T1 and T2), there is little observable evidence showing the test has had any effect on their instruction. Through direct observations, it is apparent that passing the test was not pursued by the participants in the way that was done by those of average universities. Such a difference, according to T1, is due to the fact that average universities rely on the CET

scores to upgrade the rankings or reputation of their schools.

In comparison, there is almost a general consensus among teachers of average schools (S2 and S3) that the change in the CET could, either adversely or positively, induce washback effects on teaching. At the time of the group interview, they demonstrated a better knowledge of both the test and curriculum. When compared to S1 (e.g., T1 and T2), S2 (e.g., T3 and T4) manifests a few differences in the ways the participants delivered their lessons. It is clear through observations that they devoted relatively more time and gave greater emphasis to test-relevant exercises (e.g., listening, fast reading, etc.) than those of the key school.

The other picture is that teachers of the same school or the same local context also hold divergent views as to the nature of the washback effects of the CET on teaching. Some believed that the test promotes their teaching, while others contended that it impedes it. But, regardless of their views about this issue, the classroom performances of the participants, as described in the preceding chapter, exhibit significant differences. Typically, two patterns of teaching were spotted: 1) same textbook, different practice (as in the cases of T3 and T4; T5 and T6): 2) different textbooks, similar practice (as in the cases of T2 and T3 in terms of their emphasis on language forms and reading). For example, one teacher (T3) was using the textbook in a formal way, dealing with it as a means of reinforcing linguistic knowledge, whereas the other teacher (T4) was using the textbook to share with her students issues of cultural differences and meaning. The ways they dealt with their lessons are strikingly different.

Drawing on the research design described by Watanabe (2004b), the existence of the washback effects of the CET is under question, since the conditions described by him (teaching, learning, and/or textbooks are similar in the courses which are taught by two different teachers ) were not met.

The data obtained from the survey also show that the test does not cause all teachers to experience the same type or degree of instructional change. As reported earlier, while 55% of respondents thought that the revised test had caused them to shift their instructional focus from linguistic knowledge to language use and 40% of them held that it can help them improve the way they teach, there correspondingly exist 45% and 60% of them who reported being unaffected by the test. The lack of agreement in teachers' responses to the questions may result from the individual

differences among them and among their schools.

Among the factors that are related to teaching variation, the following are cited by washback researchers: teacher ability, teachers' understanding of tests, teaching experience, adequacy of training (Davison, 2008; Tan, 2008; Tavares & Hamp-Lyons, 2008; Turner, 2008, 2009; Urmston & Fang, 2008;) as central in leading to the differences in their perceptions, interpretations and reactions. The data from this study add evidence to the literature that the 'learner factor' (e.g., learner variation in beliefs, English level, learning habits, etc.) also functions as a contributing factor to teacher and school variability in washback. Detailed information regarding the 'learner factor' can be traced at the end of this chapter.

A careful analysis of the data reveals that in addition to the 'learner factor' and other contributing factors listed above, closely related to the above-mentioned differences among teachers and schools is a variance in teachers' attitudes to change, conceptions of teaching and learning, teacher knowledge base, experience and proficiency, etc., which will be addressed successively in the discussions to follow.

Of particular note is the considerable attention paid to teacher differences in beliefs and behaviours by other education researchers. For instance, Woods (1996) noted that aspects of the teachers' BAK networks may take many different forms and each teacher's system differs from other teachers' systems not only in terms of its individual 'components', but also in terms of the interrelationships among the BAKs. According to him, such differences may account for their different reactions to change as well as the different ways they approach teaching. The points made by Woods (1996) seem to make better sense when the data of this study are further examined.

### *6.2.3 Negative Views on Tests and Innovation*

As the first research question of the study concerns teacher beliefs about the CET and its impact on teaching, in this section the negative views about tests and innovation maintained by the participants of this study are analyzed by taking account of the mainstream negative beliefs toward the CET and CEC detailed in Section 3.3.

The long-standing controversy over the CEC and the CET has been expounded in the preceding chapters. In light of the common concern expressed by the large number of Chinese EFL researchers and practitioners about the negative effects of the CET on CE teaching, here I intend to revisit the issue by drawing on the empirical data from this study to see if there are improvements in teachers' attitudes toward the test after

changes have been introduced to it.

After an analysis of both the case-study and survey data, I find that there is consistency in teachers' interpretations and accounts of the impact of the revised CET on teaching. Overall, teachers' attitudes toward the CET and its impact on teaching, as demonstrated in Table 25 and Figure 28 in the survey data, are (slightly) skewed toward the negative side. The teachers' responses to the survey questions (Q49, Q23) show that 76% and 69% of them continued to regard the revised version of the test as a constraint on their instruction. Their responses to another survey question (Q19) reveal that half of them (49%) were sceptical about its validity. A close examination of the data, however, reflects an interesting finding. That is, these teachers conceptualized and interpreted tests, test impact and washback as three distinct concepts. The survey data indicate that on the one hand, the majority of participants took a critical view of the CET and its impact on teaching in terms of its influence on 'how to teach'. However, on the other hand they hold positive opinions about its impact on learning (65% of them) and its washback effects on teaching in terms of its influence on 'what to teach', with percentages of 65% and 61% respectively. Here we should note that their notion of the term 'test impact' or 'washback' departs from the concept in our discussion in the language testing field. Being unaware of the implied meaning of washback as well as the instrumental value of testing and the underlying purpose of the relevant reforms newly introduced in CE education, it appears that they simply interpreted 'test impact' literally and categorized it as a variable constraining their teaching. Such results are buttressed by the data obtained from the EFA. The data indicate that teacher beliefs about the impact of the CET on teaching (BI) and teacher beliefs about washback (BW) are two separate variables (factors), each with a set of its relevant items loading on it. The results provide some interesting information regarding how teachers interpret 'test impact' and 'washback'. The two terms were interpreted by them as two different concepts. The results seem to imply that a confusion exists in teachers' understanding of the terms. In view of the teachers' confusion about the terms, care must be taken in interpreting the terms. Specifically, a division should be drawn between notions of 'test impact' (the alleged negative impact of a test) and 'washback' (the instructional changes induced by testing to teaching).

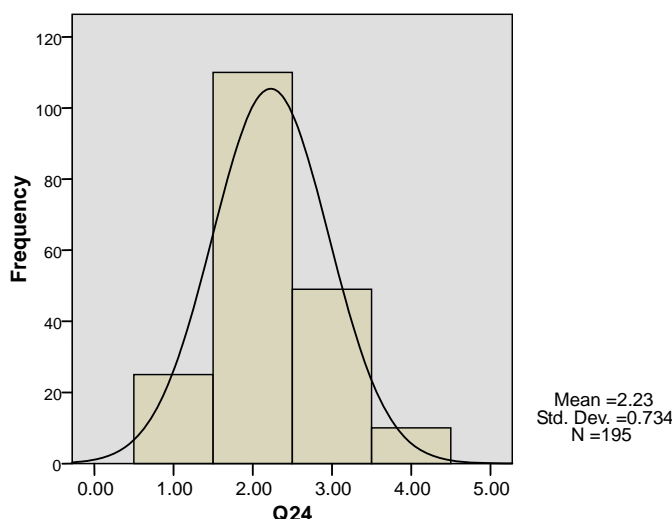


Figure 28. Teachers' perception of the CET impact on teaching.

To explain why teachers were so negative about the test, its reforms and its impact on teaching, I reviewed the comments made by the participants at different interview sessions and found that discrepancies or contradictions often emerged between what the teachers articulated and the realities or facts. The following excerpt from T3 serves as an example.

*“It is impossible for me to carry out communicative activities in class, for the simple reason that we cannot get rid of the CET. Without the test, I will probably increase the student-talk time”.*

As discussed earlier (see 6.2.1), T3 attributed her reluctance and failure to carry out communicative activities to the existence of the CET. However, in analyzing her comments, we find that her complaints about the test seem to be groundless, for her blame on the “impact of the CET” has little to do with the test itself. With the updated changes in the test, the CET should not take the sole blame for the lack of communicative intent in her teaching. The above excerpt helps us to make sense of such discrepancies. We should be mindful that the data in this study contain many instances similar to this one.

Through analysis of the data, I also find that teachers' beliefs and views with regard to the CET, its impact and washback can generally be classified into the following three patterns: 1) being unaware of the notion of washback, giving criticisms to the test and its impact which are test-irrelevant, automatically resisting reforms; 2) being unaware of the notion of washback, placing test-related blame on the test and its

impact; 3) with a little awareness of the notion of washback, holding mixed feelings toward the new test – being willing to change to correspond to the test reform and meanwhile being a little at a loss as to how to change. There is substantial evidence in the case-study data reflecting these patterns.

The above results illustrate that teachers' resistance to change and their negative perceptions of the CET still prevail. Based on the case-study data, the contrasting views the teachers held about the test ('for' or 'against' attitude towards CET) imply that they held incompatible epistemological stances toward teaching, learning and how to teach, as the teachers were often found to be guided by their own theories and ideologies while delivering their lessons. Such data, to a certain extent, correspond to those derived through the CFA and SEM which suggest a direct, positive impact of teacher knowledge of pedagogy on teacher perceptions of the test impact on teaching and which also reveal a significant impact of BKE network on teacher views about test impact (see Figure 27). Following an examination of the data, an inference that can be made is that teachers' reluctance to change and their negative attitudes toward the CET and its impact may have to do with the joint effect of perceptions of teaching and learning, teacher training as well as the knowledge structures that they have already possessed. Also in light of Model 4, one important conclusion that stands out is that the respondents' negative attitudes toward the CET can be attributed to their inadequate pedagogical knowledge.

Such an inference seems to be in line with Woods' (1996) position that teachers tend to interpret pedagogical change according to their BAK networks. Meanwhile, the conclusion seems to be in accord with that of Chapman and Snyder (2000) which highlighted the role of teachers' pedagogical skills in the formation of their beliefs and construction of their behaviors. It is also consistent with that of Andrew et al. (2002) regarding the main reason for teachers' mistranslation of innovators' messages. A similar conclusion drawn from their study is that the prevalent teacher resistance to change may result from the lack of knowledge on the part of teachers about how to change their teaching methods to align with the new curriculum.

Furthermore, the above results confirm the conclusions drawn both by language testing researchers (Andrew et al., 2002; Alderson & Hamp-Lyons, 1996; Cheng, 1999; Shohamy et al., 1996; Smith et al., 1994; Watanabe, 1996) and Chinese EFL researchers (Cen, 1999; Huang, 2002; Liu & Dai, 2003) that teachers' perceptions of washback do not necessarily overlap those of test developers or policy makers.

Nevertheless, it must be noted that although the results reflect that the majority of teachers perceived the CET as adversely affecting teaching (69%,  $M=2.23$ ,  $SD=0.71$ ), there is also evidence showing that the majority of them (65%,  $M=2.67$ ,  $SD=0.66$ ) viewed the test as an active driving force in promoting learning. The finding indicates that while the participants were negative about the CET, they were not as critical of it as some researchers and critics had asserted. The findings contradict both those by the CET constructors (Jin, 2000; Yang & Jin, 2001; Yang, 1999, 2000; Yang, 2003) which dwelt on the facilitating role of the test in Chinese CE teaching and learning, and those by its opponents which attacked it for posing a hindrance to CE teaching and learning (Jing, 1999; Niu, 2001; Liu, 2002, etc.). They also depart from those of earlier research conducted in other EFL contexts, which argue that the teachers perceive the implementation of change as being of benefit neither to themselves nor to their students (Chen, 2002, Watanabe, 1996).

By adopting a dynamic perspective, as Woods (1996) suggested, the notion of teachers' 'resistance to change' can be seen in a different light. In accordance with the evolving nature of teachers' BAKs, a change introduced must evolve in a way that is appropriate and natural to the teachers' evolving BAK networks. However, washback, as a power-coercive strategy, may not be compatible to the natural evolutionary processes of teacher change, especially the change in teachers' BKEs. The existence of such incompatibilities may help explain why teachers choose to resist the change when it is imposed on them. In view of this consequence, it is important for us to bear in mind that teacher change can only be encouraged but not mandated. As a way to deemphasize the power of tests, we must seek ways that involve securing teacher consent, participation and active involvement in the change process. What we should be reminded of is that apart from Davison (2008), Tavares and Hamp-Lyons (2008) and Turner's (2001, 2002, 2005, 2008, 2009) studies, little attention was reported to have been paid to teachers' active involvement in innovation. As Turner (2008, 2009) cautioned us, 'teacher involvement' is a factor that should not be overlooked in washback research.

Based on Woods (1996), Kennedy (1988) and Fullan (1991), such an over-emphasis on the power of tests and absence of attention to teachers' involvement may account for why intended washback did not occur in most studies. Therefore, to ensure the success of instructional reforms, there is a need to de-emphasize the power of tests on the one hand and encourage teacher participation on the other.



6.3 Research question 2.: *How is the ‘teacher factor’ manifested in such a washback effect? What aspects of the ‘teacher factor’ (e.g., beliefs, knowledge, past experiences) contribute to the way that teachers interpret and react to washback?*

The following two major themes presented in Sections 6.3.1 and 6.3.2 relate to the ‘teacher factor’ manifested in washback and they also shed light on issues pertaining to teacher perceptual and behavioral change. Hence, they will be discussed in relation to Question 2.

6.3.1 *Conflicting Views and Contradictions Related to the Teacher Factor in Washback*

As mentioned in the previous section, a recurring theme has emerged from the data showing that gaps exist between teachers’ articulated beliefs and the realities. In-depth analysis of the case-study and survey data shows that the beliefs held by the participating teachers of this study may sometimes be conflicting or lack agreement. For instance, while on the one hand, they claimed that the revised CET has enhanced their understanding of the goal of teaching and has caused real improvement in their instructional practice (see 5.5.2.2.1.2), they accused it of being an impediment to their methodological innovation in the classroom, on the other. On more than one occasion, I was confused not only by the inconsistencies of the views and comments expressed by the participants during the case-study process, but also by the incompatibilities between what they articulated that they did and what they actually did in the classroom. As one of the main research questions in this study is to ascertain the shape of washback in existence in China, such issues need to be clarified.

To achieve a better understanding of the underlying reasons for the innumerable gaps, contradictions, paradoxes, this theme needs to be further exemplified and highlighted. Below are some concrete examples.

6.3.1.1 *Contradictions in Teachers’ Articulated Beliefs on Different Occasions*

The results of this study show that contradictions exist between what teachers articulated on one occasion and what they responded on the other. For example, T1’s responses in the survey run counter to his arguments at the interviews. As manifested in Table 48, when asked whether the CET could produce washback effects on teaching (e.g., whether it would help to improve the way he taught, shift his

instructional focus from linguistic knowledge to language use, and give direction as to what aspects of language needed to be taught ), all his answers were positive.

Table 48 *Washback Effects of the Revised CET Based on the T1 and T6's Accounts (Y=Yes, N=No)*

Participants	Improve the way you teach?	Shift instructional focus from linguistic knowledge to language use?	Give direction as to what aspects of language needed to be taught?
T1	Y	Y	Y
T6	Y	Y	Y

The data from the interviews, whereas, have produced conflicting evidence, for he reiterated on different interview occasions that the CET exerts no influence on teaching and learning in his school.

Similarly, a disparity is shown between T6's radical remarks concerning the negative impact of the test on teaching and learning at the group interview and his positive responses to the same set questions on the survey.

Based on the findings discussed above, one plausible explanation for the conflicting views teachers expressed with respect to the test and its impact is that the teachers were unaware of the instrumental role of testing or the implied meaning of washback. Another possible reason may relate to the highly intricate interrelationships across teachers' belief systems, their knowledge base and experience (as manifested by Model 3). Before pursuing further the reasons for the conflicting views, it is important to examine some more examples of such contradictions.

#### 6.3.1.2 *Contradiction Between How Teachers' Perceive They Are Teaching and the Way They Are Teaching*

Incompatibility also exists between teachers' accounts of what they did and what they actually did. A comparison between the participants' responses in the survey and their practices in the classroom reveals innumerable instances of such a mismatch. In general, discrepancies are shown between how teachers perceived they were teaching and how they were teaching in terms of their teaching approaches, medium of instruction, and teaching focus, etc.

One example is that in the survey, when asked whether the CET has shifted their instructional focus from linguistic knowledge to language use, four participants

checked the answer “Yes”. Since their responses were positive, then a cut should be expected in the amount of time they devoted to linguistic knowledge.

*Table 49 Comparison Between Participants’ Survey Responses and Observation Data in terms of Percentage of the class time devoted to language forms (%)*

Participants	R 1	R2	R3	Shift instructional focus from linguistic knowledge to language use?
T1	47.62	48	40	Y
T2	43.37	55.56	51.84	N
T3	50.51	42.59	56.41	Y
T4	3.62	17.26	28.88	N
T5	43.79	34.72	21.84	Y
T6	64.62	18.31	64.78	Y

However, the results from classroom observations indicate that except for one participant (T4), little change was found in the instructional focus of the other three participants. Their classroom behaviors were still characterized by mechanical manipulation of the formal aspects of language (reference to vocabulary, grammar, and pronunciation). As shown in Table 49 and Figure 11 (see Section 5.4.5.2), a great proportion of class hours (more than 40%) were still spent on them. It was apparent that such a practice contradicts their responses on the survey.

Another example is that even though the teachers at S1 argued at the group interview that they never conducted test-relevant exercises, in R3 it was found that they did, which was alleged by them to be a school practice. Little agreement exists between what they said and what they did.

### *6.3.1.3 Contradiction Between the Teachers’ Perceived Teaching Approach and the Approach They Employ*

Another interesting finding concerns the mismatch between teachers’ accounts of the teaching approach they favour and the one they utilize. More than half of the teachers (54%) responded, in the survey, that they strongly favour CLT over the combined approach and the structural approach, and the interview data also show that most of the participants were interested in applying the principles of CLT as encouraged by the CECR. Nonetheless, the observable data of this study seem to remind us that teachers’ statements in regard to methodological issues should be viewed in an objective way. Based on the results from case studies, despite the participants’ profession both in the survey and at the interviews that they strongly

favour communicative activities over traditional activity types and despite their denial where employing the structural approach in their instruction is concerned, in real practice communicative activities (e.g., information gap, debates, lectures, role plays, discussions and group activities) were seldom noted in their classes. Only T4, T5 and T6 were occasionally seen experimenting with interactive activities in their classes. Other participants, in sharp contrast, continued to adhere to the structural approach (e.g., explanation of vocabulary in a decontextualized manner, introduction of text-related background information using slide shows, explanation of texts in the course-books paragraph after paragraph, either by paraphrasing the sentences in which they think there are language points or by translating them, mechanical translation exercises or rote practice). Such a finding illustrating the disjunction between teachers' stated aims and their classroom practice coincides with Brown (1994) and Wall (1999). A similar example cited by Wall is that although teachers claimed that they had changed their way of teaching once they began using the new textbooks, and that they were then using a 'communicative approach', their classrooms were in fact very formal and they spent much of their time teaching formal aspects of language rather than developing the skills that the textbook promoted.

In light of such a discrepancy, Brown (1994) warns us that there are certain caveats in the field of language teaching when it comes to discussing CLT. According to him, some people simply give "lip service" to the principles of CLT without actually understanding the characteristics of the theory.

The disjunctions between teachers' beliefs and their statements about beliefs, and between teachers' articulated beliefs and their behaviours, have often been noted and referred to not only in the language testing literature (Cheng, 1999; Wall, 1999; Watanabe, 2004b), but also in English language teaching (ELT) literature (Pajares, 1992; Richardson, 1996; Samuelowicz & Bain, 1992; 2001; Woods, 1996).

All of these researchers were aware of the complexities and contradictions between what people say and what they practice, and they respectively pinpointed such a discrepancy. For instance, Alderson (2001) analyzed the reason for such a mismatch, saying that much of teacher-thinking is vague, muddled, rationalized, prejudiced, or simply uninformed (Alderson, 2001, p 3). From a different perspective, Woods (1996) pointed out that individuals can carry out actions which seem to be inconsistent with what they say their beliefs are. Similar to Woods' point (1996), other researchers

cautioned us that there may be a mismatch between teachers' awareness or judgment of their beliefs and their real beliefs.

Drawing on their respective research, the above-mentioned researchers provided convincing explanations for these mismatch(es). Based on Ernest's (1989) analysis, there are three key causes for the mismatch. First, if teachers' espoused beliefs are not integrated with other knowledge and beliefs, especially pedagogical knowledge, only a limited basis for their enactment exists. Second, a mismatch may exist between teachers' awareness or judgment of his or her beliefs and his or her real beliefs. Third, the mismatch may result from the powerful influence of the social context (e.g., the expectations of colleagues and superiors). A similar thread of comment was made by Woods (1996, 2003). He posited that a teacher's BAK and behaviour do not necessarily correspond. The reason for the discrepancy, based on his explanation, is that there are two types of knowledge – declarative knowledge (factual knowledge) and procedural knowledge (action-related knowledge). The latter was interpreted by him as a tacit and unconscious knowledge. According to him, it is this knowledge that guides our actions. In his explanation, beliefs may not always be entirely consciously accessible to teachers. What he implied is that what teachers say they believe may not always be the factor which influences their actions, which may stem from some patterns they are unable to make explicit. Another reason, according to him, is that when talking about generalized beliefs, teachers may answer according to what they would like to believe, or what they would like the audience to think they believe. From Woods' (1996) perspective, in such a subordinate power relationship with supervisors, evaluators, theorists, and researchers, teachers may prefer to claim "allegiance to beliefs consistent with what they perceive as the current teaching paradigm rather than consistent with their unmonitored beliefs and their behaviours in class" (p. 71). Woods' (1996) explanations are similar in part to the ones provided by Ernest (1989).

Along these lines, Samuelowicz and Bain (1992, 2001) asserted that the teachers based their teaching practices on the explicit or implicit theories they held about teaching and learning. He posited that they might have both 'ideal' and 'working' conceptions of teaching. The findings they obtained seem to suggest that the aims of teaching expressed by academic teachers coincide with the 'ideal' conception of teaching, whereas their teaching practices, including assessment, reflect their working conception of teaching.

The viewpoints made by these researchers shed light on the findings of this study. Through analysis of the data, three reasons were found to possibly relate to the mismatch between the teachers' claims of what they do and what they actually do in the classroom. The first reason may associate with the dynamic evolution of a teacher's BAK over time. Due to the evolving nature of a teacher's BAK, certain changes were traced during the process of research in the teachers' beliefs and perceptions of the CET, teaching and learning. A case in point is that the attitude T6 maintained toward the CET was getting less negative when we compare his statements at the group interview to those at the individual interviews. It is possible that his beliefs about the test have been refined and thus have evolved during the research process. Another example is that relatively more changes were observed with T4 and T5 in terms of the activities they organized in the classroom. At one interview, both of them admitted that they had given little thought to matters of teaching methods in the past. However, during the last individual interview, they both took an initiative to share, to reflect on their experiences as well as their various perceptions. The viewpoints they articulated at the last interview demonstrate evidence of internal cognitive shifts in their beliefs and philosophies about teaching, learning and testing.

The second possible reason, which has been discussed earlier, is that despite teachers' preference for concepts such as CLT and learner-centeredness, they lack the hands-on knowledge to implement or incorporate them in the classroom. The paradoxes or contradictions of teachers' belief systems as well as discrepancies between their articulated beliefs and their behaviors seem to suggest that they knew what they were expected to do, but were at a loss at how to make corresponding changes.

The third reason may relate to the Chinese context. As portrayed in Chapter 3, the Chinese educational context is centralized, exam-led, and knowledge-focused. In such a context, teachers may give their opinions according to what they would like others to think they believe. Especially on occasions such as group interviews before the presence of their colleagues and superiors, they would prefer to say something which is consistent to the mainstream beliefs about teaching or pronounce the sort of beliefs consistent with their 'ideal' conceptions of teaching rather than working conditions of teaching. It is possibly due to this reason that the views aired by the teachers differ from occasion to occasion.

One implication from the above finding is that we need to take an objective view of

teachers' claims. On the one hand, we must be aware that being covert and invisible, teachers' beliefs can only be inferred from their actions rather than from their statements or claims. On the other hand, we need to triangulate the data to confirm their claims.

Another implication is that we should provide teachers with more opportunities to reflect on their teaching and recount their dilemmas, for the process of reflection is considered to be a crucial step for developing their expertise. Based on Woods (1996), it is the dilemmas that cause teachers to analyze and reflect on their beliefs, and to consider the various options for achieving their teaching goals. In this regard, rather than focusing on the negative impact of the discrepancies and mismatches, we need to view them as an indispensable step for teacher change or teacher professional development.

### *6.3.2 Different Aspects of the Teacher Factor and Their Interrelationships*

The theme discussed above associates with teacher and school variability in the face of washback. In addition to the different amounts and types of washback effects manifested in different school settings and different cases, teacher differences also exhibit in teachers' conceptions of teaching and learning, their interpretations of classroom behaviors in terms of the activities, teaching styles and methods adopted by them, their knowledge of contemporary ELT theories, their proficiency, and the kind and amount of training they have received. The data from case studies indicate that these conceptions and interpretations seem to play a fundamental role in teachers' decision about the way they teach. The survey data further illustrate that the 'teacher factor' is a multidimensional construct composed of such aspects as teachers' various beliefs, knowledge and experience. They also indicate that close interrelationships exist among these aspects, which enable them to form a network to affect teacher performance in the classroom.

In view of these data, the themes to be elucidated in the following sections concern the 'teacher factor' under study. As limitations of space prevent a full description of each aspect in a thorough way, in this section I will first highlight three aspects which are found to play a significant role in how teachers conduct their lessons: 1) teacher beliefs about teaching and learning; 2) teacher knowledge; 3) teacher English proficiency. Then, I will discuss the interrelationships that exist among various aspects of the 'teacher factor'.

### 6.3.2.1 *Teacher Views on Teaching and Learning*

An examination of Model 3 revealed that teacher beliefs can have a significant, direct, positive effect on the way they structure their classroom events. The findings seem to make sense, since other data sources also show that the way teachers approach teaching correlates with their beliefs about teaching and learning. While the survey questions may help us learn about whether or not the participants hold a transmission-based view of teaching and learning, they cannot delineate an accurate picture of these beliefs, such as their connotations and their embodiment in the classroom. To get an overall understanding of the multifaceted complexities and realities intertwined in the process of teaching, in what follows, the survey data should be complemented by evidence from case studies.

As presented and summarized in the previous chapter, the results from observations show that similar to the mainstream classroom practice in the Chinese EFL context (which has been discussed in Chapter 3), the lessons conducted by most of the participants in this study are also characterized as being knowledge-oriented (e.g., focus on formal linguistic knowledge such as sentence patterns, prescriptive grammar rules, vocabulary items, and background information concerning texts in course-books, reliance on the structural or grammar-based approach, etc.), text-oriented (e.g., teaching done through explanation of texts in course-books, accompanied by a meticulous analysis of language points such as grammar and vocabulary in a decontextualized manner, paraphrases of sentences which they deem as difficult and extensive translation either at sentence level or discourse level), and teacher-centred (e.g., the format of teaching is teacher lecturing and presenting to the whole class). Their teaching styles are didactic rather than interactive; deductive rather than inductive. Activities that best represent features of CLT such as creation of a context that facilitates learning (Littlewood, 1981), the provision of opportunities for students to use the language to interact and to share real information (Wesche & Skehan, 2002), learner communicative involvement in the negotiation of meaning (Savignon, 1991) and learner choice and autonomy (Candlin, 1978) were hardly noted in their classes. These results coincide with those from the survey, which reflect that a little more than half of the teachers (52 %,  $M=2.41$ ) maintained that learning a language is a process of accumulating knowledge. They can be taken to mean that the beliefs about teaching and learning held by the majority of participants are still transmission-centered.



Both their responses given on the survey and the ideas the participants articulated at the interviews provide insight into why their practice in the classroom displayed an explicit focus on knowledge and why rote practice was utilized.

When asked what their rationale was for giving special attention to language forms, a few participants touched upon the role of 'language input' in their discussions at the interviews. It seemed that underlying their focus on form are their conceptions of language input. Below are the rationales provided by the participants for such a focus.

For instance, in reflecting on why he asked his students to recite clips of some passages, T6 recounted:

*It is impossible for students to express their ideas freely. Thus, I assigned them 42 topics and in each class, each student is supposed to be able to recite 3 paragraphs of these passages. I think input is more important. Without input, it will not be possible to have output. If we want to improve the students' listening skills, we need to develop their overall skills. Since the skills are integrated, it is hard for us to enhance their listening when they have trouble reading.*

T5 also provided her rationale for her emphasis on language points in class. She said:

*Language is a means of communication. If a student wants to speak, he needs to have a good command of the language. Otherwise what do we expect him or her to talk about. So, language use should take place under the condition that there is some real content. Thus, we should give them more input.*

Similarly, to justify his practice, T1 provided an interesting comment:

*The reason why students often go out of their ways to memorize vocabulary is that their problem with it has never really been tackled. So, they are in need of further input in it.*

The rationale provided by T3 serves as yet another example of her emphasis on why she placed the most value on language knowledge:

*Seldom have I thought of such issues as teaching methods. I simply teach using my own way of teaching. Personally I think that teaching priority should be given to developing students' abilities in reading, because I find the biggest barrier the students have while reading is that their vocabularies are limited. Consequently, they have difficulty understanding the passages they read, and furthermore they cannot use even the simplest expressions when they are writing compositions. Therefore, linguistic knowledge should further be stressed.*

It is apparent that with respect to input, the views taken by the participants are extraordinarily similar. These same patterns of interpretation can be traced in the responses of the other teachers as well (e.g., the group interview participants). In general, the teachers thought that at their stage of CE study, what the students should do was to build up a good foundation in English and therefore, language forms were fundamental to language teaching and learning. The results reflect that contrary to the constructivist principle of teaching and learning, students are considered by the teachers to be receiving language knowledge in class rather than constructing it. To most of the participants, teachers are still at the core of the teaching process.

In order to further defend why they teach the way they do, T3 and T6 articulated some interesting beliefs. See the following excerpt from T3's remarks.

*The students are used to listening to structural analysis, and they find it hard to change their traditional way of learning after they enter college. Each text consists of a large number of new words. If we do not explain them, the students do not know how to use them.*

Here is another excerpt taken from T6's comments.

*Each teacher has his/her own way of teaching. As long as students' interest in learning is evoked, the method counts as a good one.*

The above statements and comments made by the participants not only threw light on the meanings teachers attach to their beliefs about teaching and learning, but also provided us with their professional stances on teaching and how to teach.

An in-depth analysis of the results indicate that although there is nothing wrong with emphasizing the need of rich input or adequate exposure to the target language, the teachers seemed to ignore what sort of input is needed for a learner and how input becomes intake (or uptake) or how the new knowledge they have imparted to the students can transform into their own.

Above all, two other important variables that promote SLA – interaction and output – seemed to have been overlooked by the teachers as well. According to Long (1985, 1996), the most effective way in which input is made comprehensible is through interactional adjustments. He postulated that comprehensible input gained through interactional adjustments such as negotiating meaning and modifying output is central to SLA. The participants, however, were unaware of the role of output emphasized by SLA theorists and researchers (Long, 1985, 1996; Gass & Selinker, 2001, Swain, 1995). Such a finding is corroborated by the results reported in Section

5.5.3.3.2 which indicate that teacher knowledge exhibits a direct, positive effect on the way teachers teach.

Given the profound impact that knowledge produces on teacher practice, in what follows, the issue will be examined more closely.

#### 6.3.2.2 *Teachers' Knowledge*

Contrary to the findings by other researchers (Anderson, 1993; Burnaby & Sun 1989; Campbell & Zhao, 1997; Cao, 1999; Gattbonton & Gu, 1994; Han, 1999; Keith & Zhao, 1993; Rao, 2001) which reported on Chinese EFL teachers' skepticism about the feasibility, practicality and effectiveness of employing CLT in their classrooms as well as their resistance to CLT, the survey results of this study show that more than half of the teachers (54%) favor CLT over the structural approach. Also, unlike earlier research findings which claimed that EFL teachers were unclear of or did not have knowledge about how to implement the new curriculum or how to teach in their respective settings (Chapman & Snyder, 2000; Chen, 2002), the results from this study suggest that the teachers had an increasing awareness as well as upgraded beliefs about teaching and learning. As evidenced by the survey data, there are a number of teachers (48%) who no longer relied on a transmission-based view of language learning in their instructional practice. The data also show that 56% of teachers have a better understanding of their ultimate goal of teaching as well as the importance of language use. Meanwhile, as demonstrated in their responses to Q34, the majority of them (68%) claimed that they favor activities such as role-play, language games, information gap, and group work or pair work, which shows that they have become increasingly aware of the importance of language use as well as the importance of teaching communicatively.

However, through interviews with my participants, I find that the teachers still have pedagogical misconceptions (e.g., reliance on input and negligence of output as mentioned above) or confusions about how to go about things in the classroom. Some of them seemed to have only acquired a superficial or surface understanding of the theory of CLT. For example, according to T3, communicative competence is nothing but a method simply emphasizing verbal ability while ignoring other skills. These same patterns of misinterpretation are found in many of the responses of the other teachers as well. Another case in point is that some participants and interviewees interpreted the concept literally and viewed it as failing to incorporate grammatical

competence. Supporting evidence is found in the survey data, as well. The teachers' responses to Q43 indicate that as many as 54 % of the respondents had the same misconception. These results suggest that despite the favorable attitudes teachers showed toward CLT and their willingness to incorporate aspects of the methodology into their classrooms, their knowledge of the theory is inadequate. Such a finding overlaps that of Shu (2004), for she also discovered that many teachers hold the view that 'communicative ability refers to spoken ability'. The finding suggests that there is a wide gap between Chinese EFL teachers' interpretations of CLT principles and those set forth by ELT researchers over the years. Consistent with this view, Hu (2002) pointed out that due to lack of exposure to CLT, many teachers had little idea on how to apply its principles.

In light of the above finding, one possible reason for the teachers' adherence to the traditional teaching approach is that in spite of their growing methodological awareness of the ELT or SLA theories, they still lacked an adequate understanding of the principles and strategies underpinned by these theories as well as the hands-on ability to implement or promote them in their classrooms. The lack of uptake by the teachers of these theories may be the key reason leading to the discrepancies and contradictions discussed in 6.2.4. Based on this finding, it is possible to make the inference that only when teachers achieve a solid knowledge base will they take steps to innovate their methodologies or teaching approaches. The above assertions were reinforced by Davies and Pearse (2000) who accounted for teachers' adherence to the Grammar-translation Method saying that it relates to their lack of knowledge and ability to apply CLT. Thus, an assumption can be made that an adequate understanding of the ELT principles and theories is fundamental to its successful implementation. Empirical evidence for the assumption will be further discussed in Section 6.3.2.4.

In the EFL contexts, as the classroom is most likely the only place where students can develop their communicative competence, the teaching approaches adopted by teachers seem vitally important. Nevertheless, in view of the close relationship between teaching approaches and teacher knowledge base as well as the inadequacy shown in the teachers' pedagogical knowledge, it appears that they need to undergo training to learn more about theories on SLA, sociolinguistics, psycholinguistics, language testing and other disciplines that can guide their teaching practice. The importance of teacher training has been highlighted by Borg (2006) and Richardson

(1996), who make the assertion that teacher knowledge about teaching may be influenced by experience with schooling and instruction, and experience with formal and pedagogical knowledge.

With such training, they may also arrive at a better understanding of their own beliefs and may ensure congruence in different sets of beliefs as well as congruence between their beliefs and their practices.

Furthermore, they must acquire subject matter knowledge, as suggested by Reagan and Osborn (2002) and Shulman (1987), in order to teach more effectively. Meanwhile, as EFL teachers, they must develop a high degree of language competence in terms of the four language skills as well as an awareness of sociocultural aspects of language and language use. Based on Shulman (1987), this content knowledge is crucial, for it provides an essential foundation for the teacher's pedagogical knowledge and skills for teaching English.

In addition to the inadequacy which lies in the teachers' knowledge base, another problem identified in this study is teacher proficiency. The issue will be dealt with below.

#### 6.3.2.3 *Teacher Proficiency*

As described in the preceding chapter, during classroom observations, most of the participants were found to be frequently using Chinese, the students' native language. By observation, the percentage of their classroom instruction conducted in English ranges from as little as 10% to as much as 50% or more, with an average use of below 40% (see the 'x' mark in Table 50). This is in sharp contrast to the percentage each of them checked in his or her response in the survey, which is much lower than that observed (see the '√' mark). This can be taken as an additional instance showing the disjunction between what they articulated that they did and what they actually did in the classroom.

Table 50 *Comparison between What the Participants Claim (x) and What They Do (✓) in terms of Their Medium of Instruction*

Participants	Category					
	0~10%	11~20%	21~30%	31~40%	41~50%	51 or more
T1	x	✓				
T2					x	✓
T3			x	✓		
T4					x	✓
T5						x✓
T6						x✓

With respect to the limited amount of English teachers used in the classroom, students could hardly have sufficient exposure to the language in the instructional contexts. Since CE class is the sole occasion for the learners to use English, such a feature of instruction is called into question and merits great attention from authorities, schools and teachers at large.

These results have provided evidence to support the contention that the teachers' proficiency level is the key problem in implementing instructional innovations. Such a point coincides with the evidence provided by Chen (1999), Gu (2005), Xia (2002) and Shu (2004), which contended that the central problem to the CE teaching in China lies in the shortage of qualified teachers. Shu (2004) pinpointed the consequence of such a problem, claiming that classes delivered by these teachers cannot provide learners with enough input. The interview data in this study lend support to the above contention by providing added evidence that the teachers are not proficient in their oral skills. From this finding, we can make an inference about why the teachers were seldom seen incorporating communicative activities in their classes. The possible reason is that due to their low proficiency, they may find it hard to deal with the spontaneous questions that the dynamic nature of communicative activities requires them to answer.

In sum, the interviews conducted in this study generated serious discussions about the issues related to the teacher factor. Apart from teacher beliefs, knowledge base, and proficiency, teacher commitments, training (or guidance) and habits were also deemed by the teachers as important issues affecting their instructional practice. However, the pursuit of these issues is well beyond the scope of this study. It is possible that these issues will be dealt with in another paper.

The next section will revisit the most significant and noteworthy result discussed in

the preceding chapter – the interrelationships of the ‘teacher factor’.

#### 6.3.2.4 *Teacher Beliefs, Knowledge, Experience and Their Interrelationships*

Inferential analyses of the data (e.g., EFAs, CFAs, and SEM) presented in Section 5.5.3 exhibit some interesting findings. One is that the ‘teacher factor’ is a multidimensional construct consisting of a set of interrelated components such as teacher beliefs, knowledge base and experience (see Model 2). A close examination of the construct and its relationships to teacher practice as well as teacher beliefs about test impact shows that the construct, as a whole, significantly relates to both teachers’ classroom behaviors and their assumptions of test impact on teaching. Additional interesting findings are that each component of the construct contributes differentially to the way that teachers teach and the manner in which they perceive test impact. As elucidated in the previous couple of sections and demonstrated in Model 3, teacher beliefs and knowledge of the methodologies were proven to count far more than all the other factors in the extent to which they implement such methodologies in their instruction and thus affecting the way that the teachers deliver their lessons. In addition to its direct causal relationship to teacher performance, teacher knowledge of pedagogy was also found to be significantly related to teacher assumptions of CET impact on teaching. Its critical role was manifested in Model 4. Both models (1.3 and 1.4) reveal that teacher knowledge of pedagogy plays a substantially greater role in how teachers perceive test impact and how they approach teaching than do any other components of the construct.

These findings are significant and noteworthy in the following ways. First, they provide empirical evidence in support of Samuelowicz and Bain (1992, 2001) in their claim that teacher practices significantly correlate to their theories of teaching and learning, and they also provide support for Chapman and Snyder (2000) and others in that teacher behaviours may correspond to their knowledge of teaching pedagogy. In the meantime, it appears that the finding helps to account for the first theme discussed in this chapter – why there is disjunction between the washback rhetoric and classroom reality. In light of this finding, a possible reason for the failure of tests to function as a means of innovation is that the teachers do not have explicit beliefs about teaching and learning and necessary pedagogical skills to meet new demands of the innovation. In addition, the finding seems to shed new light on the incongruence, paradoxes or discrepancies that lie in teachers’ claims and behaviors discussed in

6.2.4 and 6.3.2. For instance, it was mentioned earlier that even though the majority of participants showed favorable attitudes toward communication-oriented EFL lessons, communicative involvement was rarely observable in their classrooms. Drawing on this finding, despite their preference and willingness to carry out communicative activities, there might be certain pedagogical factors that restricted the teachers when it came to implementing them.

Second, the findings substantiated Woods (1996, 2001) in that teacher knowledge base, experience and beliefs, though interrelated and interwoven in a teacher's BAK network, may relate to one another in complex and non-linear ways. Due to the interrelated and interactive nature of the network, although a "one-way" relationship was not established between other factors (e.g., experience and other beliefs) and teacher performance in the classroom, their role should not be overlooked. Each of them may contribute, in an indirect or non-linear way, to how teachers conduct their lessons. In the meantime, the findings seem to bolster Woods' (1996, 2001) assertion that the network plays an important role not only in teachers' interpretation of their beliefs and behaviors, but also in their organization of thoughts, decisions and actions.

Third, some findings of the current study coincide with those of the earlier research. For example, beliefs about testing tend to follow beliefs about teaching and learning (Glaser & Silver, 1994) and positive washback effect is seen to occur with teachers who maintain a positive attitude towards their teaching (Cheng, 1999). Similar to these assumptions, significant positive correlations are shown between BI and BTL on the one hand; and between the behavioral dimension of washback in terms of the test-related activities teachers conduct in the classroom and BTL on the other.

Fourth, the establishment of direct causal relationships between teacher pedagogical knowledge and teacher assumptions of test impact on teaching, and between teacher pedagogical knowledge and teacher practice may imply that the factor functions as a crucial condition for instructional change. The finding and those illustrated above seem to suggest that to facilitate meaningful perceptual and behavioural change on the part of teachers, other factors may need to integrate with this one. At the same time, the finding indicating that the way teachers interpreted the impact of the CET on teaching correlates to their knowledge base also seems to contribute to our understanding of another theme outlined in this chapter – negative views on tests and innovation. As mentioned above, one possible reason leading to the



teachers' prevalent negative attitudes toward the CET and innovation (teachers' accounts of the constraint of the CET) is that they may not have been equipped with adequate knowledge of teaching pedagogy to carry out the testing reform as well as the instructional innovation. The finding is not only consistent with the data from case studies of this research, but is also in accord with those documented in the literature, for they assumed that teacher knowledge and expertise appear to be the major influences in teachers' beliefs about washback.

In view of the above findings, a conclusion that can be drawn is that the network of the 'teacher factor' underlies the washback phenomenon.

## 6.4 Other Factors in the Washback Phenomenon

### 6.4.1 *Learner Factor*

Another notable theme that has emerged from data analysis is that the 'learner factor' is considered by teachers to play an essential role in their decisions concerning change. There are numerous examples in the data illustrating this theme. In general, teachers blame tests, learner factor, large class, etc., some of which have been manifested and indicated above. Here, I only probe a bit more into the learner factor. As recounted by the participants of this study, in addition to the impact of the CET, the 'learner factor' also constitutes a constraint on their teaching. Based on their accounts, the factor chiefly involves: 1) learner beliefs, 2) learner proficiency, 3) learner habits, etc.

#### *1) Learner Beliefs*

Both the group and individual interview data indicate that learner beliefs have a part to play in teacher beliefs. It has been mentioned that during the interviews, teachers repeatedly talked about their dilemma as to how to teach. The following excerpt from T4 illustrates her dilemma.

*Sometimes I intended to lead my students to perform some speaking activities. To my disappointment, the students had little interest in them and later after class they even reminded me that the CET does not measure their ability in speaking. They expressed that they would expect me to cover more knowledge of vocabulary in class.*

Based on her account in the above excerpt, in spite of her personal desire to undertake activities to enhance students' skills in spoken English, under the pressure from her students, she had no choice but to teach the way she was expected to.

Other participants expressed similar dilemma(s). When discussing the relationship between testing and teaching, T2 convinced us that it was her students rather than herself who deemed it to be a conflict. Likewise, T6 blamed his students for being so utilitarian and clinging to their deeply-rooted views about tests. He proceeded to remind us that students vary from school to school. The following excerpt is taken from his statements at one of the interviews.

*Students of different schools have totally different learning objectives. Both tests and our teaching methods should be tailored according to the levels as well as needs of our students.*

To probe into the implication of the above data, an inference that can be drawn is that Chinese EFL learners seemed to have negative attitudes toward instructional innovation or change and these attitudes not only determined their passive behaviors in the classroom, but also affected their teachers' perceptions and beliefs about teaching and testing. Also drawn from these results is the inference that to ensure a change in teacher practice, a change has to be brought about in the learners' beliefs. Before concluding this discussion on learner beliefs, however, it is important to note that the teachers seemed to be unaware of the dynamic nature of change. That is, it takes time for students to get used to a new method. If they were persistent enough, their students might understand the benefits the new method would bring to them and might take the initiative in changing their beliefs and practices.

## 2) *Learner Proficiency*

In addition, learner proficiency was also regarded by the participants as a hindrance to their instructional change. On more than one occasion, T2, T5 and T6 expressed their concern about the weaker students. In T2's words, it was impractical to leave the floor to her students in class. Below is an explanation from T2:

*Once I asked several students to give presentations in front of their class. Even though they had made some sort of preparations beforehand, other students could not understand them because of their accent as well as their low voice.*

This excerpt suggests that learner proficiency also plays a determining role in what teachers do in the classroom.

## 3) *Learner Habits*

Another aspect of the learner impact on teachers' instructional patterns is learner habit. One habit cited and complained about by the participants was 'learner

dependence on teachers'. In their opinion, Chinese CE learners are too teacher-dependent. The reason is that they have been so accustomed to their role as passive listeners in class that they only enjoy sitting in the classrooms absorbing knowledge.

When I asked the participants to comment retrospectively on the teacher-centered activities carried out by them in their instruction, they argued that the activities were utilized to correspond to the learning habits of their students. The following comments articulated by T1 aided our understanding of their point(s).

*Unlike learners in western countries, Chinese learners are teacher-dependent. They prefer to be led or guided by their teachers, when they learn something new. If I do not analyze the text and provide them with all the explanations, they will end up learning nothing.*

Another learner habit that the participants talked about was the students' tendency to be test-oriented. This tendency was addressed in 'learner beliefs' as well.

Taken together, the above evidence seems to coincide with other evidence described in the washback literature (McNamara, 2000). McNamara also claims that whether or not positive washback is achieved will depend on local conditions in classrooms, the established traditions of teaching, the immediate motivation of learners, etc. (2000). In this regard, when examining washback, more research attention should be directed to the 'learner factor' as well.

While some of the sources of blame listed above are not as ungrounded as those found by Watanabe in his study, they are considered secondary and less decisive when compared to the role that internal factors such as teacher beliefs, knowledge and proficiency play in their classroom practice. Such a view was buttressed by Alderson (2001). According to him, there are many factors that affect the impact a test will have, and how it will be used, misused and abused. But he reminded us that it was the 'teacher factor' that plays a determining role in ensuring the success of a testing innovation. Although the agenda of the 'learner factor' as well as other factors also plays a part in washback, it is not as crucial and decisive as the teacher variable and thus is not of central concern in this study.

The results discussed above have also provide evidence to support the contention by Watanabe (2004b) and Chapman and Snyder (2000) that teachers tend to place undue blame on the presence of examinations for what they are doing. A close examination of the data in this study shows that the attribution theory they drew on in

interpreting teachers' tendency of attribution also holds true in my study. Based on the attribution theory (Willaims & Burden, 1997), most people tend to externalize reasons for failure while internalize reasons for success.

### 6.5 Principal Findings

In summary, this study has yielded a number of interesting findings related to the 'teacher factor' in washback. One significant finding is that teachers' integrated BKE network, and particularly their pedagogical knowledge and beliefs about teaching and learning, exerts a powerful influence in the way they interpret and react to washback as well as the extent to which testing and instructional innovations take hold. This finding supports the conclusions or statements made by other researchers that teacher variables make a difference in how washback or other educational innovations unfold (Alderson, 2001; Andrews, 2004; Burrows, 2004; Cheng & Curtis, 2004; Wall & Alderson, 1993; Greene, 2006, 2007, Turner, 2008, 2009).

In light of the causal relationship between teachers' BKE and teacher practice and case-study data in this study, without transforming teacher conceptions from a transmission view to a constructivist view (Samuelowicz & Bain, 1992; 2001; Winitzky and Kauchak, 1997) or experiential view of teaching and learning (Nunan, 1999; 2001) as well as upgrading teacher knowledge which can be provided in training (Turner, 2009), teacher behavioral change will not be a possibility.

In addition, the 'teacher factor' is an integrated concept, for various aspects of teachers' BKE are inter-correlated. The data confirms Woods' (1996) assumption of teachers' interconnected BAK networks.

Also, considering the essential role that pedagogical knowledge plays in determining teachers' claims and performances, it appears that a fundamental change in teacher classroom behaviors would necessitate teacher pedagogical knowledge.

The second interesting finding is that contradictions, incompatibilities and mismatches exist not only between teachers' beliefs and their articulated beliefs, and between what they articulated on one occasion and what they expressed on another, but also between what they articulated that they did and what they actually did. As discussed above, these discrepancies may be ascribed to the evolving nature of teachers' BAK networks over time, the Chinese context, and the possibility of lack of knowledge of teaching pedagogy or inaccessibility of their beliefs to teachers. According to Richards (2008), the dimension of teacher pedagogical knowledge is

often tacit, experiential (gained through experience) and contextual. Thus, a suggestion that can be made is that guidance be provided to teachers to make this knowledge explicit.

The mismatches may also be attributed to, as pinpointed by Ernest (1989), the fact that teachers' espoused beliefs are not integrated with other aspects of knowledge and beliefs, especially pedagogical knowledge.

The third notable finding is that the way teachers perceive and react to examination change not only varies from individual to individual, but also differs from school to school. This finding has been sufficiently exemplified in the study. Such a finding overlaps with that of earlier research, which also argues teacher as well as context variability in washback. There are two possible reasons for the differences. One is that each teacher has an individual system of interwoven BKE that underlies everything that he or she does and says, which differs from that of others. Another is that teacher beliefs and behaviors are both cultural-specific and context-dependent. In contexts where there is ongoing training, and constant guidance provided to teachers over time, as evidenced in studies by Davison (2008), Muñoz and Álvarez (2010), Tavares and Hamp-Lyons (2008), Turner (2009), and Urmston and Fang (2008), positive washback seems to be an attainable goal. However, in contexts where little or limited teacher support is provided, positive washback may not be expected.

Thus, the shape washback takes has to do with the socio-cultural environment and school setting where washback is nurtured.

What I want to mention is that it is this context-bound feature that determines the way I presented the results from case studies.

Another conspicuous finding is that tests neither constitute a facilitator nor pose a constraint on instructional innovation. Both the Chinese experience and international experience discussed in this study show that the impact of tests was often intensified and misinterpreted. On the one hand, the extent and power of the washback effects may possibly be overestimated by washback proponents. In view of the interrelationships of the 'teacher factor' and its unique relationship to teacher practice as well as the above-mentioned pedagogical, social and personal complexities embedded in the washback phenomenon, testing reforms can only address the symptoms of the existing problem(s) (e.g., content of the teaching) with the core of the problem(s) – deeply-seated teacher beliefs, teacher pedagogical knowledge and teacher practice – left unaffected (at least in the short term). As these critical change

agents remain unaffected, testing innovation on its own may not be an effective mechanism for improving teachers' instructional practice, unless adequate training is provided to teachers on assessment and pedagogy. Nor might it fulfill its intended goal of effecting teacher conceptual and behavioral change, for the process of teacher evolution in terms of beliefs, knowledge and behaviors takes time. As elucidated by Vygotsky (1978) and Woods (1996, 2003), it involves a process of internalization and a "lag time" for cognitive changes to take place. Given the absence of this process, it came as no surprise that in this study, the intended innovation effects did not appear to manifest themselves.

On the other hand, the negative impact of the CET appears to have been exaggerated by the Chinese CE teachers and opponents of the CET. Although it appears that the impact of tests in China is so devastating that teachers are compelled to teach to the tests, actually no true linkage can be established between the two variables. Substantial evidence from this study shows that what teachers do in the classroom is not as closely related to the presence of tests as assumed.

Overall, the findings of this study corroborate the results from other EFL contexts, which report that the attempt and effort to introduce positive washback simply by changing the exam format or contents will not necessarily produce an expected outcome (Andrews et al., 2002; Chapman & Snyder, 2000; Cheng, 1997, 1998; Cheng & Curtis, 2004; Qi, 2004, 2007; Wall & Alderson, 1993; Wall, 1999; Shohamy, 1993; Shohamy et.,1996; Watanabe, 1996a, 2004b). The salient themes that have emerged from this study are consistent with those of Hong Kong (Cheng, 1997; Cheng, 1998), Uganda (Chapman & Snyder, 2000), and Sri Lanka (Wall and Alderson, 1993). The findings of the study lend support to Chapman and Snyder's (2000) assertion that a test has far fewer mechanisms for influencing and shaping individual teachers' pedagogical practice than most policy-makers wish.

#### 6.6 Further Analysis of the Findings – Washback Nested in Complex Webs of Factors

Why has the undertaking of inducing positive washback always been such a difficult one? The findings of this study seem to partially answer this question. An in-depth analysis of the complete data set of this research illustrates that the complexity of the undertaking not only lies in the construct of the test and its

connection with the intricate relationship between the ‘teacher factor’ and teacher practice(s), but also resides in the involvement of the complex interplay of a large number of factors both within and beyond the school throughout the process of implementing instructional change. To put it another way, like other innovations, promoting washback is entangled in complex webs of factors (e.g., contextual factors and various components of the internal ‘teacher factor’ – teacher beliefs about teaching, learning, testing and how to teach, teacher knowledge, teacher proficiency, experience, etc.) and multilayered and intertwined relationships among them. Since it involves changing beliefs, knowledge and behaviors on the part of teachers, there must be changes fostered both in external environment and internal conditions to ensure them to happen. A change in beliefs is itself no easy task. Such a contention is consistent with that of Woods (1996, 2003). Both Woods (1996, 2003) and other L2 researchers maintain that not only are teacher beliefs intertwined with other variables of the ‘teacher factor’ such as knowledge and experience, which interact dynamically in all phases of the language teaching process, but they also relate to action in complex and indirect ways. It might be for this reason that their evolution is a gradual, complex process. Furthermore, a change in deeply-rooted beliefs is more intricate and demanding. It might also be for this reason, as Andrews et al. (2002) concluded, that the washback effect is often delayed rather than immediate.

In view of the above complexities underlying the washback phenomenon, the idea of using tests as a strategy to drive teaching seems to oversimplify our understanding of the process of instructional change. Such an insight into the intertwined and intricate nature of washback seems to cast light on a number of findings or themes discussed in this chapter.

### *Summary*

The findings of this study show that tests constitute neither a facilitator nor a constraint on instructional innovation. The idea of using tests as a strategy to drive teaching may oversimplify our understanding of the process of teacher conceptual and instructional change. These findings show that the type of washback in existence and the characteristics of the ‘teacher factor’ displayed in innovation in China are similar to those of many cultures. Given some of the similarities and congruence in findings in studies conducted in different EFL contexts as well as in different research areas, there seem to be some shared meanings that permeate research on teacher beliefs and

practice. In light of this, the problems involved in fostering teacher instructional change (as discussed in Chapter 6) are neither unique to the testing field, nor particular to the Chinese EFL context. In this regard, although my study focuses on the educational context and testing culture of China, its theoretical and practical significance goes well beyond the Chinese context and testing innovation.

Considering the complexities intertwined in the phenomenon of washback, we should neither overestimate nor underestimate the power of its effect, because either extreme may lead to negative results.

Implications from the above findings are that researchers need to take an objective view of teachers' claims. Furthermore, we need to triangulate the data to confirm these claims. Meanwhile teachers' accounts of the impact of tests can be examined in light of the attribution theory so that we might be in a better position to help teachers to change their tendency to attribute examinations, from being seen as "external" and "uncontrollable" factors to being seen as "internal" and "controllable" factors. Considering such research results, authorities need to reexamine their belief that a test is an effective tool to invoke teaching and learning. Other efforts need to be made to ensure the successful implementation of the reforms.

As a result of the above data analysis, I realize that there are at least four features of change washback studies must consider.

First we need to note that instructional change is based on the premise that teachers are guided by advanced theories of teaching and learning and equipped with adequate knowledge of teaching pedagogy. Second, we need to be aware of the evolving nature of innovation as well as the dynamic aspects of teachers' BKEs as well as the interwoven characteristic of teachers' BKEs in the process of instructional change. Third, from a dynamic perspective, washback, as a power-coercive strategy, may not be compatible to the natural evolutionary processes of teacher change, especially the change in teachers' BKEs. Fourth, we need to change our traditional views of discrepancies. As posited by innovation researchers, rather than viewing conflicts, differences and discrepancies as forces that hinder the processes of change, they should be perceived as an impetus for change and an indispensable part of teacher evolution.



---

## Chapter 7: Conclusion

### 7.1 Introduction

This exploratory study examined various teacher beliefs, knowledge, proficiency and performances in the classroom. It also investigated the role that the ‘teacher factor’ plays in washback. This concluding chapter summarizes the major findings of this research. In order to limit the discussion to a more manageable degree, it reports on the findings in relation to issues of L2 teaching, learning, testing and teacher change.

It then discusses the theoretical and methodological implications of the study. Finally, it concludes by outlining the study’s strengths, limitations, strengths and contributions, and direction for future research.

As this study drew on the multidisciplinary conceptual framework developed from multiple sources, the results presented in this study have a number of implications for general educational research as well.

### 7.2 Major Findings

As outlined in Chapter 6, this study provides information about the nature and variability of washback. Owing to the individual teacher factors (or the ‘teacher factor’) and contextual factors underlying the washback phenomenon, the revised CET seems incapable of helping teachers initiate a deeper self awareness of teaching. Nor is it capable of effecting fundamental teacher conceptual and behavioral change. The results seem to suggest that seeking immediate change in teacher beliefs, knowledge, and behavior is impractical, which confirms other findings (Freeman & Johnson, 1998; Richards, 2008; Woods, 1996) that claim that there is often little immediate evidence for change in teachers’ practice(s). Drawing on the theoretical framework outlined above, the process of initiating teacher change is complex and developmental and it may develop and emerge over time through the process of teaching and training.

While there is ample evidence in the data showing that teachers’ resistance to change and their negative perceptions of the CET still prevail, there is also sufficient evidence indicating that the test neither constitutes a facilitator nor poses a constraint on instructional innovation.

There is also adequate evidence showing that the CET has had a different impact on

different teachers, since great variation was noted in their interpretations of and reactions to washback. In addition to individual variability in terms of the degree of washback, school variability was also noted during the research process. One instance is that the CET has exerted more influence on average schools than on key schools. The results reveal that while the participants appear to have interacted with institutional contexts (or context of work) in different ways, they seem to respond and react to washback in accordance with the rules and norms that prevail in their institutional contexts. Thus, a conclusion that can be drawn is that both personal teacher factors and institutional contextual factors are involved in the process of washback.

In addition to providing information about washback, this research has provided information about how the 'teacher factor' is manifested in the washback process. As mentioned above, the networked 'teacher factor' was found to underlie the washback phenomenon in terms of how teachers interpret test impact and implement reforms or change their practice. Nested within this interrelated network are attributes which include various teacher beliefs, pedagogical knowledge and professional training (or experience). One significant finding is that the interconnected attributes seem to form a core that threads itself through the process of teaching as well as the process of washback. The limited washback in this study might be attributed to the combined reasons of teachers' misconceptions of teaching and learning, inadequate knowledge, insufficient in-service training, and guidance for change.

Of the various aspects of the 'teacher factor', teacher pedagogical knowledge was found to affect teacher practice in a deterministic way. Consistent with the findings by Andrews (2003), Shulman (2000), and Turner (2008, 2009) that view teacher pedagogical knowledge as an essential component of teacher professionalism, this study has shown that this dimension of knowledge exerts a strong influence on how teachers interpret testing innovation and change their practice. This finding also reinforces Richards' (2008) assertion that such knowledge contributes to the formulation of teachers' working principles that guide their teaching behavior and functions as the source of teachers' practices. Another salient finding is that discrepancies, contradictions, and mismatches exist (1) between teachers' articulated beliefs and the realities; (2) between what they articulated on one occasion and what they expressed on another; (3) between what they articulated that they did and what they actually did. One typical example is that their claimed perceptions and

knowledge of CLT are in conflict with their applications and actual knowledge of it. A possible explanation for the conflict is that despite the participants' claimed familiarity with this theory, they might not have internalized the principles of the theory to influence their decision making or actions in the classroom situation.

The results from the SEM showed that in addition to pedagogical knowledge, the way teachers approach their classroom activities is shaped by their 'personal theories of teaching and learning'. As evidenced in the remarks made by the participants concerning language input (T1, T3, T5, T6 in Section 6.3.1), teachers' personal and subjective understandings of teaching and learning seem to function as a source of teachers' practices as well. In light of this finding, we seem to be able to make sense of why the participants tended to carry out teacher-centered activities, since they were found to have a belief system that was teacher-oriented rather than student-oriented. The findings suggest that teachers may interpret and react to washback on the basis of what they already know and believe about teaching, learning, and testing.

In view of these findings, to induce positive washback may require, on the part of teachers, a specialized knowledge base and upgraded beliefs about teaching and learning.

Another principal finding of this study has to do with the Chinese context. In the third chapter, it was discussed that China has an educational context which is centralized, knowledge-focused, and examination-based. It also has a socio-cultural environment in which teachers have to cope with large classes and a lack of freedom to choose their own textbooks and content of teaching. Classroom observations have shown that the traditional outdated teaching norms typical of the Chinese educational context are still present. The participating teachers in this study were found to hold a knowledge transmission perspective and to adopt transmission-oriented teaching styles. As indicated above, a little more than half of them (54% and 52%) still viewed teaching and learning as a process of imparting and receiving knowledge. They also viewed students as passive recipients of transmitted knowledge rather than active participants in the construction of meaning. It appears that the Chinese educational and socio-cultural environments may have a part to play in nurturing the above-mentioned teaching patterns and norms. Thus, it was imperative to adopt a socially and institutionally situated view about teacher learning and knowledge and consider the effects of both social and institutional contexts upon teacher practice. As pointed out by other language testing researchers (Turner, 2005), washback cannot be

understood apart from the socio-cultural environments in which it is embedded.

The findings summarized above have implications for both language testing research and other types of research (e.g., general education, language education, and educational innovation), for the findings of this study share a number of similarities to those of other research. For instance, one of them is that they all aim at improving teacher practice. Another is that they share a common focus on the ‘teacher factor’ or ‘teacher role’ in the classroom. To be specific, teachers are seen as the major change agents who play a central role in shaping classroom events (Borg, 2006; Davison, 2008; Davison & Hamp-Lyons, 2008; Turner, 2008, 2009). The following section discusses the implications of these findings.

### 7.3 Implications of the Study

The results of the present study indicate that overcoming the barriers to change is no simple task. It requires the joint efforts of authorities, test developers, students and particularly teachers. Its implications are twofold. On the one hand, the Chinese examination authorities and test developers could focus on improving test design to facilitate language teaching and learning. The simple reason, as pinpointed by Muñoz and Álvarez (2010), is that “students’ successful performance on assessment tasks greatly depends on how well teachers and test developers design those tasks” (p. 37). As illustrated and exemplified in studies by Colby-Kelly and Turner (2007), Davison, 2008, Davison and Hamp-Lyons (2008), Rea-Dickens (2004), and Saif (2006), considerations and efforts directed at alignment of assessment, curriculum, teaching and learning practices do make a difference. In this study the CET tasks were found not to correspond highly to tasks in the language use (TLU) domain. Thus, it would appear beneficial for the CET constructors to address issues and problems inherent in the test. As the new curriculum, the CECR, places an emphasis on facilitating students’ communicative competence, test constructors need to, as Shohamy (1993) suggests, do their job to ensure the construct of test validity and increase the match between the curriculum and the test. To this end, organized efforts need to be directed to more task-based test designs guided by contemporary language testing theories (Bachman & Palmer, 1996; Bailey, 1996). In this way, test papers would be designed in such a way that not only knowledge of English, but also the ability to use English, is tested. As Bachman and Palmer (1996) suggested, task characteristics need to be considered in order to ensure that test tasks correspond in specific ways to language

use tasks. Evidence needs to be provided to demonstrate that the test score reflects the area(s) of language ability they want to measure.

In addition, it would be helpful for test developers to be aware of the conditions for initiating teacher conceptual and behavioral change. In this connection, it would appear important for them to bear in mind that what may possibly affect teachers is not only *what* is included in the test, but also *how* the objectives of the test as well as training is provided to the teachers. As stated above, it would also appear important if test developers are aware that an over-emphasis on the power of tests and absence of attention to teachers' involvement may account for why intended washback did not occur in most studies. To ensure the success of instructional reforms, it is highly suggested that they de-emphasize the power of tests on the one hand and encourage teacher engagement on the other.

What I need to clarify here is that my goal of analyzing the role of the 'teacher factor' in fostering washback is not to dissuade test designers from using tests as instruments to innovate teaching, but rather to raise their awareness of the important role of this factor in engineering washback so that they may work out better ways to improve the existing test designs.

On the other hand, it would be beneficial if adequate attention were directed at another one of the stake-holders, language teachers. Given the involvement of the 'teacher factor' in effecting teacher behavioral change, consideration needs to be given to all issues and conditions concerning the observable (i.e., practices) and unobservable dimensions (i.e., beliefs) of teaching. As the chief implementers of reforms, it would appear beneficial for teachers to be aware that innovation and change are a necessary part of teacher development (Bailey 1992; Willis & Willis 1996). It would also appear essential that they be aware of the ways that can be used to promote their awareness. In the meantime, it might be crucial that support be provided to them to promote their awareness of the issues relevant to instructional change.

As mentioned above, one salient finding of the study is related to discrepancies and mismatches in teachers' beliefs and actions in the classroom situation. Based on the theoretical framework developed in the study, the discrepancies may be attributable to the interplay of teachers' BKE network(s). More specifically, they may be attributed to the incompatibilities between teacher declarative knowledge and their procedural knowledge (working knowledge). In view of the discrepancies, it would be highly

important for researchers to be concerned not only with the declarative dimension of teacher beliefs and knowledge (i.e., understanding of CLT, the CET, CECR), but also with the procedural dimension of them (i.e., the underpinnings of CLT). It would also be crucial for teachers to be alerted to such discrepancies, for both Woods (1996) and Ashton (1996) made the argument that teachers' misconceptions about teaching and learning derived from experience are difficult to change and must be challenged in powerful ways. From Woods' (1996) perspective, teachers' beliefs, knowledge and practice evolve through awareness of these discrepancies and resolution of conflicts. An important suggestion provided by him is that teachers should be provided with more opportunities to reflect on their teaching and recount their dilemmas, for the process of reflection is considered to be a crucial step for developing teacher expertise. Based on Woods (1996), it is the dilemmas that cause teachers to analyze and reflect on their beliefs, and to consider the various options for achieving their teaching goals. In this regard, rather than focusing on the negative impact of the discrepancies and mismatches, we need to view them as an indispensable step for teacher change or teacher professional development.

Apart from the powerful impact of the nested 'teacher factor' on teacher performances, two attributes of the 'factor' (teacher pedagogical knowledge and beliefs about teaching and learning) appear to play a deterministic role in shaping teacher practice. In light of this finding, teachers may benefit from developing awareness of what their pedagogical knowledge and beliefs about teaching and learning are and they may also benefit from reflecting on the gap between their BAKs and those of others. The objective of teachers reflecting on BAKs, according to Woods (1996), is to facilitate teachers' readiness. Based on his account, an awareness of one's own BAK may make it easier for one to accept others', to understand how they differ, and to decide that the difference can be worked through in areas of conflict. Other researchers (Borg, 1998, 2003; Griffiths & Tann, 1992; Kennedy 1987, Prawat, 1991; Richards & Lockhart, 1994; Roberts, 1998; Wallace, 1991) also suggest ways to promote teacher awareness. Like Woods, they also deemed reflection as a pre-requisite of teacher development and held that teachers should be encouraged to engage in reflection, research, or systematic inquiry. The benefit of reflective teaching, from their interpretation, is that it may allow teachers to make tacit beliefs and practical knowledge explicit. Here, what should be noted is that underlying the view that underscores the important role of reflection is a constructivist view about

learning.

Given the myriad of misconceptions about teaching and learning, the limited English proficiency, and the insufficient knowledge and training that Chinese EFL teachers have, it would appear that reflective activities such as training in English teaching methodology, test and curriculum development, research (i.e., observation, introspection and inquiry), and micro-teaching or peer-teaching might be beneficial to them, for such opportunities may allow them to actively reflect on their beliefs, knowledge, and practices, and then modify and reinterpret them. Furthermore, these reflective activities may also allow them to enhance teacher professionalism.

The process of reflection may also allow them to reconstruct their personal theories of language teaching and learning, and recount their dilemmas (Bailey, 1992; Crandall, 1994; Flowerdew, et al. 1992; Freeman, 1998; Freeman & Richards, 1996; Shulman, 1992; Woods, 1996). What merits attention is that the above researchers regard research as a reflective tool.

In addition to the reflective activities, the participants in this study may benefit from opportunities to link theory with practice. To be concrete, they may benefit from hands-on guidance that serves to integrate the constructivist theory with actual classroom practice.

However, neither guidance nor training is sufficient. Teacher conceptual and behavioral change may involve collaborative and/or autonomous learning. I would suggest that teachers concentrate on improving their methodological skills to achieve effective teaching.

It is hoped that this study can help provide teachers with a basis for reflection about language teaching. Since the current teaching methods adopted by Chinese CE teachers, as shown in the data of this research, do not enable them to accomplish the goals stated in the 2004 revised draft of the CECR, it would prove useful for teachers to modify their instructional behaviors to better meet the students' needs.

First, there seems to be a need for teachers to be sensitized to the potentially bidirectional nature of washback (either positive or negative). In view of the tendency that they are likely to exaggerate the power of test impact, it would be beneficial if they are aware of the nature of washback as well as the intentions underlying the educational innovation. It might also be beneficial if they are aware that there have already been immense changes in external learning, teaching and testing environment in Chinese EFL education.

Second, as mentioned above, there seems to be a need for teachers to develop their awareness about their own BAKs. It would also be beneficial for them to be aware of the differences in the methodologies that different teachers employ as well as the gap that exists in terms of their perceptions of test impact and its washback effects on them. It is expected that once they see and reflect on the differences and gap in their BAKs as well as in their teaching behaviors, they may be aware of the real problems underlying their teaching practice and probably develop a more reasonable view about their teaching and then be in a position to change their methodologies to lessen the gap.

Third, it would be helpful for teachers to be aware of the urgent need to update their knowledge of language teaching and learning theories, since the traditional transmission model of education still prevails in their classrooms and language ability is still viewed by some teachers as a set of finite components – grammar, vocabulary, pronunciation, etc. Thus, it appears that there is a need that some “input” related to language learning and teaching theories should be provided to them to make them aware that language acquisition concerns more than knowledge of grammar and vocabulary. Above all, their outdated teaching concepts and philosophies might need to be transformed and replaced by more up-to-date ones.

As teachers’ awareness of the above issues is increased, hopefully they would be motivated to reform their own curriculum and make conscious efforts to modify their English teaching methodologies. With a greater awareness and more accurate perception of the nature of washback effects and the conditions under which they operate, teachers may eventually alleviate negative washback and replace it by positive washback.

One way of promoting teachers’ awareness and facilitating their involvement in the change process, as suggested by Davison (2008), Tavares and Hamp-Lyons (2008) and Turner (2008, 2009) is through providing well-tailored guidance, support and training to them. Such guidance and support may trigger their reflection on some of their internal factors as well as the issues that emerged in this study.

In addition, considering the common resistance encountered in various educational innovations as well as the negative perceptions the Chinese EFL teachers maintain towards the CET (as found in this study in **6.2.3** and **6.3.1**), it would be helpful if more interaction were undertaken between the interpretative processes of the teacher and test developers. It might be helpful for test developers to give a rationale for



carrying out a certain type of change. Meanwhile, it also seems to be of help if the objectives and purposes of the CET are made available to all test-takers. With an enhanced understanding of the rationale, objectives and purposes of the CET, teachers may reframe their conceptions of the testing reforms.

What warrants our attention is that an awareness of what is going on in the language teaching and testing circle is only an initial step to helping teachers identify and define their implicit beliefs. We must be mindful that instructional change occurs as a gradual progress and it is the product of long-term comprehension of different contexts for teaching. To bring about such a change, the Chinese EFL teachers may also need to attain adequate proficiency and build a sound knowledge base. For instance, there seems to be a need for them to enhance their language skills and develop a much more fine-grained understanding of the principles involved in CLT or task-based approaches. In addition to guidance on the testing dimension of change, it appears that teachers are badly in need of guidance to correctly interpret theories of SLA and translate them into effective instructional practice. We should note that despite the increase in the number of teacher training programs such as in-service teacher training, summer workshops, and study abroad programs (that have been provided by the Chinese governments to EFL teachers), the findings of this study and those of others (Shu, 2004; Wu, 2001) suggest that the majority of teachers are still inadequately trained in language teaching methodology. Furthermore, some of these programs were said to be inadequate in the sense that they were not targeted to meet teachers' specific needs. In this regard, teacher training or teacher education programs in China could possibly offer more well-tailored opportunities to teachers for professional development so that they will acquire adequate proficiency and be well prepared to face the challenges of the reforms.

It needs to be noted that conditions that hinder the Chinese educational reforms found in this study remain prevalent. Only by tackling these causal conditions can any meaningful changes be made in education. Only by so doing, will the goal of quality education be attained in the Chinese university setting.

#### 7.4 Recommendations

The results of the study suggest that improvements should be made in future reform efforts in terms of test design, teacher training, and research methods. Some recommendations are given with the intent to facilitate the implementation of

educational reforms or innovations. This section outlines a set of recommendations in relation to the findings of this study.

- ♦ Tests objectives need to be clear and transparent. As a means of promoting teacher pedagogical change, they need to reflect a shift from transmission, product-oriented theories to constructivist, process-oriented theories and pedagogies of teaching and learning. The rationale for including this objective is that only when teachers achieve a good comprehension of the objectives, content and methodology of a test, will they be in a better position to change their perceptions and behaviors which conform to its innovation.
- ♦ Future testing innovation endeavors need to be accompanied by ongoing training and/or appropriate teacher guidance and support on assessment and instructional practices over time. As teacher evolution is a transformative process (or as Vygotsky (1978) terms it, a ‘process of internalization’), it takes time for cognitive and behavioral changes to take place. Thus, teacher training for language teachers should provide adequate opportunities to engage them in reflective activities (e.g., classroom research, action research, classroom observation, team teaching, test and curriculum development), since these activities, as noted by Richards (2008), are the principal sources for constructing a knowledge base.
- ♦ Teacher training needs to help teachers raise their awareness in terms of (1) what their pedagogical knowledge is and whether there is a gap between their pedagogical knowledge and that defined by L2 theorists; and (2) what their BAKs are and how they differ from those of others. As discussed above, this is considered to be a crucial step for developing teacher expertise. From Woods’ (1996) point of view, teachers’ beliefs, knowledge and practice evolve through awareness of these discrepancies and resolution of conflicts. He puts forward the suggestion that teachers be provided with more opportunities to reflect on their teaching and recount their dilemmas. According to him, it is the dilemmas that cause teachers to analyze and reflect on their beliefs, and to consider the various options for achieving their teaching goals. In this regard, rather than viewing conflicts, differences and discrepancies as forces that hinder the processes of

change, they should be perceived as an impetus for change and an indispensable part of teacher evolution. Therefore, teacher training programs need to help teachers increase their awareness of various discrepancies that may exist in their perceptions and behaviors. Meanwhile, guidance needs to be provided to teachers, as suggested by Richards (2008) and Woods (1996), to make their tacit and inaccessible beliefs and knowledge of teaching pedagogy explicit and accessible.

- ♦ Research on washback needs to focus on the longer-term effect of teacher change (i.e., change over time). Owing to the nature of the nested ‘teacher factor’, research needs to examine the phenomenon of washback in the growing professionalism of the field. Consideration should be given to how teachers’ beliefs (cognition underpinning teacher practice) are conceived, how their knowledge structures are formed, and how teachers’ beliefs and knowledge inform their practice.
- ♦ The notion of washback needs to be redefined or reconceptualized. Rather than dwelling on debates over whether a test can induce positive or negative washback, an emphasis should be given to the more complex issue of how a test interacts with the ‘teacher factor’ in the course of washback. To deemphasize the power of tests, consideration needs to be given to how to secure teacher consent, participation and active involvement in the change process. As discussed in Section 6.2.3, apart from the studies conducted by Davison (2008), Muñoz and Álvarez (2010), Tavares and Hamp-Lyons (2008) and Turner (2001, 2002, 2005, 2008, 2009), it appears little attention has been paid to teachers’ active involvement in innovation. However, ‘teacher involvement’ seems to be a factor that should not be overlooked in washback research (Turner, 2008, 2009). Based on Woods (1996), Kennedy (1988) and Fullan (1991), such an over-emphasis on the power of tests and absence of attention to teachers’ engagement may account for why intended washback did not occur in most studies. Therefore, to ensure the success of instructional reforms, there is a need to de-emphasize the power of tests on the one hand and encourage teacher participation on the other.
- ♦ Research from general education needs to be drawn on in research on washback. It appears that studies concerning the role of the ‘teacher factor’ in the context of

language testing are still scant. To understand how the ‘teacher factor’ functions, we have to turn to parallel work in general education, L2 teacher education, and cognitive psychology (Allwright & Bailey, 1991; Chaudron, 1988; Edge & Richards, 1993; Freeman 1996; 1998, Freeman & Johnson, 1998; Nunan, 1989; Richards & Nunan 1990; Richards, 2008; Van Lier, 1988). The rationale for drawing on other areas of research is that there is a substantial body of literature in these areas that focus on beliefs, knowledge, proficiency, experience, and practice. The insights gained from these areas may help us tease apart the ‘teacher factor’ in washback.

- ♦ Efforts need to be exerted to encourage classroom-based performance assessment. (e.g., role plays, interviews) and personal-response assessments (e.g., self-assessments, conferences) (Brown & Hudson, 1998), for these assessments are considered to be a bottom-up process. The application of these bottom-up, process-oriented assessment approaches may help balance those which are top-down, product-oriented (see Colby- Kelly & Turner, 2007).
- ♦ Great importance must be attached to the understanding of the socio-cultural, educational and institutional context in which teaching, learning, and testing occurs when planning/implementing reforms.
- ♦ Greater emphasis should be placed on alignment of assessment and the central agents of change, since it appears that considerations and efforts directed at alignment of assessment, curriculum, teaching and learning practices (as Colby-Kelly and Turner (2007) calls it, ‘assessment bridge’) do make a difference.

### 7.5 Suggestions for Future Research

Considering the complex nature of washback and in view of what was found in this study (that exams alone have a limited impact on how teachers teach), I am now in a position to make a few suggestions for future research. I suggest, therefore, a shift of focus from discussions of the existence or nature of washback in the field to the study of the role of the internal factors (the ‘teacher factor’) as well as of the external factors (e.g., the ‘learner factor’) in the process of washback. In addition to asking whether

washback exists or not and whether it is negative or positive, we should now first ask how a test can be developed to contribute to fundamental teacher methodological changes.

Future studies could integrate teacher beliefs about personal constraints within the context of language instructional innovation, giving teachers opportunities to discuss and reflect not only about their beliefs about teaching, learning and how to teach, but also about personal constraints, dilemmas and challenges encountered in carrying out instructional practice.

Since the study of the ‘teacher factor’ in washback involves an examination of teachers’ thought processes, it appears that it would require a shift in research methods. According to Richards and Farrell (2005) and Woods (1996), stimulated recall would give us a clearer picture of what can trigger reflection. They view it as an appropriate method for revealing the underlying beliefs or motives of teachers. According to them, this method may allow teachers to describe or articulate what they know about teaching and learning and how they know it. Thus, it would be beneficial for future research to conduct studies using this method to further examine the values and beliefs underpinning teachers’ classroom practice.

As mentioned in Section 5.5.3, this study has chosen to examine only some aspects of the ‘teacher factor’ (i.e., teacher BKEs). In this regard, it would be imperative and informative if other aspects are integrated into future research. One suggestion is that future research can broaden the scope to include aspects such as teacher proficiency, teacher background, teacher motivation, etc.

The results of this study have demonstrated that both teachers’ internal factors and contextual factors (institutional and societal) can contribute to washback. However, while the study has added an important dimension to our understanding of the phenomenon of washback, it is difficult to ensure that it has adequately covered all the factors that contribute to washback in China. The results of the study have demonstrated that there are several other factors which need exploring. Therefore more extensive studies need to be conducted to explore how various forces are combined to produce varying cases of washback.

In addition, future research needs to further examine the ‘student’ variable in washback. In fact, this is a concern commonly shared by other researchers (Alderson, 2004; Bailey, 1996; Messick, 1996; Muñoz & Álvarez, 2010). They all hold that for beneficial washback to take place, the student role and/or student awareness of

examination objectives and their connection to educational goals must be taken into consideration.

### 7.6 Limitations of the Study

This study attempted to relate washback research to issues of conceptual and behavioral change in order to advance our understanding of the washback phenomenon. Although it intends to move beyond a simple understanding of tests as predictors of teacher behaviors to an investigation of the ‘teacher factor’ (teacher beliefs, knowledge and behaviors in context) as well as its interrelationships, it is hard to ensure that a study such as this will adequately cover all the factors that contribute to or inhibit washback in China. The data indicate that there are further factors which need exploring.

Owing to the length of this thesis, only the salient results could be reported in some cases (e.g., some of the results from the EFAs, CFAs and SEM analyses). Future reporting could expand on the details of these data.

Owing to the long-term effects of the reform in China as well as the emergent nature of beliefs, it is too early to provide a detailed evaluation of the effects of the revised CET on teaching. As more remains to be learned about the ‘teacher factor’ and its relationship to washback, further research would benefit from the use of more prolonged observation and dynamic and contextualized research tools.

### 7.7 Strengths and Contributions of the Study

Traditional washback research focuses on whether there is intended washback brought about by examination change. This study, however, represents one of the first few attempts (another two being Tan, 2008; Turner, 2008, 2009) in the field of language testing to investigate how the intertwined or networked ‘teacher factor’ fits into the washback phenomenon. One of the strengths of the study is that it has provided detailed empirical evidence showing how teachers’ diverse beliefs, knowledge, and experience are combined to influence teacher practice. This evidence yielded insights into the complexity of the process of testing change by helping tease apart some of the previously less-researched factors.

Another strength of the study, drawing on Woods’ (1996) conceptual framework, lies in the use of structural equation modeling for understanding washback.

Adopting a MM approach to both data collection and analyses is another strength of the study. Concretely, throughout the process of research, careful consideration was

given to issues of data quality and soundness of the study. The steps that were taken include: (1) ensuring that the study was firmly grounded in the conceptual framework proposed by educational and ELT theorists; (2) basing the study on an extensive literature review, which helped broaden understanding of the research approach appropriate for dealing with the research problems that were confronted; (3) devising a solid research design – MMR – that helps maximize the possibility of addressing my research questions thoroughly; and (4) applying the strategies of MMR – using multiple sites and purposive sampling, prolonged engagement, persistent observation, triangulation and etc.

The findings of the study contribute to the literature on general education, L2 teacher education, and cognitive psychology relating to how to initiate teacher conceptual and behavioral change.

### 7.8 Concluding Remarks

As indicated at the beginning of this paper, the CET has been widely accused of its “adverse effects” (so-called “negative washback”) on Chinese CE teaching and learning. However, the results of this study reveal that while testing, as an external force, had limited washback effects on teachers’ behaviors in the classroom, it should not bear the whole responsibility for its unintended adverse effects. In fact, the data from this study show it neither constituted a constraining factor, nor functioned as a facilitating factor to the implementation of instructional innovations. The findings add more evidence to earlier studies that although other factors such as institutional context and societal pressures are also involved in the washback phenomenon, they appear less powerful when compared to the inner forces, such as the ‘teacher factor’. This study has demonstrated the salient role of the ‘teacher factor’ as it is the teachers who meet the daily challenges of addressing the diverse needs of the students.

The findings of the study show that teachers’ intertwined beliefs, knowledge systems, and prior experiences serve as a framework through which they make sense of their classrooms events. The BKE framework, especially their theoretical knowledge about teaching, is shown to make a difference in how washback or instructional innovations take shape. Specifically, the framework is found to determine not only how teachers interpret washback and implement reforms, but also how they would change their classroom practice. It appears that these findings reinforce those by Andrews (2003), Shulman (2000) and Turner (2008; 2009) in that

the 'teacher factor' is seen as an essential component of teacher professionalism.

In light of the above findings, a simple examination of the washback phenomenon with little consideration of the 'teacher factor' in terms of teachers' beliefs, knowledge base, and proficiency level as well as experience of teaching and learning, cannot fully explain critical washback issues.

When the results of this study are compared to research carried out in other EFL contexts, two main characteristics are found to exist in current testing and educational innovations. One characteristic relates to the way Chinese EFL teachers perceive and react to testing and ELT innovation. This is similar to that exhibited in other contexts. Such a characteristic reflects the fact that while tests in different contexts may affect teaching and learning at varying degrees, and moreover washback may take different forms, the rules or mechanisms underlying the washback phenomenon as well as the complexities implicit in it seem to make little difference. The second characteristic is that testing innovation shares many common features with other sorts of educational innovations. Four shared salient features include: (1) complexity and unpredictability exhibited in implementing teacher conceptual and behavioral change, especially pedagogical change; (2) prevalence of teachers' resistance to change or innovation; (3) individual differences in terms of teachers' attitudes, beliefs and actions in the process of change; and (4) discrepancies between teachers' actual behaviors and their claimed beliefs about teaching. A detailed discussion of these features can be found above.

It is due to the above characteristics that, to get a clearer and more concise picture of the role of 'teacher factor' in the washback phenomenon and to bring about the deeper, underlying level of change, research on washback must go beyond the discussion of tests and test impact and adopt a cross-cultural and interdisciplinary perspective to address the essential issues relevant to this study. As elucidated above, future testing innovation endeavors should be accompanied by ongoing teacher training and guidance as well as teacher active involvement in the change process.

In summary, it is hoped that the recurring features implicit in the change process discussed in this study will provide the basis for improvement for further innovation endeavors. It is also hoped that this study will give a strong impetus to the study of the 'teacher factor' in washback in the future.



## REFERENCES

- Abelson, R. P. (1979). Differences between belief systems and knowledge systems. *Cognitive Science*, 3, 355–366.
- Alderson, J.C. (2001). *The Shape of things to come: will it be the normal distribution?* In European Language Testing in a Global Context – Proceedings of the ALTE Barcelona Conference, July, 2001.
- 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 Associates.
- 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, 115-129.
- Allwright, D. & Bailey, K. (1991). *Focus on the language classroom*. New York: Cambridge University Press.
- Anderson, J. (1993). Is a Communicative Approach Practical for Teaching in China? Pros and Cons. *System*, 21 (4). 471-480.
- Andrews, S. (1994). The washback effect of examinations: Its impact upon curriculum innovation in English language teaching. *Curriculum Forum* 1 (4).
- Andrews, S., Fullilove, J. & Wong, Y. (2002). Targeting washback – a case-study, *System*, 30 (2), 207-223.
- Andrews, S. (2004). Washback and curriculum innovation. In L.Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods*, pp. 37-50. Mahwah, NJ: Lawrence Erlbaum Associates.
- Arbuckle, J. L. (2006). *Amos 7.0 User's Guide*. Amos Development Corporation.
- Bachman, L. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Brookhart, S. & Freeman, D. (1992). Characteristics of Entering Teacher Candidates. *Review of Educational Research*. 62 (1), 37-60.
- Bailey, K.M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing* 13(3), 257-279.

- Bailey, K. M., Bergthold, B., Braunstein, B., Jagodzinski Fleishman, N., Holbrook, M. P., Tuman, J., Waissbluth, X., & Zambo, L. (1996). The language learner's autobiography: Examining "apprenticeship of observation." In D. Freeman & J. Richards (Eds.), *Teacher learning in language teaching*, pp. 11-29. Cambridge, UK: Cambridge University Press.
- Bailey, K. M. (1999). *Washback in language testing*. TOEFL Monograph Series, Ms. 15. Princeton, NJ: Educational Testing Service.
- Barcelos, A. M. F. (2003). Researching beliefs about SLA: A critical review. In P. Kalaja and A. M. F. Barcelos (Eds.), *Beliefs about SLA: New research approaches*, pp. 7-33. Dordrecht: Kluwer Academic Publishers.
- Bentler, P. M. & Chou, C. (1987). Practical issues in structural modeling. *Sociological Methods and Research*, 16, 78-117.
- Bentler P.M. (1992): Structural equation models in medical research. *Statistical Methods in Medical Research*, 1 (2), 159-181.
- Bogdan, R. & Biklen, S. K. (1998). *Qualitative Research for Education: An Introduction to Theory and Methods*. London: Allyn & Bacon.
- Borg, S. (2003). Teacher cognition in language teaching: a review of research on what language teachers think, know, believe, and do. *Language Teaching*. 36 (2), 81-109.
- Borg, S. (2006). *Teacher cognition and language education: Research and practice*. London: Continuum.
- Breen, M.P. & Candlin, C. (1980). The essentials of a communicative curriculum in language teaching. *Applied Linguistics*, 1 (2), 89-112.
- Brookhart, S. M. & Freeman, D. J. (1992). Characteristics of entering teacher candidates. *Review of Educational Research*, 62 (1), 37-60.
- Brewer, J. & Hunter, A. (1989). *Multimethod research: a synthesis of styles*. Newbury Park, California: Sage.
- Brown, H. D. (1994). *Teaching by Principles: An Interactive Approach to Language Pedagogy*. (1st ed.). N. J. : Prentice Hall Regents
- Brown, J. D. (1997). Designing surveys for language programs. In D. Griffiee & D. Nunan (Eds.), *Classroom teachers and classroom research*, pp. 109-122. Tokyo: The Japan Association for Language Teaching (JALT).
- Brown, J. D. & Hudson, T. (1998). *The Alternatives in Language Assessment*. *TESOL Quarterly*, 32 (4), 653-675.
- Brumfit, C. (1991). Problems in defining instructional methodologies. In K. de Bot, R. P. Ginsberg, & Kramsch, C. (Eds.), *Foreign language research in cross-cultural*

- perspective*, pp.133-144. Amsterdam/Philadelphia: John Benjamins.
- Burnaby, B. & Sun, Y. (1989). Chinese Teachers' Views of Western Language Teaching: Context Informs Paradigms. *Tesol Quarterly*, 23 (2), 219-238.
- Burns, A. (1990). Focus on language in the communicative classroom. In G. Brindley (Ed.). *The Second Language Curriculum in Action*. Sydney: NCELTR.
- Burrows, C. (2004). Washback in classroom-based assessment: A study of the washback effect in the Australian adult migrant English program. In L.Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods*, pp.113-128. Mahwah, NJ: Lawrence Erlbaum Associates.
- Cai, J. (2002). Pressures in current college English teaching. *Foreign Language Teaching and Research (Bimonthly)*, 34 (3), 228-230.
- Cai, J. (2003). The Reasons for Carrying out Fundamental Reforms on Traditional College English Teaching. *China University Teaching*, 11, 25-30.
- Cai, J. (2005). Some Thoughts on College English Teaching. *Foreign Language Teaching and Research*, 37 (2), 83-91.
- Cantu, D. A. (2001). *An Investigation of the Relationship Between Social Studies Teachers' Beliefs and Practice*. Lewiston, New York: Mellen Press.
- Campbell, K. P. & Zhao, Y. (1993). The Dilemma of English Language Instruction in the People's Republic of China. *Tesol Journal*, 2 (4), 4-6.
- Cen, J. (1999). The status quo of foreign language teaching at the tertiary level. *Foreign Language Teaching and Research*, 117 (1), 3-6.
- CEC. (1985). *The College English Curriculum* (for both art and science students) [M]. Shanghai Foreign Language Education Press.
- CEC. (1999). *The College English Curriculum*. [revised version][M]. Shanghai: Shanghai Foreign Education Press; Beijing: Higher Education Press.
- CECR (2004). *The College English Curriculum Requirements* (trial version) [M]. Beijing: Higher Education Press.
- CET News. (2005). [http://www.cet.edu.cn/gaige\\_news01.htm](http://www.cet.edu.cn/gaige_news01.htm)  
[http://www.cet.edu.cn/gaige\\_news02.htm](http://www.cet.edu.cn/gaige_news02.htm)  
[http://www.cet.edu.cn/gaige\\_news02\\_2.htm](http://www.cet.edu.cn/gaige_news02_2.htm)  
[http://www.cet.edu.cn/gaige\\_news02\\_3.htm](http://www.cet.edu.cn/gaige_news02_3.htm)
- CET Sample. (2005). [http://www.cet.edu.cn/cet\\_pen2.htm](http://www.cet.edu.cn/cet_pen2.htm)
- Chapelle, C. A. (1998). Construct definition and validity inquiry in SLA research. In L.

- F. Bachman & A. D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research*, pp. 32-70. New York: Cambridge University Press.
- Chapman, D. W. & Snyder, C. W. (2000). Can high stakes national testing improve instruction: Reexamining conventional wisdom. *International Journal of Educational Development*, 20, 457-474.
- Chen, G. (1999). How can the pot of English be brought to a boil? *Foreign Language World*, 74 (2), 26-29.
- Chen, L. (2002). *Washback of a public examination: Impact of the Basic Competency Test on Taiwan junior high school English teaching*, presented in Edward F. Hayes Graduate Research Forum and paper appeared in Edward F. Hayes Graduate Research Proceedings, 2, 91-99.
- Chen, S. (1995). An Analysis of the CET 4 Backwash Effects. *Shantou University Journal (Humanities Quarterly)*, 2 (3), 78-82.
- Chen, X. (1999). Reflection on Communicative Teaching in College English. *Journal of Chongqing University (Social Sciences Edition)* 5 (3), 86-87.
- Chen, Z. (2004). The need for respecting facts in academic discussions—Discussing with Mr. Runqing Liu on issues of the CET 4/6. *Foreign Language World*, 99 (1), 74-77.
- Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, 11(1), 38-54.
- Cheng, L. (1998). Impact of a public English examination change on students' perceptions and attitudes toward their English learning. *Studies in Educational Evaluation*, 24 (3), 279-301.
- Cheng, L. (1999). Changing assessment: Washback on teacher perceptions and actions. *The Teacher and Education*, 15, 253-271.
- Cheng, L. (2001). Washback studies: Methodological considerations. *Curriculum Forum*, 10 (2), 17-32.
- Cheng, L. (2003). Looking at the impact of a public examination change on secondary classroom teaching: A Hong Kong case study. *Journal of Classroom Interaction*, 38 (1), 1-10.
- Cheng, L. (2004). The washback effect of a public examination change on teachers' perceptions toward their classroom teaching. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods*, pp.147-170. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Cheng, L. & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In Cheng, L., Watanabe, Y., and Curtis, A. (Eds.), *Washback in language testing: Research contexts and methods*, pp. 3-17. Mahwah, NJ: Lawrence Erlbaum Associates.

- Cheng, L. & Qi, L. (2006). Description and Examination of the National Matriculation English Test, *Language Assessment Quarterly*, 3 (1), 53-70.
- Cheng, Sh. (1997). On College English Test Band IV and Band VI. *Journal of Guangzhou Normal University (Social Science Edition)*, 3 (3), 94-100.
- Cheng, Sh. (2004). The dilemma caused by the norm. *Journal of Guangzhou University (Social Science Edition)*, 3 (1), 88-96.
- Clark, J. L. (1987). *Curriculum renewal in school foreign language learning*. Oxford, England: Oxford University Press.
- Clift, R. (1987). English teacher or English major: Epistemological differences in the teaching of English. *English Education*, 19, 229-236.
- Colby-Kelly, C. & Turner, C.E. (2007). AFL research in the L2 classroom: taking assessment to the next level. *Canadian Modern Language Review*, 64 (1), 9-38.
- Cook, T. D. & Campbell, D.T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Cooper, J. (1990). Cooperative Learning and College Teaching: Tips from the Trenches. *Teaching Professor*, 4 (5), 1-2.
- Crandall, J. (2000). Language teacher education. *Annual Review of Applied Linguistics*, 20, 34-55.
- Creswell, J. W. (1994). *Research Design: Qualitative and Quantitative Approaches* (2<sup>nd</sup> ed.). Thousand Oaks, Calif: Sage Publications.
- Creswell, J & Plano Clark, V.L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks , CA : SAGE.
- Creswell, J. (2009). *Research design : qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage.
- Crook, D. (1998). *Some Problems of Chinese Education as Seen Through the Eyes of a Foreigner*, Presented at the International Symposium on Teaching English in the Chinese Context, Beijing, China and Paper Appeared in Proceedings of ELT in China (2<sup>nd</sup> ed.), pp. 29-38. Beijing: Foreign Language Teaching and Research Press.
- Dai, W. (2001). The construction of the streamline ELT system in China. *Foreign Language Teaching and Research*. 33(5), 322-327.
- Davison, C. (2008). *Using summative assessments for formative purposes: The ultimate justification for learners and teachers*. Presented at 30th Annual Language Testing Colloquium, (LTRC 08). Hangzhou.
- Davison, C. & Hamp-Lyons (2008). *Applying school-based assessment for learning in a formal examination system*. Workshop presented at the 30<sup>th</sup>

- Language Testing Research Colloquium, Hangzhou, China.
- Davies, A. (1968). *Language testing symposium: A psycholinguistic approach*. Oxford: Oxford University Press.
- Davies, P. M. & Pearse, E. (2000) *Success in English Teaching*. Oxford: Oxford University Press.
- Donnelly, W. & Yu, Zh. (2002). Testing the Test: CET and Communicative Validity. *Journal of Chendu University of Information Technology*, 17 (2), 138-148.
- Dong, Y. (2003) Reading & writing-Top priority in China's college English education. *Foreign Language World*, 1, 2-6
- Ellis, R. (1990). *Instructed Second Language Acquisition*. Cambridge: Basil Blackwell.
- English, Fenwick W. (2000). *Deciding what to teach and test: developing, aligning, and auditing the curriculum*. Thousand Oaks, CA: Corwin Press, Inc.
- Ernest P. (1989). The knowledge, beliefs and attitudes of the mathematics teacher: A model. *Journal of Education for Teaching*, 15 (19), 13-33.
- Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational Research*, 38, 47-65.
- Feng, Q. (2000) A discussion of the problems of College English teaching and reform strategies. *Foreign Language Teaching* (3), 55-61.
- Frederiksen, J. R. & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, 18 (9), 27-32.
- Freeman, Y.S. & Freeman, D.E. (1992). *Whole language for second language learners*. Portsmouth, NH: Heinemann.
- Freeman, D. (1996). Renaming experience/ reconstructing practice: Developing new understandings of teaching. In Freeman, D. & Jack C. Richards (Eds.), *Teacher learning in language teaching*, pp. 221-41. New York: Cambridge University Press.
- Freeman, D. (2002). The hidden side of the work: teacher knowledge and learning to teach. A perspective from North American educational research on teacher education in English language teaching. *Language Teaching*, 35, 1-13.
- Freeman, D. & Richards, J. C. (1993). Conceptions of teaching and the education of second language teachers. *TESOL Quarterly*, 27 (2), 193-216.
- Freeman, D. & J. Richards (1996). *Teacher learning in language teaching*.

- 
- Cambridge: Cambridge University Press.
- Fullan, M.D. (1982). *The meaning of educational change*. NY: Teachers College Press.
- Fullan, M. (1993). *Change forces: Probing the depths of educational reform*. Bristol, PA: Falmer.
- Fullan, M. (1999). *Change Forces: The Sequel*. (1<sup>st</sup> ed. ). London: Falmer Press.
- Fullan, M. (2001). *Leading in a culture of change*. San Francisco: John Wiley and Sons.
- Fullan, M.G. & Stiegelbauer, S. (1991). *The New Meaning of Educational Change*. (2<sup>nd</sup> ed.). New York: Teachers College Press.
- Gao, Y. (2003). Need for Scientific Evidence in Foreign Language Teaching and Learning Reforms. *Foreign Language Teaching and Research*, 3, 222-223.
- Garson, (2007). Factor Analysis. Retrieved 11 January 23, 2007, from: <http://www2.chass.ncsu.edu/garson/pa765/factor.htm>.
- Gass, S. M., & Selinker, L. (2001). *Second language acquisition: An introductory course*. (2nd ed.). Mahwah, NJ: Erlbaum.
- Gatbonton, E. & Gu, G. (1994). Preparing and Implementing a Task-based ESL Curriculum in an EFL Setting: Implications for Theory and Practice. *Tesol Canada Journal*, 11 (2), 9-29.
- Glaser, R. and Bassok, M. (1989). Learning Theory and the Study of Instruction. *Annual Review of Psychology*, 40, 631-666.
- Glaser, R. and Silver, E. (1994). Assessment, Testing, and Instruction: Retrospect and Prospect. *Review of Research in Education*, 20, 393-419.
- Glaser, B.G., and Strauss, A.L. (1990). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine.
- Green, A. (2006). Watching for washback: Observing the influence of the International English Language Testing System Academic Writing Test. *Language Assessment Quarterly* 3 (4), 333-368.
- Green, A. (2007). *Studies in language testing: Vol. 25. IELTS washback in context – Preparation for academic writing in higher education*. Cambridge: Cambridge University Press.
- Greene, J. C. (2007). *Mixed methods in social inquiry*. San Francisco, CA: Jossey-Bass.
- Greene, J. C., Caracelli, V. J., & Graham W. F. (1989). Toward a Conceptual

- Framework for Mixed-method Evaluation Designs. *Educational Evaluation and Policy Analysis*, 11 (3), 255-274.
- Grossman, P. (1990). *The making of a teacher: Teacher knowledge and teacher education*. New York: Teachers College Press.
- Grossman, P., & Richert, A. (1988). Unacknowledged knowledge growth: A re-examination of the effects of teacher education. *Teaching and Teacher Education*, 4 (1), 53-62.
- Gu, X. (2003). Counter-evidence of the Comment “Time-consuming with Low efficiency” on college English teaching and learning. *Foreign Language World*, 3, 49-53.
- Gu, X. (2005). *Positive or Negative? An Empirical Study of CET Washback on College English Teaching and Learning in China*. Unpublished doctoral dissertation, Shanghai Jiao Tong University, China.
- Gu, W. & Liu, J. (2005). Test Analysis of College Students' Communicative Competence in English. *Asian EFL Journal*, 7 (2), 90-116.
- Guo, Y. (2003). Evaluation of College English Test. *Journal of Donghua University*, 20 (3), 144-146.
- Hamp-Lyons, L. (1997). Washback, impact and validity: ethical concerns. *Language Testing*, 14 (3), 295 – 303.
- Han, C. (1999). An analysis of the practicality of the CLT. *Foreign Language World*, 74 (2), 8-10.
- Han, B. (2002). The Urgent Need for Carrying out Testing Reforms on the Language Proficiency of Chinese Tertiary-level Students. *Foreign Language Teaching and Research*, 6. 410-411.
- Han, B., Dai, C. & Yang, L. (2004). On the Problem of the CET. *Foreign Languages and Their Teaching*, 2, 17-23.
- Hargreaves, A., Earl, L., Moore, S. & Manning, S. (2001). *Learning to Change: Beyond Participants and Standards*, San Francisco: Jossey-Bass.
- Harvey, P.A. (1985). Lesson to be learned: Chinese approaches to language learning. *ELT Journal*, 39 (3), 183-186.
- Hayes, B. & Read, J. (2004). IELTS test preparation in New Zealand: Preparing students for the IELTS academic module. In L.Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods*, pp. 97-111. Mahwah, NJ: Lawrence Erlbaum Associates.
- Huang, D. (2002). *A preliminary investigation of CET-4 washback*. Paper presented at the International Conference on Language Testing and Language



- Teaching. Shanghai: Shanghai Jiao Tong University, 224-233.
- Huang, D. & Jin, G. (2001). Washback Effect of Language Testing. *Journal of Southwest Jiaotong University (Social Sciences)*, 2 (3), 107-110.
- Hughes, A. (1988). Introducing a needs-based test of English language proficiency into an English-medium university in Turkey. In Hughes, A. (ed.), *Testing English for university study*, pp.134-146. ELT Documents #127. Modern English Publications in association with the British Council.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Jin, Y. (2000). The Washback Effect of the CET-SET on Teaching and Learning. *Foreign Language World*, 80 (4), 56-61.
- Jin & Yang (2006). The English Proficiency of College and University Students in China: As Reflected in the CET. *Language, Culture and Curriculum*, 19 (1), 21-36
- Jing, Sh. (1999). Reasons for the Lack of Efficiency in Chinese College English Teaching and Learning. *Foreign Language Teaching and Research*, 117 (1), 21-23.
- Jing, Sh. (1999). *Why can't the pot of English teaching be brought to a boil?* China Youth Daily.
- Johnson, K. E. (1999). *Understanding language teaching: Reasoning in action*. Boston: Heinle and Heinle.
- Johnson, K. E. & Golombek, P. R. (2002). (Eds.) *Teachers' narrative inquiry as professional development*. NY: Cambridge University Press.
- Johnson, R. B. & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Education Researcher*, 33 (7), 14-26.
- Kalaja, P. (1995). Student beliefs (or metacognitive knowledge) about SLA reconsidered. *International Journal of Applied Linguistics*. 5 (2), 191-204.
- Kalaja, P. (2003). Research on students' beliefs about SLA within a discursive Approach. In Kalaja, P., & Barcelos, A. M. F. (Eds.), *Beliefs about SLA: New Research Approaches*, pp. 87-108. Dordrecht: Kluwer Academic Publishers.
- Karavas-Doukas, E. (1996). Using attitude scales to investigate teachers' attitudes to the communicative approach. *English Language Teaching Journal*, 50 (3), 187-198.
- Keiny, S. (1994). Teachers professional development as a process of conceptual change. In I. Carlgren, G. Handal, & S. Vaag (Eds.). *Teachers minds and*

- actions. *Research on teachers' thinking and practice*, pp. 232–246. London: Falmer Press.
- Kennedy, C. (1987). Innovating for a change: Teacher development and innovation. *English Language Teaching Journal*, 41 (3), 163-170.
- Kennedy, C. (1988). Evaluation of the management of change in ELT project. *Applied Linguistics*, 9 (4), 329-342.
- Kim, J. O. & Mueller, C.W. (1978). *Introduction to factor analysis: What it is and how to do it*. Newbury Park. CA: Sage University Press.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling*. (2<sup>nd</sup> ed.). New York: The Guilford Press.
- Kong, Y. & Nie, J (2002). The compound dictation in the CET and its washback effect of on teaching and learning. *Foreign Language World*, 88 (2), 51-57.
- Larsen-Freeman, Diane. 1986. *Techniques and principles in language teaching*. Oxford: Oxford University Press.
- Leng, H. (1997). New bottles, old wine: Communicative language teaching in China. *English Teaching Forum*, 35 (4), 38-41.
- Lincoln, Y. S. & Guba, E. G. (1985). *Naturalistic Inquiry*. Beverly Hill, CA: Sage.
- Liu, D. (1995). *English Language Teaching In Schools In China* . Amity Foundation Teachers Project – Teaching Resources in The Amity Teacher Project Handbook IV.
- Liu, R. (2002). Issues on the Reform of English Curriculum – from a Discrete Curriculum to An Uniformed Curriculum requirement. *Foreign Language Teaching and Research*, 6, 403-404.
- Liu, X. Y. (1988). An analysis of factors that hinder TESOL innovations in China. *TESL Reporter*, 21, 70-77.
- Liu, R. & Dai, M. (2003). *The Status Quo of Chinese College English Teaching Reform and Research on Its Development Strategies* [M]. Beijing: Foreign Language Teaching and Research Press.
- Long, M.H. & Sato, C.J. (1983) Classroom foreigner talk discourse: forms and functions of teachers' questions. In H.W. Seliger & M.H. Long (Eds.). *Classroom Oriented Research in Second Languages*. Rowley, MA: Newbury House.
- Long, M. H. (1985). *Input and Second Language Acquisition Theory*. In S. M. Gass & C. G. Madden (eds.), *Input in Second Language Acquisition*, pp. 377-393. Rowley, MA: Newbury House.

- Long, M. H. (1996). The Role of Linguistic Environment In Second Language Acquisition. In W.C. Ritchie & T. K. Bhatia (eds.), *Handbook of Second Language Acquisition*, pp. 413-468. San Diego: Academic Press.
- Luo, X. & Xiao, Y. (2002). On the Reform of the CET. *Foreign Language Teaching and Research*, 34 (4), 294-299.
- Ma, Z. 1991. *Aspects of Chinese Culture*. Tianjin: Nankai University Press.
- Ma, Y. (2000) Some thoughts on the CET reform. *Higher Education in Building Material Science*, 19 (3), 93-96.
- Ma, Y. & Jin, H. (2000). CET's misguidance to College English teaching. *Journal of Architectural Education in Institutions of Higher Learning*, 34 (2), 54-55.
- McNamara, T. (2000). *Language Testing*. Oxford, Oxford University Press.
- Markee, N. (1997). *Managing Curriculum Innovation*. Cambridge: Cambridge University Press.
- Maxwell, J.A. (1996). *Qualitative Research Design - An Integrative Approach*. Thousand Oaks, London: Sage.
- Maykut, P. & Morehouse, R. (1994). *Beginning Qualitative Research: A Philosophic and Practical Guide*. London: The Falmer Press.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 1 (23), 13-23.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13 (3), 241-256.
- Miles, M., & Huberman, A. (1994). *Qualitative data analysis: A sourcebook of new methods* (2<sup>nd</sup> ed.). Beverly Hills, CA: Sage.
- Mitchell, R. (1988). *Communicative Language Teaching in Practice*. London: CILT.
- Newman, I. (2000). A conceptualization of mixed methods: A need for an inductive/deductive approach to conducting research. *Journal of Research in Education*, 10 (1), 81-84.
- Muñoz, A. & Álvarez, M. (2010). Washback of an oral assessment system in theEFL classroom. *Language Testing*, 27 (1), 33-49.
- Nespor, J. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies*, 19 (4), 317-328.
- Newman, I. (2000). A conceptualization of mixed methods: A need for an

- 
- inductive/deductive approach to conducting research. *Journal of Research in Education*, 10 (1), 81-84.
- Niu, Q. (2001). Problems with the current College English Tests in China. *Foreign Language Teaching and Research (bimonthly)*, 33 (2), 140-143.
- Noble, A. J. & Smith, M. L. (1994a). *Old and new beliefs about measurement-driven reform: "The more things change, the more they stay the same"* (CSE Tech. Rep. No. 373). Los Angeles: UCLA, National Center for Research on Evaluation, Standards, and Student Testing.
- Noble, A. J., & Smith, M. L. (1994b). *Measurement-driven reform: Research on policy, practice, repercussion* (CSE Tech. Rep. No. 381). Los Angeles: UCLA, National Center for Research on Evaluation, Standards, and Student Testing.
- Nunan, D. (1987). Communicative language teaching: Making it work. *English Language Teaching Journal*, 41, 136-145.
- Nunan, D. (1992). *Research methods in language learning*. Cambridge: Cambridge University Press.
- Nunan, D. (1997). Language teaching and research. In D. Griffee & D. Nunan (Eds.), *Classroom teachers and classroom research*, pp. 13-22. Tokyo: The Japan Association for Language Teaching (JALT).
- Nunan, D. (1999). *Second language teaching and learning*. Boston: Heinle & Heinle.
- Nunan, D. (2001). Tradition and change in the ELT curriculum. In W. Hu (Eds), *Foreign Language Teaching and Research*, pp. 50-58. Beijing: Foreign Language Teaching and Research Press.
- Oxenham, J. (1984). *Education versus qualifications?* London: Allen and Unwin.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62 (3), 307-332.
- Patton, M.Q. (1990). *Qualitative Evaluation and Research Methods* (2<sup>nd</sup> ed.). Newbury Park, CA: Sage.
- Patton, M.Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.
- Pearson, I. (1988). *Tests as levers of change*. In D. Chamberlain & R.J.Baumgartner (Eds.), *ESP in the classroom: Practice and evaluation* (pp. 98-107). ELT Documents #128. London: Modern English Publications in association with the British Council.
- Popham, W. J. (1987). The merits of measurement driven instruction. *Phi Delta Kappan*, 68, 679-682.

- Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Towards a theory of conceptual change. *Science Education*, 66 (2), 211-227.
- Prodromou, L. 1995. The backwash effect: from testing to teaching. *ELT Journal* 49 (1): 13-25.
- Purpura, J.E. (1998) Investigating the Effects of Strategy Use and Second Language Test Performance with High- and Low-Ability Test Takers: a Structural Modelling Approach, *Language Testing*, 15 (3), 333-379.
- Purpura, J.E. (1999) *Learner Strategies and Performance on Language Tests: a Structural Equation Modeling Approach*, Cambridge University Press, Cambridge.
- 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 Associates.
- Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing*, 22 (2), 142-173.
- Qi, L. (2007) 'Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes English test in China', *Assessment in Education: Principles, Policy & Practice*, 14 (1), 51-74.
- Qian, J. (2003). On the Impeding Factors of CET Reform and Development. *Journal of Zhongzhou University*, 20 (4), 91-92.
- Rao, Z. (2001). Chinese students' perceptions of communicative and non-communicative activities in EFL classroom. *System*, 30, 85-105.
- Reagon, T. G. & Osborn, T. A. (2002). *The Foreign Language Educator in Society. Toward a Critical Pedagogy*. (1<sup>st</sup> ed.). Mahwah, New Jersey: Lawrence Erlbaum Sociates, Publishers.
- Reichardt, C.S. & Cook, T.D. (1979). Beyond qualitative versus quantitative methods. In T.D. Cook & C.S. Reichardt (Eds.), *Qualitative and quantitative methods in evaluation research*. Beverly Hills, London: Sage.
- Rea-Dickens, P. (2004). Editorial: Understanding teachers as agents of assessment. *Language Testing*, 21(3), 249 -258.
- Richards, J. (2008). Second Language Teacher Education Today. *RELC Journal*, 39 (2), 158-177.
- Richards, J. C., & Farrell, T. (2005). *Professional development for language teachers: strategies for teacher learning*. Cambridge: Cambridge University Press.

- Richards, J. & Lockhart, C. (1994). *Reflective teaching in second language classrooms*. New York: Cambridge University Press.
- Richards, J., & Rodgers, T. (2001). *Approaches and methods in language Teaching* (2nd ed.). Cambridge: Cambridge University Press.
- Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. In J. Sikula, T. J. Buttery, & E. Guyton (Eds.), *Handbook of research on teacher education* (2nd ed.), pp. 102-119. New York: Macmillan.
- Saif, S. (2006). Aiming for positive washback: A case study of international teaching assistants. *Language Testing*, 23 (1), 1-34.
- Samuelowicz, K., and Bain, J.D. (1992). Conceptions of teaching held by academic teachers, *Higher Education*, 24, 93–111.
- Samuelowicz, K., and Bain, J.D. (2001). Revisiting Academics' Beliefs about Teaching and Learning, *Higher Education*, 41, 318.
- Sato, K., & Kleinsasser, R. C. (2004). Beliefs, practices, and interactions of teachers in a Japanese high school English department. *Teaching and Teacher Education*, 20, 797-816.
- Savignon, J. S. (1983). *Communicative Competence Theory and Classroom Practice: Texts and Contexts in Second Language Learning*. (1st. ed). Massachusetts: Addison- Wesley Publishing Company.
- Savignon, S. (1991). Communicative Language Teaching: State of the art. *TESOL Quarterly*, 25, 261-277.
- Savignon, S. (2003). Teaching English as communication: a global perspective. *World Englishes*, 22, 55–66.
- Savignon, S. (2005). Communicative Language Teaching: Strategies and goals. In E. Hinkel (ed.), *Handbook of research in second language teaching and learning*, pp. 635-651.
- Shavelson, R.J., & Stern, P. (1981). Research on teachers' pedagogical thoughts, judgements, decisions, and behavior. *Review of Education Research*, 51(4), 455-498.
- Shepard, L.A. (1991). Psychometricians' beliefs about learning. *Educational Researcher*, 20 (6), 2–16.
- Shepard, L.A. (1993). The place of testing reform in educational reform: A reply to Cizek. *Educational Researcher*, 22 (4), 10–14.
- Shih, C. (2007). A new washback model of students' learning. *Canadian Modern Language Review*, 65(1), 135 -162.

- Shohamy, E. (1993). *The power of tests. the impact of language tests on teaching and learning*. Washington, DC: NFLC Occasional Papers.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13 (3), 298-317.
- Shu, D. (2004). *FLT in China: Problems and Suggested Solutions*. Shanghai: Shanghai Foreign Language Education Press.
- Shulman, L.S. (1986a). Paradigms and research programmes in the study of teaching: a contemporary perspective, in M.C. Wittrock (Ed.) *Handbook of Research on Teaching*, pp. 505-526, (3rd ed.). New York: Macmillan.
- Shulman, L. (1986b) Those who understand: knowledge growth in teaching, *Educational Researcher*, 15, 4-14.
- Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Education Review*, 57 (1), 1-21.
- Skehan, P. (1998). Task-based instruction. *Annual Review of Applied Linguistics*, 18, 268-286.
- Smith, M.L.(1991). Put to the test: The effects of external testing on teachers. *Educational Researcher* 20 (5), 8–11.
- Smith, M.L., Noble, A.J., Cabay, M., Heinecke, M., Junker, M.S. & Saffron, Y. (1994). *What happens when the test mandate change?* Results of a multiple case study. CRESST Report 380. Arizona, USA: Arizona State University.
- Solomon, P. (2002). *The assessment bridge: Positive ways to link tests to learning, standards, and curriculum improvement*. Thousand Oaks, CA: Corwin.
- Spada, N. & Fröhlich, M. (1995). *The Communicative Orientation of Language Teaching Observation Scheme (COLT)*. The National Centre for English Language Teaching and Research (NCELTR). Macquarie University, Sydney, Australia.
- Stern, H.H. (1992). *Issues and options in language teaching*. Oxford: Oxford University Press.
- Stobart, G. (2003). The impact of assessment: Intended and unintended consequences. *Assessment in Education*, 16 (2), 139-140.
- Strauss, A., & Corbin, J., (1990): *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. London: Sage.
- Strauss, A. & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2<sup>nd</sup> ed.). Thousand Oaks: Sage.

- Swain, M. (1995). Three functions of output in second language learning. In G. Cook and B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H.G. Widdowson*, pp. 125-144. Oxford: Oxford University Press.
- Tan, M. (2008). *Bilingual high-stakes mathematics and science exams in Malaysia: pedagogical and linguistic issues*. Paper presented at the Language Testing and Research Colloquium 2008, Hangzhou.
- Tarnopolsky, O.; Karajeva, T. and Zhaing, X. (2001). Teaching English as a Foreign Language in China: Teachers and Students' Attitudes to the Communicative Approach, Learner-Centering, and Learners' Autonomy. *Journal of Shandong Agricultural University (Social Science)*, 3 (2), 89-94.
- Tatto, M. T. (1998). The influence of teacher education on teachers' beliefs about purposes of education, roles, and practice. *Journal of Teacher Education*, 49 (1), 66-77.
- Tashakkori, A. & Teddlie, C. (1998). *Mixed Methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.
- Tavares, N. & Hamp-Lyons, L. (2008). *ESL teachers learning to assess interactively: The challenges and teacher development issues*. Presented at 30th Annual Language Testing Colloquium, (LTRC 08). Hangzhou.
- Teddlie, C. & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative techniques in the social and behavioral sciences*. Thousand Oaks, CA: Sage.
- Thompson, A. (1992). Teachers' beliefs and conceptions: A synthesis of research. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning*, pp. 127-146. New York: Macmillan.
- Ting, Y. R. (1987). Foreign Language Teaching in China: Problems and Perspectives. *Canadian an International Education*, 16, 48-61.
- Turner, C.E. (2001). The need for impact studies of L 2 performance testing and rating: Identifying areas of potential consequences at all levels of the testing cycle. In A. Brown et al. (Eds.), *Experimenting with uncertainty: Language testing essays in honour of Alan Davies*, pp.138-149. Cambridge, Cambridge University Press.
- Turner, C.E. (2002). *Investigating high-stakes test impact at the classroom level*. Presented at 24th Annual Language Testing Colloquium, (LTRC '02). Hong Kong.
- Turner, C.E. (2005). Professionalism and high-stakes tests: Teacher perspectives when dealing with educational change introduced through provincial exams. *TESL Canada Journal*, 23 (2), 54-76.



- Turner, C.E. (2008, June). *The specificity of the “research approach” in classroom studies: Probing the predictability of washback through teacher conceptual and instrumental evidence in Quebec high schools*. Paper presented at the Language Testing and Research Colloquium 2008, Hangzhou.
- Turner, C.E. (2009). Examining washback in second language education contexts: A high stakes provincial exam and the teacher factor in classroom practice in Quebec secondary schools. *International Journal of Pedagogies and Learning*, 5(1), 103-123.
- Urmston, A. (2003). 'Learning to Teach English in Hong Kong: The Opinions of Teachers in Training'. *Language and Education*, 17 (2), 112-137
- Urmston, A. & Fang, F. (2008). *Stakeholder involvement in the language assessment evaluation process: A Hong Kong case study*. Presented at 30th Annual Language Testing Colloquium, (LTRC 08). Hangzhou.
- Vernon, P.E. (1956). *The measurement of abilities* (2<sup>nd</sup> ed.). London: University of London Press.
- Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.
- Wall, D. (1996). Introducing new tests into traditional systems: Insights from general education and from innovation theory. *Language Testing*, 13, 334-354.
- Wall, D. (1997). Impact and washback in language testing. In C. Clapham & D. Corson (Eds.), *Language Testing and Assessment, Encyclopedia of Language and Education*, Vol. 7 (Vol. 7). Dordrecht: Kluwer.
- Wall, D. (1999). *The impact of high-stakes examinations on classroom teaching: A case using insights from testing and innovation theory*. Unpublished doctoral dissertation, University of Lancaster, England.
- 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.
- Wall, D. and Horák, T. (2007) 'Using baseline studies in the investigation of test impact', *Assessment in Education: Principles, Policy & Practice*, 14 (1), 99-116
- Wang, J. (2008). *Identifying the main impact on teacher classroom behaviour: Examining the effects of the CET, teacher beliefs, teacher knowledge bases and teacher proficiency*. Paper presented at the Language Testing and Research Colloquium 2008, Hangzhou.
- Watanabe, Y. (1996a). Does grammar translation come from the entrance

- examination? Preliminary findings from classroom-based research. *Language Testing*, 13 (3), 318-333.
- Watanabe, Y. (1996b). Investigating washback in Japanese EFL classrooms: Problems of methodology. In G. Wigglesworth & C. Elder (Eds.), *The language testing circle: From inception to washback*, pp. 208-239. Melbourne, Victoria, Australia: Applied Linguistics Association of Australia.
- Watanabe, Y. (2004a). 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 Associates.
- Watanabe, Y. (2004b). Teacher factors mediating washback. In L.Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods*, pp.129-146. Mahwah, NJ: Lawrence Erlbaum Associates.
- Watson-Gegeo, K. A. (1988). Ethnography in ESL: defining the essentials. *TESOL Quarterly*, 22 (4), 575-592.
- Wesche, M. and P. Skehan (2002). Communicative, task-based and content-based language instruction. In R. Kaplan *The Oxford Handbook of Applied Linguistics*. Oxford: Oxford University Press.
- Williams, M. & Burden, R. (1997). *Psychology for Language Teachers*. New York: Cambridge University Press.
- White, R. (1987). Managing innovation. *English Language Teaching Journal*, 41 (3), 211-218.
- Widen, M.F., O'Shea, T., & Pye, I. (1997). High-stakes testing and the teaching of science. *Canadian Journal of Education*, 22, 428-444.
- Winitzky, N., & Kauchak, D. (1997). Constructivism in teacher education. Applying cognitive theory to teacher learning. In V. Richardson (Ed.), *Constructivist teacher education. Building a world of new understandings* (pp. 59-84). London: Falmer.
- Woods, D. (1996). *Teacher Cognition in Language Teaching*. Cambridge: Cambridge University Press.
- Woods, D. (2003). The social construction of beliefs in the language classroom. In Kalaja, P., & Barcelos, A. M. F. (Eds.), *Beliefs about SLA: New Research Approaches*. Dordrecht: Kluwer Academic Publishers.
- Woods, D. (2006) Who Does What in The 'Management of Language Learning'? Planning and the Social Construction of the 'Motivation to Notice'. In Gieve, S. and I. Miller (eds). *Understanding Language Classroom*. London: Palgrave Macmillan.

- Wu, Q. (2005). Speech made at the press conference on the upcoming reform of the CET 4/6. *Foreign Language World*, 2, 2-4.
- Wu, Sh. (2003). College English Teaching and Learning: Problems and Solutions. *Foreign Language Teaching and Research*, 3, 226-228.
- Wu, Y. (2001). English Language Teaching in China: Trends and Challenges. *TESOL Quarterly*, 35 (1), 191-194.
- Xia, J. (2002). A report on investigating the College English teachers' knowledge of and beliefs in foreign language education and their roles in teaching, researching and professional development. *Foreign Language World*, 5, 35-41.
- Xia, J. (2005). Interpretation of College English Curriculum Requirements (For Trial Implementation). *Foreign Language World*, 2, 12-14
- Xu, Ch. & Gao, P. (2001). Thoughts on the CET 4/6. *Journal of Nantong Teachers College (Philosophy & Social Science Edition)*. 17 (1), 134-136.
- Yang, H. (1999). Language Testing and Language Teaching. *Foreign Language World*, 73 (1), 16-25.
- Yang, H. (2000). A look at the actual language competence of Chinese students through the CET 4/6. *Foreign Language World*, 77 (1), 46-52.
- Yang, H. (2003). The 15 years of the CET and its impact on teaching. *Journal of Foreign Languages*. 145 (3), 21-29.
- Yang, H. & Jin, Y. (2001). An interpretation of the CET 4/6 scores. *Foreign Language World*, 81 (1), 62-68.
- Yang, H. and Weir, C.J. (1998) *The Validity of the College English Test*, Shanghai Foreign Language Education Press.
- Yates & Muchisky. (2003) On Reconceptualizing Teacher Education. *TESOL Quarterly*, 37 (1), 135-147.
- Ye, J (1998, a). An investigation and reflection on the washback effects of the CET. *Journal of Shanghai Jiaotong University*, 12 (2), 132-136.
- Ye, J (1998, b). An investigation and reflection on the washback effects of the CET. *Foreign Language World*, 71 (3), 40-43.
- Yu, H. (1999). What kind of personnel of foreign languages are needed in the 21st Century? – An issue beginning with the negative effects of the CET-4 and CET-6. *Journal of Jinling Polytechnic College*, 14 (4), 33-36.
- Zhang, Sh. (2003). Separating teaching from testing – a must for the CET innovation. *Foreign Language Teaching and Research*, 5, 385-386.

Zhang, X. (2004). Comment on the doubts about the CET. *Foreign Language World*, 100 (2), 65-69.

## APPENDIX A –English Version of the CECR

### College English Curriculum Requirements

(For Trial Implementation)

With a view to keeping up with the new developments of higher education in China, deepening teaching reform, improving teaching quality, and meeting the needs of the country and society for qualified personnel in the new era, ***College English Curriculum Requirements*** (*Requirements* hereafter) is drawn up, providing colleges and universities with the guidelines for English instruction to non-English major students.

Because institutions of higher learning differ from each other in terms of teaching resources, students' level of English upon entering college, and the social needs they face, colleges and universities should formulate, in accordance with the ***Requirements*** and in the light of their specific circumstances, a scientific, systematic and individualized College English syllabus to guide their own College English teaching.

#### I. Character and Objective of College English

College English, an integral part of higher learning, is a required basic course for undergraduate students. As a systematic whole, College English has as its main components knowledge and practical skills of the English language, learning strategies and intercultural communication; it takes theories of foreign language teaching as its guide and incorporates different teaching models and approaches.

The objective of College English is to develop students' ability to use English in an all-round way, especially in listening and speaking, so that in their future work and social interactions they will be able to exchange information effectively through both spoken and written channels, and at the same time they will be able to enhance their ability to study independently and improve their cultural quality so as to meet the needs of China's social development and international exchanges.

#### II. Teaching Requirements

As China is a large country with conditions varying from region to region and from college to college, the teaching of College English should follow the principle of providing different guidance for different groups of students and instructing them in accordance with their aptitude so as to meet the specific needs of the individualized teaching.

The requirements for undergraduate College English teaching are set at three levels, i.e., basic requirements, intermediate requirements, and higher requirements. All non-English majors are required to attain to one of the three levels of requirements after studying and practicing English at school. The basic requirements, a goal that all college graduates must achieve, are meant for students who have or have not completed Band 7 of the ***Senior High School English Standards*** prior to entering college. Intermediate and higher requirements are respectively set for those who, having laid a good foundation of English, can afford time to learn more of the language, and have completed Bands 8 or 9 of the ***Senior High School English Standards*** upon entering college. The three levels of requirements, which incorporate

## **APPENDIX A –English Version of the CECR**

knowledge and practical skills of the English language, learning strategies and intercultural communication, embody qualitatively and quantitatively the objective of College English teaching.

The basic requirements are the minimum level that all non-English majors have to reach before graduation. Institutions of higher learning should set their own objectives in the light of their specific circumstances, strive to create favorable conditions, and encourage students to adjust their objectives in line with their own performance and try to meet the intermediate or higher requirements.

### **The three levels of requirements are set as follows:**

#### **Basic requirements**

1. Listening: Students should be able to follow classroom instructions, everyday conversations, and lectures on general topics conducted in English. They should, by and large, be able to understand Special English programs spoken at a speed of about 130 words per minute (wpm), grasping the main ideas and key points. They are expected to be able to employ basic listening strategies to facilitate comprehension.
2. Speaking: Students should be able to communicate in English in the course of learning, to conduct discussions on a given theme, and to talk about everyday topics with people from English-speaking countries. They should be able to give, after some preparation, short talks on familiar topics with clear articulation and basically correct pronunciation and intonation. They are expected to be able to use basic conversational strategies in dialogue.
3. Reading: Students should be able to read, in the main, English texts on general topics at a speed of 70 wpm. With longer yet less difficult texts, the reading speed should be at 100 wpm. They should be able to read, in the main, English newspapers and magazines published in China, grasping the main ideas, and understanding major facts and relevant details. They should be able to understand texts of practical styles commonly used at work and in life. They are expected to be able to employ effective reading strategies while reading.
4. Writing: Students should be able to complete writing tasks for general purposes, e.g., describing personal experiences, impressions, feelings, or some events, and to undertake practical writing. They should be able to write within 30 minutes a short composition of 120 words on a general topic or an outline. The composition should be basically complete in content, appropriate in diction and coherent in discourse. Students are expected to be able to have a command of basic writing strategies.
5. Translation: With the help of dictionaries, students should be able to translate essays on familiar topics from English into Chinese and vice versa. The speed of translation from English into Chinese should be 300 English words per hour whereas the speed of translation from Chinese into English should be 250 Chinese characters per hour. The translation should read smoothly. Students are expected to be able to use appropriate translation techniques.
6. Recommended Vocabulary: Students should acquire a total of 4,500 words and 700 phrases (including those that have been covered in high school English courses), among which 2,000 are active words (see Appendix III: Active Word List). Students should not only be able to comprehend the active words but be proficient in using them when expressing themselves in speaking or writing.

---

## APPENDIX A –English Version of the CECR

### Intermediate requirements:

1. Listening: Students should be able to follow, in the main, talks and lectures by people from English-speaking countries, to understand longer English radio and TV programs produced in China on familiar topics spoken at a speed of around 150 wpm, grasping the main ideas, key points and relevant details. They should be able to understand, by and large, course in their areas of specialty taught by foreign teachers in English.
2. Speaking: Students should be able to hold conversations in fairly fluent English with people from English-speaking countries, and to employ fairly well conversational strategies. They should, by and large, be able to express their personal opinions, feelings and views, and to state facts, events and reasons with clear articulation and basically correct pronunciation and intonation.
3. Reading: Students should, in the main, be able to read essays on general topics in newspapers and magazines published in English-speaking countries at a speed of 80 wpm. With longer texts for fast reading, the reading speed should be 120 wpm. Students should be able to skim or scan reading materials. When reading summary literature in their areas of specialty, students should be able to get a correct understanding of the main ideas, major facts and relevant details.
4. Writing: Students should be able to express personal views on general topics, compose English abstracts of theses in their own specialization, and write short English papers on topics of their specialty. They should be able to describe charts and graphs, and to complete within 30 minutes a short composition of 160 words. The composition should be complete in content, clear in organization and coherent in discourse.
5. Translation: With the help of dictionaries, students should be able to translate texts on familiar topics in newspapers and magazines published in English speaking countries, to translate on a selective basis articles of popular science relevant to their own specialty. The speed of translation from English into Chinese should be 350 English words per hour whereas the speed of translation from Chinese into English should be 300 Chinese characters per hour. The translation should read smoothly, convey the original meaning and be free from serious mistakes in understanding or expression.
6. Recommended Vocabulary: Students should acquire a total of 5,500 words and 1,200 phrases (including those that have been covered in high school English courses and the Basic Requirements), among which 2, 500 are active words (including the active words that have been covered in the Basic Requirements). (see Appendix III: Active Word List)

### Higher Requirements:

1. Listening: Students should be able to understand longer dialogues and passages, and grasp the key points even when sentence structures are complicated and views are only implied. They should, by and large, be able to understand radio and TV programs produced in English-speaking countries. They should be able to understand lectures related to their areas of specialty and grasp the gist and main points.
2. Speaking: Students should be able to conduct dialogues or discussions with certain

## APPENDIX A –English Version of the CECR

degree of fluency and accuracy on general or specialized topics, and to make concise summaries of extended texts or speeches in difficult language. They should be able to deliver papers at academic conferences and participate in discussions.

3. Reading: Students should be able to read rather difficult texts, and understand their meanings. With the help of dictionaries, they should be able to read original versions of English textbooks and articles in newspapers and magazines published in English-speaking countries, and to read literature related to their areas of specialty without much difficulty.

4. Writing: Students should be able to express their opinions freely on general topics with clear structure, rich content and good logic. They should be able to write brief reports and papers of their areas of specialty, and to write within 30 minutes expository or argumentative essays of 200 words on a given topic. The text has complete content, logical thinking, and clear expression of ideas.

5. Translating: With the help of dictionaries, students should be able to translate fairly difficult English texts on popular science, culture, and reviews in newspapers and magazines published in English-speaking countries into Chinese, and translate Chinese introductory texts on the conditions of China or Chinese culture into English. The speed of translation from English into Chinese should be 400 English words per hour whereas the speed of translation from Chinese into English should be 350 Chinese characters per hour. The translation should convey the idea with accuracy and smoothness and be basically free from mistakes and misinterpretation.

6. Recommended Vocabulary: Students should acquire a vocabulary of 6,500 words and 1,700 phrases, among which 2,500 are active words (including the active words that have been covered in the Basic Requirements and Intermediate Requirements). (See Appendix III: Active Word List)

In developing competence in listening, speaking, reading, writing and translation at the three levels mentioned above, college and universities should lay more stress on the cultivation and training of listening and speaking abilities. A good command of vocabulary, especially of active words, constitutes the basis for the improvement of students' ability to use English in an all-round way. Therefore, teaching plan for this component should be specified in the College English syllabus of each school.

Moreover, colleges and universities should cover components of learning strategies and intercultural communication in their teaching so as to enhance students' abilities of independent learning and of communication.

### III. Course Designing

Colleges and universities should take into account the school's circumstances and follow the guidelines of the **Requirements** in setting the goals of their College English teaching and designing College English course systems. The course system, which is a combination of required and selective courses in comprehensive English, language skills, English for practical uses, language and culture, and English of specialty, should ensure that students at different levels receive adequate training and make improvement in their ability to use English.

In designing College English courses, requirements for competence in listening and speaking should be fully considered, and corresponding teaching hours and credits should be adequately allocated. Moreover, the extensive use of advanced information technology should be encouraged, computer-and Web-based English teaching should



## **APPENDIX A –English Version of the CECR**

be promoted, and students should be provided with favorable environment and facilities for language learning. With regard to computer-based courses, face-to-face coaching should be fully provided, and the hours spent on coaching should be counted in the teachers' teaching load. Students should receive credits after completing courses they learn via computers and passing the exams.

College English course designing should give a full play to the strengths of traditional classroom teaching, and encourage excellent teachers to offer courses suited to classroom teaching, thus forming a combination with computer-and Web-based courses.

College English is not only a language course that provides basic knowledge about English, but also a capacity enhancement course that helps students to broaden their horizons and learn about different cultures in the world. When designing College English courses, therefore, it is necessary to take into full consideration the development of students' cultural capacity and the teaching of knowledge about different cultures in the world.

All the courses, whether computer-based or classroom-based, should be fully individual-oriented, taking into account students with different starting points, so that students who start from lower level will be taken good care of while students whose English is better will find room for further development. College English course designing should enable students to have a solid foundation in the English language while developing their ability to use English, especially their ability to listen and speak in English. It should ensure that students make steady progress in English proficiency throughout their undergraduate studies and make it conducive to students' individualized learning so as to meet the needs of their development in different specialties.

### **IV. Teaching Model**

In view of the marked increase in student enrolments and the relatively limited resources, colleges and universities should remould the existing unitary teacher-centered pattern of language teaching by introducing new teaching models with the help of multimedia and network technology. The new model should be built on modern information technology, particularly network technology, so that English language teaching will be free from the constraints of time or place and geared towards students' individualized and autonomous learning. The new model should combine the principles of practicality, knowledge and interest, mobilize the initiative of both teachers and students, and attach particular importance to the central role of students in the teaching and learning process. This model should technically attain to a high level of interactivity, feasibility and operability. In addition, it should take into full account and incorporate into it the strengths of the current model while fully employing modern information technology.

Colleges and universities should each design a computer-based or Intranet-or campus-network-based multimedia listening and speaking teaching model that suits their own needs in line with their own conditions and student situation. Those in more favorable situations may deliver listening and speaking course via the Internet. The teaching of reading, writing and translation can be conducted either in the classroom or online. In either case, however, enough teachers should be guaranteed for

## APPENDIX A –English Version of the CECR

instructions or lectures.

An important indicator of the successful reform of the teaching model is the development of individualized study methods and the autonomous learning ability on the part of students. The new model should enable students to select materials suited to their individual needs, make up for the limitations of the conventional classroom teaching of listening and speaking, and track down, record and check the progress of learning as well as teaching and coaching, so that students will be supervised to learn on their own initiative, and their listening and speaking abilities and other language abilities will be improved at a faster pace.

For effective reforms of the teaching model, the proportion of teaching hours or credits for computer-based English courses should be increased. It is proposed that the credits acquired via computer-based learning account for 30%~50% of the total. Changes in the teaching model by no means call for changes in teaching practices or approaches only, but, more important, consist in changes in teaching philosophy, and in a shift from the teacher-centered pattern, in which knowledge of the language and skills are imparted by the teacher in class only, to the student-centered pattern, in which the ability to use the language and the ability to learn independently are cultivated in addition to language knowledge and skills.

For the implementation of the new model, refer to Appendix I: Computer-and Classroom-based Multimedia College English Teaching Model.

### V. Evaluation

Evaluation is a key component in College English teaching. A comprehensive, objective, scientific and accurate evaluation system is of vital importance to the achievement of the course goal. It not only helps teachers obtain feedback, improve the administration of teaching, and ensure teaching quality but also provides students with an effective means to adjust their learning strategies and methods, and improve their learning efficiency.

Evaluation consists of formative assessment and summative assessment.

Formative assessment includes students' self-assessment, peer assessment, and assessment conducted by teachers and school administrators. By keeping a record of students' in and outside of classroom activities and online self-learning data, keeping files on students' study results, conducting interviews and holding meetings, students' learning process is under observation, evaluation and supervision, thus contributing to the enhancement of their learning efficiency. Formative assessment is particularly important in computer-based teaching which is characterized by students' independent learning. (See recommended assessment forms in Appendix II)

Summative assessment refers to final tests and proficiency tests. These tests are designed to assess student's all-round ability to use English. Students' scores in reading, writing and translation should not substitute for or offset the scores in listening and speaking, or vice versa.

To evaluate the results of the set goal, colleges and universities may administer tests of their own, run tests at the intercollegiate or regional level, or let students take the national test in accordance with the different requirements set by the

**Requirements.** Whatever form the tests may take, the focus should be on the

## APPENDIX A –English Version of the CECR

assessment of students' ability to use English in communication, particularly their ability to speak and write in English.

Government education administrative offices at different levels and colleges and universities should regard the evaluation of College English teaching as an important part of the evaluation of the overall teaching quality of each school.

### VI. Teaching Administration

A system of teaching administration should be established that is accountable for the whole process of College English teaching. To ensure that the set teaching objectives can be achieved, efforts should be made to strengthen the guidance for and supervision of the teaching process. For this purpose, the following measures should be taken.

#### **1. A system of teaching and teaching administration documentation should be established.**

Documents of teaching include College English Curriculum of the colleges and universities concerned, as well as the documents stipulating the teaching objectives, course description, teaching arrangement, teaching progress, and methods of assessment for all the courses within the program. Documents of teaching administration include documents registering students' status and their academic credits, regulations of assessment, and guidelines for teaching.

#### **2. Efforts should be made to promote the credit system.**

College English program should adapt itself to the overall credit system of the colleges and universities concerned and should account for 10% (around 16) of the total undergraduate credits. The credits students acquire by computer-based independent learning should be equally acknowledged.

#### **3. A system of faculty management and development should be established.**

The quality of teachers is the key to the success of College English teaching reform and to the long-term development of the discipline. Colleges and universities should lay emphasis on the training and development of College English teachers so that they can better adapt to the new model of English teaching. Meanwhile, opportunities should be created so that the teachers can enjoy sabbaticals and engage in advanced studies, thus ensuring a sustainable improvement in their academic performance and methods of teaching.

## **APPENDIX B – Announcement of Postponement of the Administration of the New CET by the NCETC (October, 21, 2005)**

### *Summary of the Main Content of the Following Announcement Concerning CET 4:*

The administration of the new CET among pilot-school students will be postponed from its originally scheduled time – December, 2005 to June, 2006. The large-scale administration of the new CET originally scheduled on December, 2006 remains unchanged.

The original announcement (in Chinese ) is provided below.

#### 关于大学英语四、六级试点考试的通知

接教育部高教司通知(教高司函[2005]199号),为了使高校教师和学生更充分熟悉和了解新的考试要求和题型,积极稳步推进改革,对原定的试点考试时间等事宜做相应调整,详细内容见下面附件 WORD 文件,请大家下载阅读。

全国大学英语四、六级考试委员会

### **关于大学英语四、六级试点考试的通知**

接教育部高教司通知(教高司函[2005]199号),为了使高校教师和学生更充分熟悉和了解新的考试要求和题型,积极稳步推进改革,对原定的试点考试时间等事宜做相应调整,现将有关事项通知如下:

一、原定于 2005 年 12 月进行的大学英语四级试点考试调整至 2006 年 6 月进行。2006 年 12 月仍按原计划全面实施新题型四级考试,时间不变。

二、原定于 2006 年 6 月进行的大学英语六级试点考试调整至 2006 年 12 月进行。2007 年 6 月仍按原计划全面实施的新题型六级考试,时间不变。

三、有关 2006 年 6 月四级试点考试和 2006 年 12 月六级试点考试的具体要求、参加试点考试高校范围等具体事宜,届时将另行通知。

全国大学英语四、六级考试委员会

二〇〇五年十月二十一日

Appendix C: A Sample of the Revised CET

答题卡 1 (Answer Sheet 1)

学校:		准 考 证 号													
姓名:		[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]
		[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]
划线要求		[2]	[2]	[2]	[2]	[2]	[2]	[2]	[2]	[2]	[2]	[2]	[2]	[2]	[2]
		[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]
		[4]	[4]	[4]	[4]	[4]	[4]	[4]	[4]	[4]	[4]	[4]	[4]	[4]	[4]
		[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]
		[6]	[6]	[6]	[6]	[6]	[6]	[6]	[6]	[6]	[6]	[6]	[6]	[6]	[6]
		[7]	[7]	[7]	[7]	[7]	[7]	[7]	[7]	[7]	[7]	[7]	[7]	[7]	[7]
		[8]	[8]	[8]	[8]	[8]	[8]	[8]	[8]	[8]	[8]	[8]	[8]	[8]	[8]
		[9]	[9]	[9]	[9]	[9]	[9]	[9]	[9]	[9]	[9]	[9]	[9]	[9]	[9]

**Part I Writing (30 minutes)**  
Directions: For this part, you are allowed 30 minutes to write a campaign speech in support of your election to the post of chairman of the student union. You should write at least 120 words following the outline given below in Chinese:

1. 你认为自己具备了什么条件（能力、性格、爱好等）可以胜任学生会主席的工作
2. 如果当选，你将为本校同学做些什么

A Campaign Speech

.....

.....

.....

.....

.....

[illegible]

1.[Y] [N] [NG]                      2.[Y] [N] [NG]                      3.[Y] [N] [NG]                      4.[Y] [N] [NG]

5.[Y] [N] [NG]                      6.[Y] [N] [NG]                      7.[Y] [N] [NG]

8. Typical customers of a landfill are \_\_\_\_\_.

9. To dispose of a ton of trash in a landfill, customers have to pay tipping fee of \_\_\_\_\_.

10. Materials that are not permitted to be buried in landfills should be dumped at \_\_\_\_\_.

## Appendix C: A Sample of the Revised CET

### 答题卡 2 (Answer Sheet 2)

学校:		准 考 证 号															
姓名:		[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	
划线要求		[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	[1]	
		[2]	[2]	[2]	[2]	[2]	[2]	[2]	[2]	[2]	[2]	[2]	[2]	[2]	[2]	[2]	
		[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	[3]	
		[4]	[4]	[4]	[4]	[4]	[4]	[4]	[4]	[4]	[4]	[4]	[4]	[4]	[4]	[4]	
		[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]	
		[6]	[6]	[6]	[6]	[6]	[6]	[6]	[6]	[6]	[6]	[6]	[6]	[6]	[6]	[6]	
		[7]	[7]	[7]	[7]	[7]	[7]	[7]	[7]	[7]	[7]	[7]	[7]	[7]	[7]	[7]	
		[8]	[8]	[8]	[8]	[8]	[8]	[8]	[8]	[8]	[8]	[8]	[8]	[8]	[8]	[8]	
		[9]	[9]	[9]	[9]	[9]	[9]	[9]	[9]	[9]	[9]	[9]	[9]	[9]	[9]	[9]	

### Part III Section A

### Section B

- 11.[A] [B] [C] [D] 16.[A] [B] [C] [D] 21.[A] [B] [C] [D] 26.[A] [B] [C] [D] 31.[A] [B] [C] [D]  
 12.[A] [B] [C] [D] 17.[A] [B] [C] [D] 22.[A] [B] [C] [D] 27.[A] [B] [C] [D] 32.[A] [B] [C] [D]  
 13.[A] [B] [C] [D] 18.[A] [B] [C] [D] 23.[A] [B] [C] [D] 28.[A] [B] [C] [D] 33.[A] [B] [C] [D]  
 14.[A] [B] [C] [D] 19.[A] [B] [C] [D] 24.[A] [B] [C] [D] 29.[A] [B] [C] [D] 34.[A] [B] [C] [D]

### Part III Section C

Russia is the largest economic power that is not a member of the World Trade Organization. But that may change. Last Friday, the European Union said it would support Russia's (36) to become a W.T.O member.

Representative of the European Union met with Russian(37)\_\_\_\_\_in Moscow. They signed a trade agreement that took six years to(38)\_\_\_\_\_.

Russia called the trade agreement(39)\_\_\_\_\_.It agreed to slowly increase fuel prices within the country. It also agreed to permit(40)\_\_\_\_\_in its communications industry and to remove some barriers to trade.

In(41)\_\_\_\_\_for European support to join the W.T.O , Russian president Putin said that Russia would speed up the(42)\_\_\_\_\_to approve the Kyoto Protocol, and international(43) agreement to reduce the production of harmful industrial gases.(44)\_\_\_\_\_

Russia had signed the Kyoto Protocol, but has not yet approved it. The agreement takes effect when it has been approved by nations that produce at least 55 percent of the world's greenhouse gases.(45)\_\_\_\_\_

The United States, the world's biggest producer, withdrew from the Kyoto Protocol after President Bush took office in 2001. So, Russia's approval is required to put the Kyoto Protocol into effect.

(46)\_\_\_\_\_

Russia must still reach agreement with China, Japan, South Korean and the United States.

**Appendix C: A Sample of the Revised CET****大学英语四级考试试点考试样卷 (试题册)****Part I Writing (30 minutes)**

注意：此部分试题在答题卡 1 上。

**Part II Reading Comprehension (Skimming and Scanning)  
(15 minutes)**

**Directions:** In this part, you will have 15 minutes to go over the passage quickly and answer the questions on **Answer Sheet 1**.

For questions 1-7, mark

Y (for YES) if the statement agrees with the information given in the passage;

N (for NO) if the statement contradicts the information given in the passage;

NG (for NOT GIVEN) if the information is not given in the passage. For questions 8-10, complete the sentences with the information given in the passage.

**Landfills**

You have just finished your meal at a fast food restaurant and you throw your uneaten food, food wrappers, drink cups, utensils and napkins into the trash can. You don't think about that waste again. On trash pickup day in your neighborhood, you push your can out to the curb, and workers dump the contents into a big truck and haul it away. You don't have to think about that waste again, either. But maybe you have wondered, as you watch the trash truck pull away, just where that garbage ends up.

Americans generate trash at an Astonishing rate of four pounds per day per person; which translates to 600,000 tons per day or 210 million tons per year! This is almost twice as much trash per person as most other major countries. What happens to this trash? Some gets recycled (回收利用) or recovered and some is burned, but the majority is buried in landfills.



## Appendix C: A Sample of the Revised CET

### How Much Trash Is Generated?

Of the 210 million tons of trash, or solid waste, generated in the United States annually, about 56 million tons, or 27 percent, is either recycled (glass, paper products, plastic, metals) or composted (做成堆肥) (yard waste). The remaining trash, which is mostly unrecyclable, is discarded.

### How Is Trash Disposed of ?

The trash production in the United States has almost tripled since 1960. This trash is handled in various ways. About 27 percent of the trash is recycled or composted, 16 percent is burned and 57 percent is buried in landfills. The amount of trash buried in landfills has doubled since 1960. The United States ranks somewhere in the middle of the major countries (United Kingdom, Canada, Germany, France and Japan) in landfill disposal. The United Kingdom ranks highest, burying about 90 percent of its solid waste in landfills.

### What Is a Landfill?

There are two ways to bury trash:

- Dump—an open hole in the ground where trash is buried and that is full of various animals (rats, mice, birds). (This is most people's idea of a landfill!)
- Landfill—carefully designed structure built into or on top of the ground in which trash is isolated from the surrounding environment (groundwater, air, rain). This isolation is accomplished with a bottom liner and daily covering of soil.
  - Sanitary landfill—land fill that uses a clay liner to isolate the trash from the environment
  - Municipal solid waste (MSW) landfill—landfill that uses a synthetic (plastic) liner to isolate the trash from the environment

The purpose of a landfill is to bury the trash in such a way that it will be isolated from groundwater, will be kept dry and will not be in contact with air. Under these conditions, trash will not decompose (腐烂) much. A landfill is not like a compost pile, where the purpose is to bury trash in such a way that it will decompose quickly.

### Proposing the Landfill

For a landfill to be built, the operators have to make sure that they follow certain steps. In most parts of the world, there are regulations that govern where a landfill can be placed and how it can operate. The whole process begins with someone proposing the landfill. In the United States, taking care of trash and building landfills are local government

## Appendix C: A Sample of the Revised CET

responsibilities. Before a city or other authority can build a landfill, an environment impact study must be done on the proposed site to determine:

- the area of land necessary for the landfill
- the composition of the underlying soil and bedrock
- the flow of surface water over the site
- the impact of the proposed landfill on the local environment and wildlife
- the historical value of the proposed site

## Building the Landfill

Once the environmental impact study is complete, the permits are granted and the funds have been raised, then construction begins. First, access roads to the landfill site must be built if they do not already exist. These roads will be used by construction equipment, sanitation (环卫) services and the general public. After roads have been built, digging can begin. In the North Wake County Landfill, the landfill began 10 feet below the road surface.

## What Happens to Trash in a Landfill?

Trash put in a landfill will stay there for a very long time. Inside a landfill, there is little oxygen and little moisture. Under these conditions, trash does not break down very rapidly. In fact, when old landfills have been dug up or sampled, 40-year-old newspapers have been found with easily readable print. Landfills are not designed to break down trash, merely to bury it. When a landfill closes, the site, especially the groundwater, must be monitored and maintained for up to 30 years!

## How Is a Landfill Operated?

A landfill, such as the North Wake County Landfill, must be open and available every day. Customers are typically municipalities and construction companies, although residents may also use the landfill.

Near the entrance of the landfill is a recycling center where residents can drop off recyclable materials (aluminum cans, glass bottles, newspapers and paper products). This helps to reduce the amount of material in the landfill. Some of these materials are banned from landfills by law because they can be recycled.

As customers enter the site, their trucks are weighed at the scale house. Customers are charged tipping fees for using the site. The tipping fees vary from \$10 to \$40 per ton. These fees are used to pay for operation costs. The North Wake County Landfill has an operating budget of approximately \$4.5 million, and part of that comes from tipping fees.

## Appendix C: A Sample of the Revised CET

Along the site, there are drop-off stations for materials that are not wanted or legally banned by the landfill. A multi-material drop-off station is used for tires, motor oil, lead-acid batteries. Some of these materials can be recycled.

In addition, there is a household hazardous waste drop-off station for chemicals (paints, pesticides, other chemicals) that are banned from the landfill. These chemicals are disposed of by private companies. Some paints can be recycled and some organic chemicals can be burned in furnaces or power plants.

Other structures alongside the landfill are the borrowed area that supplies the soil for the landfill, the runoff collection pond and methane (甲烷) station.

Landfills are complicated structures that, when properly designed and managed, serve an important purpose. In the future, new technologies called bioreactors will be used to speed the breakdown of trash in landfills and produce more methane.

注意：此部分试题请在答题卡 1 上作答；8-10 题在答题卡 1 上。

1. The passage gives a general description of the structure and use of a landfill.
2. Most of the trash that Americans generate ends up in landfills.
3. Compared with other major Industrialized countries, America buries a much higher percentage of its solid waste in landfills.
4. Landfills are like compost piles in that they speed up decomposition of the buried trash.
5. In most countries the selection of a landfill site is governed by rules and regulations.
6. In the United States the building of landfills is the job of both federal and local governments.
7. Hazardous wastes have to be treated before being dumped into landfills.

## Part III Listening Comprehension (35 minutes)

### Section A

**Directions:** In this section, you will hear 8 short conversations and 2 long conversations. At the end of each conversation, one or more questions will be asked about what was said. Both the conversation and the questions will be spoken only once. After each question there will be a pause. During the pause, you must read the four choices marked A), B), C) and D), and decide which is the best answer. Then mark the corresponding letter on **Answer Sheet 2** with a single line through the center.

---

**Appendix C: A Sample of the Revised CET**

注意：此部分试题请在答题卡 2 上作答。

11. A) The man hates to lend his tools to other people.  
B) The man hasn't finished working on the bookshelf.  
C) The tools have already been returned to the woman.  
D) The tools the man borrowed from the woman are missing.
  
12. A) Save time by using a computer.  
B) Buy her own computer.  
C) Borrow Martha's computer.  
D) Stay home and complete her paper.
  
13. A) He has been to Seattle many times.  
B) He has chaired a lot of conferences.  
C) He holds a high position in his company.  
D) He lived in Seattle for many years.
  
14. A) Teacher and student.  
B) Doctor and patient.  
C) Manager and office worker.  
D) Travel agent and customer.
  
15. A) She knows the guy who will give the lecture.  
B) She thinks the lecture might be informative.  
C) She wants to add something to her lecture.  
D) She'll finish her report this weekend.
  
16. A) An art museum.  
B) A beautiful park.  
C) A college campus.  
D) An architectural exhibition.
  
17. A) The houses for sale are of poor quality.  
B) The houses are too expensive for the couple to buy.  
C) The housing developers provide free trips for potential buyers.  
D) The man is unwilling to take a look at the houses for sale.
  
18. A) Talking about sports.  
C) Reading newspapers.  
B) Writing up local news.  
D) Putting up advertisements.

## Appendix C: A Sample of the Revised CET

### Part IV Reading Comprehension (Reading in Depth) (25 minutes)

#### Section A

**Directions:** *In this section, there is a passage with ten blanks. You are required to select one word for each blank from a list of choices given in a word bank following the passage. Read the passage through carefully before making your choices. Each choice in bank is identified by a letter. Please mark the corresponding letter for each item on Answer Sheet 2 with a single line through the center. You may not use any of the words in the bank more than once.*

**Questions 47 to 56 are based on the following passage.**

When Roberto Feliz came to the USA from the Dominican Republic, he knew only a few words of English. Education soon became a 47. “I couldn’t understand anything,” he said. He 48 from his teachers, came home in tears, and thought about dropping out.

Then Mrs. Malave, a bilingual educator, began to work with him while teaching him math and science in his 49 Spanish. “She helped me stay smart while teaching me English,” he said. Given the chance to demonstrate his ability, he 50 confidence and began to succeed in school.

Today, he is a 51 doctor, runs his own clinic, and works with several hospitals. Every day, he uses the language and academic skills he 52 through bilingual education to treat his patients.

Roberto’s story is just one of 53 success stories. Research has shown that bilingual education is the most 54 way both to teach children English and ensure that they succeed academically. In Arizona and Texas, bilingual students 55 outperform their peers in monolingual programs. Calexico, Calif., implemented bilingual education, and now has dropout rates that are less than half the state average and college 56 rates of more than 90%. In El Paso, bilingual education programs have helped raise student scores from the lowest in Texas to among the highest in the nation.

注意：此部分试题请在答题卡 2 上作答。

## Appendix C: A Sample of the Revised CET

A) wonder	I)hid
B) acquired	J)prominent
C) consistently	K)decent
D) regained	L)countless
E) nightmare	M)recalled
F) native	N)breakthrough
G) acceptance	O)automatically
H) effective	

### Section B

**Directions:** *There are 2 passages in this section .Each passage is followed by some questions or unfinished statement. For each of them there are four choices marked A),B),C),D). You should decide on the best choice and mark the corresponding letter on **Answer Sheet 2** with a single line through the center.*

### Passage One

**Question 57 to 61 are based on the following passage.**

“Tear ’em apart!” “Kill the fool!” “Murder the referee(裁判) !”

These are common remarks one may hear at various sporting events. At the time they are made ,they may seem innocent enough. But let’s not kid ourselves .They have been known to influence behavior in such a way as to lead to real bloodshed. Volumes have been written about the way word affect us. It has been shown that words having certain connotations (含义) may cause us to react in ways quite foreign to what we consider to be our usual humanistic behavior. I see the term “opponent” as one of those words .Perhaps the time has come to delete it from sports terms.

The dictionary meaning of the term “opponent” is “adversary”; “enemy” “one who opposes your interests.” Thus, when a player meets an opponent ,he or she may tend to every action no matter how gross ,may be considered justifiable. I recall an incident in a handball game when a referee refused a player’s request for a time out for a glove change because he did not consider them wet enough .The player proceeded to rub his gloves across his wet T-shirt and then exclaimed, “Are they wet enough now?”

In the heat of battle, players have been observed to throw themselves across the court without considering the consequences the such a move might have on anyone in their way. I have also witnessed a player reacting to his opponent’s intentional and illegal blocking by deliberately

## Appendix C: A Sample of the Revised CET

hitting him with the ball as hard as he could during the course of play. Off the court, they are good friends. Does that make any sense? It certainly gives proof of a court attitude which departs from normal behavior.

Therefore, I believe it is time we elevated (提升) the game to the level where it belongs, thereby setting an example to the rest of the sporting world. Replacing the term “opponent” with “associate” could be an ideal way to start.

The dictionary meaning of the term “associate” is “colleague”; “friend”; “companion.” Reflect a moment! You may soon see and possibly feel the difference in your reaction to the term “associate” rather than “opponent”.

注意：此部分试题请在答题卡 2 上作答。

57. Which of the following statements best expresses the author’s view?

- A) The words people use can influence their behavior.
- B) Unpleasant words in sports are often used by foreign athletes.
- C) Aggressive behavior in sports can have serious consequences.
- D) Unfair judgments by referees will lead to violence on the sports field.

58. Harsh words are spoken during games because the players \_\_\_\_\_.

- A) are too eager to win
- B) treat their rivals as enemies
- C) are usually short-tempered and easily offended
- D) cannot afford to be polite in fierce competitions

59. What did the handball player do when he was not allowed a time out to change his gloves?

- A) He angrily hit the referee with a ball.
- B) He refused to continue the game.
- C) He claimed that referee was unfair.
- D) He wet his gloves by rubbing them across his T-shirt.

60. According to the passage, players in a game may \_\_\_\_\_.

- A) kick the ball across the court with force
- B) lie down on the ground as an act of protest
- C) deliberately throw the ball at anyone illegally blocking their way
- D) keep on screaming and shouting throughout the game

## Appendix C: A Sample of the Revised CET

### Part V Cloze (15 minutes)

**Directions:** There are 20 blanks in the following passage. For each blank there are four choices marked A), B), C) and D) on the right side of the paper. You should choose the ONE that best fits into the passage. Then mark the corresponding letter on **Answer Sheet 2** with a single line through the center.

注意：此部分试题请在答题卡 2 上作答。

Wise buying is a positive way in which you can	67. A) form	C) way
make your money go further. The <u>67</u> you go about	B) fashion	D) method
purchasing an article or a service can actually <u>68</u>	68. A) save	C) raise
you money or can add <u>69</u> the cost.	B) preserve	D) retain
Take the <u>70</u> example of a hairdryer. If you are	69. A) up	C) in
buying a hairdryer, you might <u>71</u> that you are	B) to	D) on
making the <u>72</u> buy if you choose one <u>73</u> look you	70. A) easy	C) simple
like and which is also the cheapest <u>74</u> price. But	B) single	D) similar
when you get it home you may find that it <u>75</u> twice	71. A) convince	C) examine
as long as a more expensive <u>76</u> to dry your	B) accept	D) think
hair. The cost of the electricity plus the cost of your	72. A) proper	C) reasonable
time could well <u>77</u> your hairdryer the most	B) best	D) most
expensive one of all.	73. A) its	C) whose
So what principles should you <u>78</u> when you go out	B) which	D) what
shopping?	74. A) for	C) in
If you <u>79</u> your home, your car or any valuable <u>80</u> in	B) with	D) on
excellent condition, you'll be saving money in the	75. A) spends	C) lasts
long <u>81</u> .	B) takes	D) consumes
Before you buy a new <u>82</u> , talk to someone who	76. A) mode	C) sample
owns one. If you can, use it or borrow it to check it	B) copy	D) model
suits your particular <u>83</u> .	77. A) cause	C) leave
Before you buy an expensive <u>84</u> , or a service, do	B) make	D) bring
check the price and <u>85</u> is on offer. If possible,	78. A) adopt	C) stick
choose <u>86</u> three items or three estimates.	B) lay	D) adapt
	79. A) reserve	C) store
	B) decorate	D) keep
	80. A) products	C) material
	B) possession	D) ownership
	81. A) run	C) period
	B) interval	D) time
	82. A) appliance	C) utility
	B) machinery	D) facility
	83. A) function	C) target
	B) purpose	D) task
	84. A) component	C) item
	B) element	D) particle
	85. A) what	C) that
	B) which	D) this
	86. A) of	C) by
	B) in	D) from



**Appendix C: A Sample of the Revised CET****大学英语四级考试试点考试样卷(听力文字稿)****Tape Script of Listening Comprehension****Section A**

**Directions:** *In this section, you will hear 8 short conversations and 2 long conversations. At the end of each conversation, one or more questions will be asked about what was said. Both the conversation and the questions will be spoken only once. After each question there will be a pause. During the pause, you must read the four choices marked A), B), C) and D), and decide which is the best answer. Then mark the corresponding letter on Answer Sheet 2 with a single line through the centre.*

11. W: Simon, could you return the tools I lent you for building the bookshelf last month?  
M: Uh, well, I hate to tell you this... but I can't seem to find them.  
Q: What do we learn from the conversation?
12. W: I'm going to Martha's house. I have a paper to complete, and I need to use her computer.  
M: Why don't you buy one yourself? Think how much time you could save.  
Q: What does the man suggest the woman do?
13. W: Bob said that Seattle is a great place for conferences.  
M: He's certainly in a position to make that comment. He's been there so often.  
Q: What does the man say about Bob?
14. W: Mr. Watson, I wonder whether it's possible for me to take a vacation early next month.  
M: Did you fill out a request form?  
Q: What is the probable relationship between the two speakers?
15. M: Do you want to go to the lecture this weekend? I hear the guy who's going to deliver the lecture spent a year living in the rain forest.  
W: Great! I'm doing a report on the rain forest. Maybe I can get some new information to add to it.  
Q: What does the woman mean?
16. W: Wow! I do like this campus: all the big trees, the green lawns, and the old buildings with tall columns. It's really beautiful.

## Appendix C: A Sample of the Revised CET

M: It sure is. The architecture of these buildings is in the Greek style. It was popular in the eighteenth century here.

Q: What are the speakers talking about?

17. M: This article is nothing but advertising for housing developers. I don't think the houses for sale are half that good.

W: Come on, David. Why so negative? We're thinking of buying a home, aren't we? Just a trip to look at the place won't cost us much.

Q: What can be inferred from the conversation?

18. M: Would you pass me the sports section, please?

W: Sure, if you give me the classified ads local news section.

Q: What are the speakers doing?

## Now you'll hear two long conversations.

### Conversation One

W: Hello, Gary. How're you?

M: Fine! And yourself?

W: Can't complain. Did you have time to look at my proposal?

M: No, not really. Can we go over it now?

W: Sure. I've been trying to come up with some new production and advertising strategies. First of all, if we want to stay competitive, we need to modernize our factory. New equipment should've been installed long ago.

M: How much will that cost?

W: We have several options ranging from one hundred thousand dollars all the way up to half a million.

M: OK. We'll have to discuss these costs with finance.

W: We should also consider human resources. I've been talking to personnel as well as our staff at the factory.

M: And what's the picture?

W: We'll probably have to hire a couple of engineers to help us modernize the factory.

M: What about advertising?

W: Marketing has some interesting ideas for television commercials.

M: TV? Isn't that a bit too expensive for us? What's wrong with advertising in the papers, as usual?

W: Quite frankly, it's just not enough anymore. We need to be more aggressive in order to keep ahead of our competitors.

M: Will we be able to afford all this?

W: I'll look into it, but I think higher costs will be justified. These investments will result in higher profits for our company.

M: We'll have to look at the figures more closely. Have finance draw up a budget for these investments.

W: All right. I'll see to it.

## Appendix C: A Sample of the Revised CET

### Questions 19 to 20 are based on the conversation you have just heard.

19. What are the two speakers talking about?
20. What does the woman say about the equipment of their factory?
21. What does the woman suggest about human resources?
22. Why does the woman suggest advertising on TV?

#### Conversation Two

W: Sir, you've been using the online catalogue for quite a while, Is there anything I can do to help you?

M: Well, I've got to write a paper about Hollywood in the 30s and 40s, and I'm really struggling. There are hundreds of books, and I just don't know where to begin.

W: Your topic sounds pretty big. Why don't you narrow it down to something like...uh... the history of the studios during that time?

M: You know, I was thinking about doing that, but more than 30 books came up when I typed in "movie studios."

W: You could cut that down even further by listing the specific years you want. Try adding "1930s" or "1940s" or maybe "Golden Age."

M: "Golden Age" is a good idea. Let me type that in ... Hey, look, just 6 books this time. That's a lot better.

W: Oh... another thing you might consider... have you tried looking for any magazine or newspaper articles?

M: No, I've only been searching for books.

W: Well, you can look up magazine articles in the Reader's Guide to Periodical Literature. And we do have the Los Angeles Times available over there. You might go through their indexes to see if there's anything you want.

M: Okay, I think I'll get started with these books and then I'll go over the magazines.

W: If you need any help, I'll be over at the Reference Desk.

M: Great, thanks a lot.

### Questions 23 to 25 are based on the conversation you have just heard.

23. What is the man doing?
24. What does the librarian think of the topic the man is working on?
25. Where can the man find the relevant magazine articles?

## Section B

**Directions:** In this section, you will hear 3 short passages. At the end of EACH PASSAGE, you will hear some questions. Both the passage and the questions will be spoken only once. After you hear a question, you must choose the best answer from the four choices marked A), B\*, C) and D). Then mark the corresponding letter on **Answer Sheet 2** with a single line through the centre.

### Passage One

In the next few decades people are going to travel very differently from the way they do today. Everyone is going to drive electrically powered cars. So in a few years people won't worry about running out of gas.

Some of the large automobile companies are really moving ahead with this new technology. F & C Motors, a major auto company, for example, is holding a press conference next week. At the press conference the company will present its new, electronically operated models.

Transportation in the future won't be limited to the ground. Many people predict that traffic will quickly move to the sky. In the coming years, instead of radio reports about road conditions and highway traffic, news reports will talk about traffic jams in the sky.

But the sky isn't the limit. In the future, you'll probably even be able to take a trip to the moon. Instead of listening to regular airplane announcements, you'll hear someone say, "The spacecraft to the moon leaves in ten minutes. Please check your equipment. And remember, no more than ten ounces of carry-on baggage are allowed."

---

## Appendix C: A Sample of the Revised CET

### Questions 26 to 28 are based on the passage you have just heard.

26. What will be used to power cars in the next few decades?
27. What will future news reports focus on when talking about transportation?
28. What is the special requirement for passengers traveling to the moon?

### Passage Two

The period of engagement is the time between the marriage proposal and the wedding ceremony. Two people agree to marry when they decide to spend their lives together.

The man usually gives the woman a diamond engagement ring? That tradition is said to have started when an Austrian man gave a diamond ring to the woman he wanted to marry. The diamond represented beauty. He placed it on the third finger of her left hand. He chose that finger because it was thought that a blood vessel in that finger went directly to the heart. Today, we know that this is not true. Yet the tradition continues.

Americans generally are engaged for a period of about one year if they are planning a wedding ceremony and party. During the time, friends of the bride may hold a party at which women friends and family members give the bride gifts that she will need as a wife. These could include cooking equipment or new clothing.

Friends of the man who is getting married may have a bachelor party for him. This usually takes place the night before the wedding. Only men are invited to the bachelor party.

During the marriage ceremony, the bride and her would-be husband usually exchange gold rings that represent the idea that their union will continue forever. The wife often wears both the wedding ring and engagement ring on the same finger. The husband wears his ring on the third finger of his left hand.

Many people say the purpose of the engagement period is to permit enough time to plan the wedding. But the main purpose is to let enough time pass so the two people are sure they want to marry each other. Either person may decide to break the engagement. If this happens, the woman usually returns the ring to the man; they also return any wedding gifts they have received.

### Questions 29 to 31 are based on the passage you have just heard.

29. What was the diamond ring said to represent?
30. Why did the Austrian man place the diamond ring on the third finger of the left hand of his would-be wife?
31. What is the chief advantage of having the engagement period?

## Appendix C: A Sample of the Revised CET

### 大学英语四级考试试点考试样卷（标准答案）

#### Key

#### Part II Reading Comprehension (Skimming and Scanning)

1. Y    2. Y    3. N    4. N    5. Y    6. N    7. NG

8. municipalities and construction companies

9. \$10 to \$40

10. drop-off stations

#### Part III Listening Comprehension

##### Section A

11. D    12. B    13. A    14. C    15. B

16. C    17. D    18. C    19. D    20. B

21. B    22. C    23. A    24. A    25. D

##### Section B

26. D    27. A    28. B    29. A    30. C

31. B    32. D    33. B    34. B    35. B

##### Section C

36. effort    37. officials    38. negotiate    39. balanced

40. competition    41. exchange    42. process    43. environmental

44. These “greenhouse gases” trap heat in the atmosphere and are blamed for changing the world’s climate.

45. But currently, nations producing only 44 percent have approved the Protocol. Russia produces about 17 percent of the world’s greenhouse gases.

46. To join the W.T.O., a country must reach trade agreements with major trading countries that are also W.T.O. members.

#### Part I V Reading Comprehension (Reading in Depth)

##### Section A

47. E    48. I    49. F    50. D    51. J

52. B    53. L    54. H    55. C    56. G

##### Section B

57. A    58. B    59. D    60. C    61. D

62. C    63. D    64. B    65. A    66. B

##### Part V Cloze

67. C    68. A    69. B    70. C    71. D

72. B    73. C    74. C    75. B    76. D

77. B    78. A    79. D    80. B    81. A

82. A    83. B    84. C    85. A    86. D

---

**APPENDIX D – Teacher Interview Protocol****Date of Interview** \_\_\_\_\_**Place of Interview** \_\_\_\_\_

Time the interview started \_\_\_\_\_ Time the interview ended \_\_\_\_\_

**Introductory statement:** Thank you for agreeing to participate in this interview. The project I'm working on is called "Examining the Role of the 'Teacher Factor' in Washback". Today I want to ask you some questions concerning your perceptions of the CET and its impact on teaching and learning. Before we start, I want to assure you that what you are going to talk about here will be used for the purpose of this research project only. I will not let any other people listen to the recording or see the transcripts.

**Questions:**

1. What do you think are the main differences between the revised CET and the old one? What do you think of the revised CET format? Do you think the objectives of the CET overlap those of the CECR?
2. How do you think of the changes introduced to the CET? Do you think it's necessary to increase the portion of listening in the test? Why (or why not)?
3. Do you think the revised CET has positive (or negative) impact on your teaching, or both? In what way?
4. What do you think are the best ways to develop students' English proficiency? What is your favorite teaching method? Why do you think it is superior to other methods? How do you try to foster your students' communicative competence, particularly their competence in speaking and listening? Do they conduct their teaching based on the guidelines of the CECR? How do you think of this version of national curriculum (the CECR) as compared to the 1999 version?
5. Has the implementation of the new CET caused you to use any different teaching methods?  
(If no) Why do you think your teaching methods do not change?
6. Have you observed any changes in your classroom, which you feel, have been brought about by the implementation of the revised CET? Could you please describe and explain them further?

**Closing remarks:** Thank you very much for sharing your views with me. We may talk again if you have something to add or clarify. I might have more questions to ask you in the next interview. I will notify you in advance when the decision is made on when our next talk will be held. Thanks again for your help and support.

Total time of Observation:

Class:

Unit of Textbook:

[illegible]

## Appendix F – The Questionnaire

I would like to ask you to help us by answering the following questions concerning your perceptions of Band Four or Band Six of the College English Test (CET4 or CET6) and its impact on your teaching. All information will be absolutely confidential. Please give your answers sincerely as only this will guarantee the success of the investigation. Thank you very much for your help.

### Part 1 Your Background Information

Please check [ ☐ ] the appropriate answer.

- (1) Your gender: [ ☐ ] female [ ☐ ] male
- (2) Your age: [ ☐ ] 21-29 [ ☐ ] 30-39 [ ☐ ] 40-49 [ ☐ ] 50-59 [ ☐ ] over 60
- (3) Number of years you have been teaching:  
[ ☐ ] 1-4 [ ☐ ] 5-9 [ ☐ ] 10-14 [ ☐ ] 15-19 [ ☐ ] over 20
- (4) Number of class hours you teach EFL per week:  
[ ☐ ] 6-8 hours [ ☐ ] 9-11 hours [ ☐ ] 12-14 hours [ ☐ ] 15 hours or more
- (5) Types of students you teach: [ ☐ ] English majors [ ☐ ] non-English majors
- (5a) Levels you teach: [ ☐ ] freshmen [ ☐ ] sophomores  
[ ☐ ] juniors [ ☐ ] seniors [ ☐ ] Other, Specify: \_\_\_\_\_
- (5b) Class types: [ ☐ ] regular undergraduates [ ☐ ] two-year programs  
[ ☐ ] continuing education [ ☐ ] Other, Specify: \_\_\_\_\_
- (5c) Number of students there are in your class:  
[ ☐ ] 20-29 [ ☐ ] 30-39 [ ☐ ] 40-49 [ ☐ ] 50-59 [ ☐ ] over 60  
[ ☐ ] Other, Specify: \_\_\_\_\_
- (6) Your academic background: [ ☐ ] Bachelors  
[ ☐ ] Advanced /Assistant Teacher Training Certificate (高级教师/助教进修班) [ ☐ ] M.A. [ ☐ ] M.Ed. [ ☐ ] PhD [ ☐ ] other, Specify: \_\_\_\_\_
- (6a) Have you ever been in an English-speaking country for one year or more?  
[ ☐ ] Yes [ ☐ ] No
- (7) School you teach in: \_\_\_\_\_
- (8) Do you have specific training in ESL? [ ☐ ] Yes [ ☐ ] No  
If Yes, what type of training is it? [ ☐ ] university training [ ☐ ] a workshop  
[ ☐ ] a certificate short program [ ☐ ] other, specify: \_\_\_\_\_  
where have you participated in it \_\_\_\_\_;  
for how long: \_\_\_\_\_



## Appendix F – The Questionnaire

### Part 2 Your Teaching in the Classroom

*In the brackets [ ], please mark the following on a six point scale as:*

**[1] 0-10% [2] 11-20% [3] 21-30% [4] 31-40% [5] 41-50% [6] 51% or more**

(9)

a. [ ] Approximately what percentage of your class time is devoted to speaking activities:

**[1] 0-10% [2] 11-20% [3] 21-30% [4] 31-40% [5] 41-50% [6] 51% or more**

b. [ ] Approximately what percentage of your class time is spent on just listening activities?

c. [ ] Approximately what percentage of your class time is spent on just reading activities?

d. [ ] Approximately what percentage of your class time is spent on just writing activities?

e. [ ] Approximately what percentage of your class time is spent on just testing activities?

(10) Approximately what percentage of your class time is conducted with the following types of participant organization?

a. [ ] teacher centered - full class

b. [ ] in large groups (5 to 15 students)

c. [ ] in small groups (3 to 5 students)

d. [ ] pair work

e. [ ] individual work

f. [ ] task-based activities

(11) [ ] Approximately what is the percentage of teacher-talk time in your classroom?

(12) [ ] Approximately what is the percentage of student-talk time in your classroom?

(13) [ ] Approximately what percentage of your classroom instruction is conducted in English?

(14) [ ] Approximately what percentage of your classroom instruction is devoted to teaching grammar?

(15) [ ] Approximately what percentage of your classroom instruction is devoted to teaching vocabulary?

### Part 3 Your ideas on Teaching, learning and Testing

*In the brackets [ ], please mark the following on a four point scale as:*

**[1] strongly disagree [2] disagree [3] agree [4] strongly agree**

(16) [ ] My textbook is well-suited to developing students' communicative competence.

(17) [ ] My textbook is well-suited to providing practice in preparation for the College English Test (CET).

## Appendix F – The Questionnaire

Please specify which textbook you use:

- (18) ☐ If the students have a good command of what is covered in the textbook, then they have little problem achieving high scores on the CET.
- (19) ☐ I believe that the test scores on the CET are an appropriate indicator of a student's English ability.
- (20) ☐ I view the CET as benefiting students (e.g., the test can motivate students, help students understand their own learning needs, etc.).
- (21) ☐ The CET has a positive influence on teaching (e.g., they help focus teaching, provide feedback on teaching).
- (22) ☐ I believe that the CET reflects the goals and objectives of the College English Curriculum Requirement.
- (23) ☐ I believe that the CET does not hinder the application of Communicative Language Teaching (CLT) (which makes communicative use of the language the focus of lessons).
- (24) ☐ I view the structural approach (which emphasizes the importance of grammatical rules and conscious understanding of the language system) as the best method to prepare students for the CET.
- (25) ☐ I feel unsure about how to go about things when teaching EFL.
- (26) ☐ My role as an English teacher in the language classroom is to transmit knowledge to my students through explaining texts and giving examples.
- (27) ☐ Learning a language is to accumulate the knowledge of grammar, vocabulary, structures and rules.

**Check (✓) what you think is appropriate:**

(28) The main goal of my teaching is to help students \_\_\_\_\_.

- ☐ succeed on the CET
- ☐ acquire language proficiency
- ☐ accumulate knowledge of grammar and vocabulary
- ☐ enhance their reading skills
- ☐ other, specify: \_\_\_\_\_

(29) What do you think is the ideal percentage of class time for teacher talk?

- ☐ 10%
- ☐ 20%
- ☐ 30%
- ☐ 40%
- ☐ 50% or more

## Appendix F – The Questionnaire

(30) What skill or area of ability do you think that the revised CET is intended to measure?

- ☐ linguistic knowledge
- ☐ language use through listening, reading and writing
- ☐ vocabulary
- ☐ reading and writing
- ☐ other, specify: \_\_\_\_\_

(31) I understand that the main objective of the revised CET is \_\_\_\_\_.

- ☐ to evaluate students' language competence.
- ☐ to evaluate students' knowledge of English grammar and vocabulary
- ☐ to choose prospective students
- ☐ to provide feedback on teaching
- ☐ other, specify: \_\_\_\_\_

(32) What methodology do you believe that you are using in your instruction?

- ☐ Communicative Language Teaching (CLT)
- ☐ the traditional structural approach
- ☐ a combined approach of CLT and the structural approach
- ☐ other, specify: \_\_\_\_\_

**Check (✓) what you think is appropriate. You may choose more than one response.**

(33) What do you think are the barriers (i.e., constraints, restrictions) to your teaching?

- ☐ the constraint of the CET
- ☐ my limited knowledge of advanced teaching methodologies
- ☐ my own English proficiency level
- ☐ textbooks
- ☐ class size
- ☐ heavy workload
- ☐ the attitude of the students
- ☐ other, specify: \_\_\_\_\_

---

## Appendix F – The Questionnaire

(34) What types of activities do you think are effective in getting students involved in the learning process?

- ☐ memorization of new words
- ☐ translation
- ☐ information gap
- ☐ role-play
- ☐ language games
- ☐ analysis of grammar
- ☐ group work or pair work
- ☐ lectures
- ☐ reading aloud
- ☐ other, specify: \_\_\_\_\_

(35) The positive effects of the CET lie in that \_\_\_\_\_.

- ☐ it motivates teachers to improve their methodology in teaching English.
- ☐ it motivates students to enhance their proficiency in English
- ☐ it forces students to devote more time to the study of English
- ☐ it is in line with my teaching objectives
- ☐ it helps upgrade teaching and learning
- ☐ other, specify: \_\_\_\_\_

(36) The negative effects of the CET lie in that \_\_\_\_\_.

- ☐ it discourages the use of advanced teaching methodologies
- ☐ it encourages memorization of vocabulary and language rules
- ☐ it forces me to use the structural approach
- ☐ it forces me to teach to the test
- ☐ other, specify: \_\_\_\_\_

(37) What teaching approach do you think you prefer?

- ☐ the grammar-translation method (语法翻译法)
- ☐ Communicative Language Teaching (CLT)
- ☐ a combined approach (a combination of CLT and the structural approach)
- ☐ other, specify: \_\_\_\_\_

## Appendix F – The Questionnaire

(38) Select the reason(s) why you prefer the teaching approach you have chosen in (37).

*(More than one choice is possible.)*

- ☐ It is an effective foreign language teaching method.
- ☐ It is the way I learned English when I was a student.
- ☐ It is easy and least stressful.
- ☐ It helps foster student comprehensive skills in English.
- ☐ It is the best method for helping students pass the College English Test (CET).
- ☐ other, specify: \_\_\_\_\_

**Check (✓) what you think is appropriate:**

(39) What do you use to guide and organize your lessons?

- ☐ the CET
- ☐ the College English Curriculum Requirement
- ☐ teachers' books
- ☐ textbooks
- ☐ other, specify: \_\_\_\_\_

(40) Communicative competence means competence in \_\_\_\_\_.

- ☐ receptive skills such as listening and reading
- ☐ productive skills such as speaking and writing
- ☐ both receptive skills and productive skills
- ☐ skills such as speaking and listening
- ☐ just oral proficiency, the ability to speak

(41) When students talk, what you value most is that \_\_\_\_\_.

- ☐ they get their messages across.
- ☐ they use correct grammar.
- ☐ they use correct vocabulary
- ☐ they use native-like pronunciation and intonation

(42) Under the College English Curriculum Requirement, the objective of College English Education is to foster students' \_\_\_\_\_.

- ☐ skills in reading and listening
- ☐ reading skills
- ☐ overall language proficiency, particularly their competence in listening and speaking
- ☐ skills in speaking and writing
- ☐ skills in reading and speaking

## Appendix F – The Questionnaire

**Please check ☐ the appropriate answer.**

- (43) Do you think that an emphasis on communicative competence might affect the instruction of grammatical elements of the language? ☐ Yes ☐ No
- (44) Do you think the revised CET has caused you to increase the amount of time devoted to teaching aural/oral aspects of English? ☐ Yes ☐ No
- (45) Are you familiar with the goals and objectives of the revised College English Test?  
☐ Yes ☐ No
- (46) Are you familiar with the types of activities on the revised CET? ☐ Yes ☐ No
- (47) Do you know what skills are assessed on the revised CET? ☐ Yes ☐ No
- (48) Do you think a student's language skills are in line with his testing skills? ☐ Yes ☐ No
- (49) Is there conflict between your teaching and testing? ☐ Yes ☐ No
- (50) Do you think that the revised CET can help you improve the way you teach? ☐ Yes ☐ No
- (51) Are you clear of what Communicative Language Teaching (CLT) is about? ☐ Yes ☐ No
- (52) Do you think that the revised CET can give you the direction as to what aspects of language needed to be taught? ☐ Yes ☐ No
- (53) Are you clear of how to implement the new College English Curriculum Requirement?  
☐ Yes ☐ No
- (54) Do your students' College English Test scores contribute to your teaching evaluation by your department head? ☐ Yes ☐ No
- (55) The revised CET has caused me to shift my instructional focus from linguistic knowledge to language use. ☐ Yes ☐ No
- (56) Did you have a lot of exposure to task-oriented activities when you were a student?  
☐ Yes ☐ No
- If Yes, do you think that such exposure benefit you a lot? ☐ Yes ☐ No
- (57) Have you been involved in workshops focusing on teaching methodology? ☐ Yes ☐ No
- If Yes, do you think that the workshops are beneficial to your teaching performance?  
☐ Yes ☐ No

**In the brackets ☐, please mark the following on a four point scale as:**

**[1] never [2] seldom [3] sometimes [4] often**

- (58) ☐ I ask my students to do translation exercises from English to Chinese or from Chinese to English.
- (59) ☐ I lead my students to do simulated tests (模拟题) to prepare for the CET.

---

## Appendix F – The Questionnaire

**Pick out four factors which you think can most affect your teaching activities and then rank them in order of importance.**

- (60)      ☐ professional training  
         ☐ personal level of English-speaking ability  
         ☐ knowledge of advanced teaching methodologies  
         ☐ personal philosophy and beliefs  
         ☐ teaching curriculum  
         ☐ past experience as a language learner (the way I learned English as a student)\_\_\_\_  
         ☐ Band 4 or Band 6 CET  
         ☐ learner expectations  
         ☐ textbooks  
         ☐ class size\_\_\_\_  
         ☐ limited class time  
         ☐ work-load  
         ☐ teaching experience  
         ☐ other factors, Specify \_\_\_\_\_

***We thank you very much for your time!!!***

## Appendix G –Informed Consent Form and Ethical Certificate

### **Informed Consent Form** (for the questionnaire)

You are being invited to participate in a research project conducted by Jing Wang, a PhD student in Second Language Education program at McGill University, Department of Integrated Studies in Education. This research is being conducted to fulfill the requirements for a doctoral dissertation, under the supervision of Dr. Carolyn E. Turner, a professor in the above-mentioned program. Here is a brief description of the research and the reasons why I am collecting data:

**Purposes:** The purpose of this study is to investigate whether tests constitute a constraint on the methodology innovation in China. The goal of this research is to arrive at an accurate description and explanation of teachers' perceptions of tests and test impact, and how the perceptions affect the way they teach.

**Procedures:** As a participant in this study, you will be asked to participate in a questionnaire, which will take you approximately 30-40 minutes. The questionnaire will ask you to answer questions about how you think about the CET and its impact on your instructional practice, and your views about EFL teaching and learning. You will have the opportunity to see the questionnaire before signing this consent form.

Any information about you that is obtained as a result of your participation in this research will be kept strictly confidential. To ensure anonymity, please do not write your name on any page of the questionnaire. By so doing, it will not be possible to match you with your data in any way.

**Conditions of Participation:** Your decision to participate in this project is completely voluntary and you may refuse to participate or withdraw your consent to participate in this study at any time without any penalty. But it could be an enriching experience for you to participate in this project. One direct benefit to you as a participant is that you will have an opportunity to express your views about testing and teaching relevant to your own classroom practice. Furthermore, the study may help you reflect on your own teaching experience. The perspectives you are going to express in this study will help other Chinese English Language Teaching (ELT) researchers and practitioners increase their awareness of the real problems with ELT in China so that Chinese ELT may be directed to the right path of effective teaching. In addition, there are no foreseeable physical and/or psychological risks to be incurred by the participation in this study. Throughout the study, you can decide for yourself whether or not to share your opinions on a given topic.

Please take your time in deciding if you would like to participate. If you need further information or have questions concerning this study, please do not hesitate to contact me by phone at 23716400, or by email (jwang87@po-box.mcgill.ca).

### **Participant's consent:**

I have read the description of the research project and hereby agree to participate. I understand that I will be asked to complete a written questionnaire. I am also aware of the purpose and procedures of this study. I have also been notified that participation is voluntary and that I may withdraw at any point during the study without any consequences to myself. I have also been told my identity will remain confidential.

If you are willing to participate in this study, please sign below.

Signature: \_\_\_\_\_ Date: \_\_\_\_\_