

**Validating the English Language Exam for university entry
in Shanghai, China**

Kin Chung Matthew Book

**A thesis
submitted to the Victoria University of Wellington
in fulfilment of the requirements
for the Degree of Doctor of Philosophy
in Applied Linguistics**

Victoria University of Wellington

2016

Abstract

The University Entrance Exam (UEE) plays a vital role in the Chinese education system. Yet despite it being a high-stakes exam for over nine million senior high school students each year, there is little research demonstrating that the UEE is of an appropriate quality. This is partly due to the very nature of the UEE. Little information about it is publicly available and there has been no significant validation research.

This study provides the first comprehensive attempt to validate the University Entrance English Exam. The study is focused on the Shanghai English Language Exam, which is one of the principal tests for university entry in Shanghai. There were two phases in the study. The first was a qualitative content analysis, which compared the past 10 years of UEE papers with English Curriculum Standards, and with Bachman and Palmer's model of language ability, in both instances with the aim of identifying what the UEE is actually testing. The second phase conducted empirical studies, namely mock exams, classroom observations and interviews, and think-aloud tasks. Six test qualities – reliability, construct validity, authenticity, interactiveness, impact, and practicality – were measured in these studies. The mock exam participants were first-year university students, whereas classroom observations, interviews and think-aloud task participants were senior high school students.

In the first phase the study found that by English Curriculum Standards, the UEE adequately tested language skills and language knowledge but failed to test affect and attitude, learning strategies, and cultural understanding by those same standards. Measured against Bachman and Palmer's model of language ability, the UEE focused on grammatical and textual knowledge and strategic competence but not functional and sociolinguistic knowledge. In the second phase the study found variable responses, with impact registering highest and authenticity registering lowest.

The UEE has been the subject of much controversy, but this study, while critical of it in a number of aspects, finds that overall its performance is above average and better than many of its critics would concede. Given the controversy, the study has important implications for stakeholders, particularly curriculum developers, examination designers, universities and schools, and teachers and students, allowing them to make far better informed decisions concerning this high-stakes exam than has previously been possible.

Acknowledgements

“I can do all things in Him who strengthens me.” Bible Philippians 4:13

「我靠那加給我力量的，凡事都能作！」聖經 腓立比書 四章十三節

First of all, I would like to thank my Lord for His love and grace in my life: without Him, I can do nothing!

I would like to express my deep gratitude to my supervisors Associate Professor Peter Gu and Dr Averil Coxhead for their patience, unfailing support, and academic guidance. Their guidance helped me throughout the research and writing of this thesis.

My sincere thanks go to Dr Malcolm McKinnon for helping with my writing in this PhD journey, Ms Madeleine Collinge from Student Learning Support Services for her valuable comments on my thesis, Elvie Majuddin from Kota Kinabalu, Malaysia for her spiritual support, Ling Huang for her company at my coffee break, and Fiona Ho and Dr Lin He for the precious time they devoted to independent checking of my data.

I would also like to thank the School of Linguistics and Applied Language Studies for providing a caring environment and my PhD colleagues and staff in LALS for their support and help. My special thanks go to Associate Professor John Macalister for valuable conversations in his office.

I am also grateful to friends in Shanghai for their help and support during my months of data collection – especially, Winton Yang, Tony Wang, Duke Hao, Linda Fu, and Derek Xu. I would like to thank Denis Coan for his kind hospitality and Professor Yao Shen from Shanghai University for his encouragement.

Last but not least, I would like to thank my family: my sisters and brothers, nieces and nephews, and great nephews in Hong Kong for their love and care throughout the writing of this thesis. I thank my godparents, Lily and Raymond Cheung for their prayers. My thanks also go to David Weekley for his love, care and support.

I wish to dedicate my thesis to my beloved late parents. As my parents always said to me, “knowledge always belongs to you – no one can take it from you!”

Table of Contents

Abstract	i
Acknowledgements	iii
List of tables	xiii
List of figures.....	xv
Chapter 1: Introduction.....	1
1.1 Background of the study	1
1.2 Rationales	2
1.3 Aims of the study and research design	3
1.4 Structure of the thesis	4
Chapter 2: Literature review.....	5
2.1 Introduction.....	5
2.2 Education and examinations in China	6
2.2.1 Examination system in China	6
2.2.2 Curriculum reform in Shanghai.....	7
2.3 English Curriculum Standards.....	9
2.3.1 Overall design of the ECS	9
2.3.2 Objectives of the ECS	11
2.4 Theories of language ability.....	14
2.4.1 History of language ability theories.....	14
2.4.2 Communicative competence	16
2.4.3 Models of language ability	18
2.5 Validity and test validation	20
2.5.1 Validity	21
2.5.2 Test validation	24
2.5.3 Interpretive and validity arguments	25
2.6 Empirical studies on validation.....	28
2.7 Chapter summary	31
Chapter 3: Methodology	33
3.1 Research questions and methodological framework.....	33
3.2 Phase 1: Qualitative content analysis	36
3.2.1 Materials.....	36
3.2.2 Stage one: English Curriculum Standards.....	36

3.2.3	Stage two: Bachman and Palmer’s model of language ability	37
3.3	<i>Phase 2: Empirical studies</i>	38
3.3.1	Mock exams	38
3.3.2	Classroom observations and interviews	39
3.3.3	Think-aloud tasks.....	40
3.4	<i>Ethical considerations</i>	41
3.5	<i>Chapter summary</i>	42
Chapter 4:	Content validity: English Curriculum Standards.....	45
4.1	<i>Introduction</i>	45
4.2	<i>Research method</i>.....	45
4.2.1	Materials	45
4.2.2	English Curriculum Standards	46
4.2.2.1	Language skills	47
4.2.2.2	Language knowledge	48
4.2.2.3	Affect and attitude.....	49
4.2.2.4	Learning strategies.....	50
4.2.2.5	Cultural understanding	51
4.2.3	Examination Specifications	52
4.2.4	Data analysis.....	53
4.3	<i>Findings and discussion: Part 1 Examination Specifications</i>.....	54
4.3.1	Listening Comprehension.....	54
4.3.2	Grammar and Vocabulary	56
4.3.3	Reading Comprehension	58
4.3.4	Writing.....	62
4.3.5	Speaking	64
4.4	<i>Findings and discussion: Part 2 English Curriculum Standards</i>.....	66
4.4.1	Language skills.....	66
4.4.1.1	Listening skill	66
4.4.1.2	Reading skill	67
4.4.1.3	Writing skill	68
4.4.1.4	Speaking skill.....	69
4.4.2	Language knowledge.....	71
4.4.2.1	Phonetics	71

4.4.2.2	Vocabulary.....	72
4.4.2.3	Grammar	73
4.4.2.4	Functions	75
4.4.2.5	Topics.....	76
4.4.3	Affect and attitude.....	78
4.4.4	Learning strategies.....	79
4.4.5	Cultural understanding	81
4.5	Conclusion	82
Chapter 5:	Content validity: Bachman and Palmer's model of language ability	85
5.1	Introduction.....	85
5.2	Research method	85
5.2.1	Materials.....	85
5.2.2	Bachman and Palmer's model	85
5.2.3	Data analysis	87
5.3	Findings and discussion.....	88
5.3.1	Grammatical knowledge.....	88
5.3.1.1	Vocabulary.....	88
5.3.1.2	Syntax	91
5.3.1.3	Phonology or graphology	93
5.3.2	Textual knowledge	94
5.3.2.1	Cohesion.....	94
5.3.2.2	Rhetorical or conversational organisation	97
5.3.3	Functional knowledge	99
5.3.3.1	Ideational functions	99
5.3.3.2	Manipulative functions	100
5.3.3.3	Heuristic functions	102
5.3.3.4	Imaginative functions.....	102
5.3.4	Sociolinguistic knowledge.....	103
5.3.4.1	Dialects and varieties	103
5.3.4.2	Registers	103
5.3.4.3	Natural or idiomatic expressions	104
5.3.4.4	Cultural references and figures of speech	105

5.3.4.5	Genres.....	105
5.3.5	Strategic Competence	106
5.3.6	Oral English Test	108
5.4	Conclusion	110
Chapter 6:	Construct validity, concurrent validity, face validity and reliability of the English Language Exam	113
6.1	Introduction	113
6.2	Research method	115
6.2.1	Participants	115
6.2.2	Instruments	115
6.2.3	Data collection	117
6.2.3.1	Mock exams procedure	117
6.2.3.2	Scoring	117
6.2.3.3	Post mock exam interviews	117
6.2.4	Data analysis.....	118
6.2.4.1	Descriptive statistical analysis	118
6.2.4.2	Classical item analysis.....	118
6.2.4.3	Validity	119
6.2.4.4	Reliability	119
6.2.4.5	Rasch analysis	119
6.3	Findings and discussion.....	120
6.3.1	Descriptive statistical analysis	120
6.3.2	Classical item analysis	122
6.3.2.1	Facility value	122
6.3.2.2	Discrimination index	123
6.3.3	Validity.....	124
6.3.3.1	Construct validity	124
6.3.3.2	Concurrent validity	126
6.3.4	Reliability.....	128
6.3.4.1	Internal consistency	128
6.3.4.2	Stability and parallel-form reliability	130
6.3.5	Rasch analysis.....	131

6.3.5.1	Test-takers' ability versus test difficulty	131
6.3.5.2	Content aspect of construct validity	135
6.3.6	Post mock exams interviews.....	138
6.3.6.1	Is the UEE able to test your language ability?.....	138
6.3.6.2	Content – the UEE and IELTS compared	138
6.3.6.3	Difficulty – the UEE and IELTS compared	139
6.4	Conclusion	139
Chapter 7:	Consequential validity of the English Language Exam.....	143
7.1	Introduction.....	143
7.2	Research method	144
7.2.1	Purpose.....	144
7.2.2	Participants	145
7.2.3	Instruments.....	145
7.2.4	Materials.....	145
7.2.5	Procedures	146
7.2.6	Post-class observation interviews.....	146
7.2.7	Interviews with UEE designer	146
7.2.8	Data analysis	147
7.3	Findings and discussion	147
7.3.1	Content of teaching	147
7.3.1.1	Grammar and Vocabulary	148
7.3.1.2	Cloze	149
7.3.1.3	Reading.....	150
7.3.1.4	Translation and Guided Writing.....	151
7.3.1.5	Listening and Speaking.....	152
7.3.1.6	Monthly tests	153
7.3.2	Teacher's beliefs and behaviour	154
7.3.3	Content of learning.....	159
7.3.3.1	Materials learned in class.....	159
7.3.3.2	Strategies learned in class.....	159
7.3.4	Learner's beliefs and behaviour	160
7.3.4.1	UEE impact on student lives.....	160

7.3.4.2	The teacher's change of focus as the UEE drew closer	161
7.3.4.3	Private English tutorials	161
7.3.4.4	English learning beliefs and behaviour if there was no UEE	162
7.3.5	Interviews with UEE designer	164
7.3.5.1	Test development process.....	164
7.3.5.2	Proficiency test versus achievement test	166
7.3.5.3	Impact on school and society	166
7.3.5.4	Traditional teaching	167
7.3.5.5	A fair and good examination	167
7.4	Conclusion	169
7.4.1	Teaching	169
7.4.2	Learning.....	170
7.4.3	Textbook.....	171
7.4.4	Curriculum.....	172
7.4.5	Assessment	172
Chapter 8:	Response validity of the English Language Exam	175
8.1	Introduction	175
8.2	Research method	177
8.2.1	Participants	177
8.2.2	Instruments	177
8.2.3	Procedure	177
8.2.4	Data Collection – Recording Details	179
8.2.4.1	Listening Comprehension	179
8.2.4.2	Grammar and Vocabulary.....	180
8.2.4.3	Cloze and Reading Comprehension	180
8.2.4.4	Writing	181
8.2.5	Data analysis.....	181
8.3	Findings and discussion.....	182
8.3.1	Response validity: Do test tasks match the Examination Specifications?.....	182
8.3.1.1	Listening Comprehension	182
8.3.1.2	Grammar and Vocabulary.....	185
8.3.1.3	Cloze.....	188

8.3.1.4	Reading Comprehension	191
8.3.1.5	Translation.....	194
8.3.1.6	Guided Writing	197
8.3.1.7	Summary	199
8.3.2	Interactiveness.....	201
8.3.2.1	Language knowledge.....	202
8.3.2.2	Strategic competence	204
8.3.2.3	Topical knowledge.....	205
8.3.2.4	Affective schemata.....	206
8.3.2.5	Summary	206
8.3.3	Testing other abilities	208
8.3.3.1	Testing mathematics calculation.....	208
8.3.3.2	Guessing the answer	208
8.3.4	Post think-aloud interviews	209
8.3.4.1	What do you think about the UEE?	209
8.3.4.2	Does the UEE test your English proficiency?	210
8.3.4.3	Is the UEE good or bad?	210
8.3.4.4	Is it a fair examination?	211
8.3.4.5	In what ways does the UEE need improvement?	211
8.3.4.6	Communicative English language	211
8.4	Conclusion	212
Chapter 9:	Summary of findings and discussion	213
9.1	Introduction.....	213
9.2	<i>Qualities of the UEE: Summary and evaluation of findings.....</i>	213
9.2.1	Reliability	213
9.2.2	Construct validity	214
9.2.3	Authenticity	216
9.2.4	Interactiveness.....	218
9.2.5	Impact.....	219
9.2.6	Practicality	221
9.2.7	Conclusion	222
9.3	<i>Major issues arising from the study.....</i>	226

9.3.1	Target language use (TLU) domains	226
9.3.1.1	Purpose of the UEE	226
9.3.1.2	TLU Domains: language instructional versus real-life	227
9.3.2	Assessing English Curriculum Standards	229
9.3.2.1	Revisiting the English Curriculum Standards	229
9.3.2.2	School-based assessment	232
9.3.3	UEE reform	233
9.3.3.1	The latest UEE reform	233
9.3.3.2	Impact on tertiary education	234
9.3.3.3	The UEE goals	235
9.3.4	Future assessment system for university entry in China	237
9.3.4.1	Assessment system for university entry in Hong Kong	237
9.3.4.2	Proposed assessment system for university entry in China	238
9.4	Chapter summary	239
Chapter 10:	Conclusion	241
10.1	Introduction	241
10.2	Major findings	241
10.3	Significance of research	243
10.4	Implications	244
10.4.1	Implications for Curriculum developers	244
10.4.2	Implications for UEE designers	244
10.4.3	Implications for universities and schools	245
10.4.4	Implications for teachers and students	245
10.4.5	Implication for textbook publishers	245
10.5	Limitations of the study	246
10.6	Recommendations for further research	247
References	249
Appendices	265
Appendix A:	Human Ethics Approval Letter	265
Appendix B:	Participants Information Sheets and Consent Forms	266
Appendix C:	2008 UEE Paper and 2011 UEE Paper	270
Appendix D:	Marking Standards for Paper II Writing Section	294
Appendix E:	Interview Questions	296

List of tables

<i>Table 1: Objectives for Skills in English Language Subject for University Entry Level</i>	13
<i>Table 2: Summary of Contrasts Between Past and Present Conceptions of Validation (Chapelle, 1999, p.258)</i>	23
<i>Table 3: Progressive Matrix for Defining the Facets of Validity (Messick, 1989, p.20)</i>	24
<i>Table 4: Data Elicitation to the Research Questions</i>	34
<i>Table 5: Components of Language Ability – Analysis Checklist</i>	37
<i>Table 6: Schedule for Research Data Collection and Analysis</i>	43
<i>Table 7: The Examination Format of the UEE English Language Exam</i>	45
<i>Table 8: Curriculum Objectives for Level 8</i>	47
<i>Table 9: Objectives for Language Skills at Level 8</i>	47
<i>Table 10: Objectives for Language Knowledge at Level 8</i>	48
<i>Table 11: Objectives for Affect and Attitude at Level 8</i>	49
<i>Table 12: Objectives for Learning Strategies at Level 8</i>	50
<i>Table 13: Objectives for Cultural Understanding at Level 8</i>	51
<i>Table 14: Objectives of English Language for Examination Specifications</i>	52
<i>Table 15: Checklist Sample</i>	53
<i>Table 16: Listening Comprehension Items versus Objectives of Examination Specifications</i> ...	55
<i>Table 17: Grammatical Features Tested in the Past 10 years’ UEE Papers</i>	58
<i>Table 18: Cloze Items versus Objectives of Examination Specifications</i>	60
<i>Table 19: Reading Comprehension Items versus Objectives of Examination Specifications</i>	62
<i>Table 20: Topic and Type of Writing Tested in the Past 10 years’ UEE Papers</i>	64
<i>Table 21: Checklist for Listening Skill of ECS</i>	66
<i>Table 22: Checklist for Reading Skill of ECS</i>	67
<i>Table 23: Checklist for Writing Skill of ECS</i>	68
<i>Table 24: Checklist for Speaking Skill of ECS</i>	69
<i>Table 25: Checklist for Phonetics of ECS</i>	71
<i>Table 26: Checklist for Vocabulary of ECS</i>	72
<i>Table 27: Checklist for Grammar of ECS</i>	73
<i>Table 28: Checklist for Functions of ECS</i>	75

<i>Table 29: Checklist for Topics of ECS</i>	76
<i>Table 30: Checklist for Affect and Attitude of ECS</i>	78
<i>Table 31: Checklist for Learning Strategies of ECS</i>	79
<i>Table 32: Checklist for Cultural Understanding of ECS</i>	81
<i>Table 33: Summary of Objective Items Achieved in the English Curriculum Standards</i>	82
<i>Table 34: The Examination Format of the UEE English Language Exam</i>	85
<i>Table 35: Components of Language Ability – A Test Analysis Checklist</i>	87
<i>Table 36: Cohesion Tested in the Past 10 years’ UEE Papers</i>	94
<i>Table 37: Topics and Text Types of Writing Tested in Guided Writing</i>	97
<i>Table 38: Text Types Tested in Reading Comprehension</i>	98
<i>Table 39: Text Types Tested in Listening Comprehension (Section B)</i>	98
<i>Table 40: Summary of the Checklist – Items Tested in the UEE</i>	110
<i>Table 41: Details of the UEE Formats</i>	116
<i>Table 42: Figures of Descriptive Statistical Analysis for the 2008 and 2011 UEEs</i>	120
<i>Table 43: Classical Item Analysis – Facility Value for the 2008 and 2011 UEEs</i>	122
<i>Table 44: Classical Item Analysis – Discrimination Index for the 2008 and 2011 UEEs</i>	123
<i>Table 45: Correlation Matrix for the 2008 UEE</i>	125
<i>Table 46: Correlation Matrix for the 2011 UEE</i>	125
<i>Table 47: Correlation Matrix for the Two UEEs and IELTS</i>	126
<i>Table 48: Reliability Coefficients for the 2008 and 2011 UEEs</i>	128
<i>Table 49: Cronbach Coefficient Alphas for the 2008 and 2011 UEEs</i>	129
<i>Table 50: Correlation Matrix for Stability and Parallel-form Reliability – 2008 versus 2011</i> .	130
<i>Table 51: Summary of Measured Person and Item for the 2008 UEE</i>	134
<i>Table 52: Summary of Measured Person and Item for the 2011 UEE</i>	134
<i>Table 53: Item Statistics – Misfit Order for the 2008 UEE</i>	137
<i>Table 54: Item Statistics – Misfit Order for the 2011 UEE</i>	137
<i>Table 55: Different Consequences between Positive and Negative Washback</i>	143
<i>Table 56: Progressive Matrix for Defining the Facets of Validity (Messick, 1989, p.20)</i>	144
<i>Table 57: Time Allocation for Different Teaching Areas in Class</i>	147
<i>Table 58: Objectives of English Language for Examination Specifications</i>	181

List of figures

<i>Figure 1: New Curriculum structure in Shanghai (SMEC, 2004, p.5)</i>	8
<i>Figure 2: The overall design of the National English Curriculum based on English Curriculum Standards for Full-time Nine-Year Compulsory Education and Senior High School (Trial Version; MOE, 2001) and English Curriculum Standards for Senior High Schools (MOE, 2003) (Source: Wang & Chen, 2012, p.93)</i>	10
<i>Figure 3: Framework of the objectives of the ECS (Source: Wang, 2007, p.97)</i>	11
<i>Figure 4: Bachman and Palmer’s model of language ability (1996, p.63)</i>	18
<i>Figure 5: Toulmin’s model of Inference (Source: Bachman, 2005, p.8)</i>	26
<i>Figure 6: Framework of the objectives of the ECS (Source: Wang, 2007, p.97)</i>	46
<i>Figure 7: Narrowing down the English Curriculum Standards</i>	83
<i>Figure 8: Bachman and Palmer’s model of language ability (1996, p.63)</i>	86
<i>Figure 9: Normal distribution curves for the 2008 and 2011 UEEs</i>	120
<i>Figure 10: Person-item map_2008</i>	133
<i>Figure 11: Person-item map_2011</i>	133
<i>Figure 12: Test qualities – validity and reliability</i>	141
<i>Figure 13: Impact of the UEE on different aspects of senior high school education</i>	169
<i>Figure 14: Interactiveness</i>	202
<i>Figure 15: Considerations of test qualities in designing and developing the UEE test tasks</i> ..	225

Chapter 1: Introduction

1.1 Background of the study

Examinations have always been a focus of attention in China. An OECD report (2010) states that teaching and learning in China are mainly driven by the examination syllabi, especially in senior high schools, and that school activities are very much exam oriented. The University Entrance Exam (UEE), commonly known as *Gaokao*, plays a vital role in the Chinese education system. This examination is a prerequisite for entrance into almost all higher education institutions at undergraduate level. In 2015, over nine million senior high school students in China participated in the UEE (China Education Online, 2015). Davey, Lian and Higgins (2007) stress the significance of this National Exam, stating that “competition is fierce, particularly for entry into prestigious universities” (p.385). Since the exam is so substantial, Chinese society tries very hard to facilitate life for the test-takers on exam days. Exam centres are often marked as quiet zones, and nearby construction and even traffic is stopped in order to minimise disruption or distraction to the test-takers during the exam. Police officers, taxi drivers, and other car owners will often transport test-takers they see walking the streets to their exam centres for free, to ensure that they are not late for this important event (Custer, 2015).

Why would English learning be so significant in China? The rapid development of information technology (much of it originating in the U.S.), and globalisation have made English the most widely learned and used foreign language in China, whilst China’s own rapid social, political, and economic development has meant an enormous number of ordinary Chinese are involved in English language learning: in 2011 the *Economist* estimated there were about 300 million English learners in China (Greene, 2011). At present, English is a compulsory subject from grade three in primary school (MOE, 2011). In first-tier cities such as Shanghai, even kindergarten children start learning English. In China, every high school graduate has had to study English for at least 10 years. Chinese, mathematics and English are the main subjects throughout high school. Once at university, students are required to study college English as a compulsory course for at least one year, and they can obtain their bachelor degree only after they pass all the English exams and get enough credits in English courses (Li, 2012). In short, English learning is extremely important in the Chinese education system.

1.2 Rationales

This study analyses in depth the English component of the University Entrance Exam. The UEE has been criticised for a number reasons.

First, both teachers and students suffer tremendous pressure in preparing for and taking the exam: the UEE is one of the most influential examinations in China. As a matter of fact, students can only take the test once a year and examinations are always the “yardstick” by which schools and education are measured in society (Dello-Iacovo, 2009, p.247).

Second, the inflexible examination format of the UEE narrows the content of learning (Hu, 2002). The UEE tests subject knowledge and theory rather than the ability of problem-solving or the carrying out of practical tasks (Davey et al., 2007). Most people seem to accept that the examination system is squeezing creativity out of students (Sudworth, 2012).

Third, the importance of the UEE influences all teaching of English in Chinese classrooms, sometimes in negative and limiting ways. Qi (2004) has found that “teaching to the test” is a widespread practice, which has hindered wider implementation of the curriculum (p.181). The format of the UEE has become the format for any English test in any school (Liu, 2010).

Fourth, according to Gu (2012), there is no communication between test developers and curriculum designers. Gu in fact suspects that the test developers do not read the curriculum standards carefully or analyse what kind of language ability should be tested. The UEE does not faithfully test the curriculum.

Fifth, some university professors, however, have queried whether the UEE is sufficient to identify the best students, especially for English majors. Yet the major purpose of the UEE is to select, on the one hand, students who have thoroughly mastered the English curriculum and on the other hand the most talented students who will study academic subjects for limited places at universities and colleges (Qi, 2010).

There have been repeated calls for reform of the UEE. But with little information on the validity of the test, nobody knows which direction reform should take. In particular there has been little research on whether or not these tests are of a suitable quality for such a high-stakes exam as the UEE. Partly on account of the nature of the UEE, very little information is publicly available. No comprehensive validation research has been reported.

This study focuses on the Shanghai English Language Exam, which is one of the tests for university entry in Shanghai. In 1985 Shanghai was the first region in China to develop its own examination papers and ever since then Shanghai has been considered the spearhead for implementing senior high school competency examinations (Hu, 2002). In addition, Shanghai was assigned the task of pioneering curriculum changes in basic education and offering relevant experience for curriculum reforms to other parts of China (Hu, 2002).

In sum, taking into account the controversy over the UEE content, it is crucial to provide policy-makers and other stakeholders with empirical evidence about the UEE so that they can make informed decisions concerning the high-stakes exam. This research is thus significant for both theoretical and practical reasons.

1.3 Aims of the study and research design

This study aims to provide the first comprehensive and open attempt at validating the University Entrance Exam English test. In Messick (1989), validation is “the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores” (p.13). Investigating the validation process requires therefore identifying the relevant theory and evidence, from which it will be possible to validate the particular test or examination that is the subject of the enquiry.

The primary purpose of the study was to explore whether the English Language Exam for university entry in Shanghai tests the language abilities specified in the Curriculum and in Bachman and Palmer’s (1996) model of language ability. The study also aimed to investigate what evidence there was for the test qualities of reliability, construct validity, authenticity, interactiveness, impact, and practicality (Bachman & Palmer, 1996).

There were two phases in this research study: 1: a qualitative and content analysis, which compared University Entrance Exam papers from the past 10 years (2005-2014) with English Curriculum Standards and Bachman and Palmer’s model of language ability, and 2: empirical studies, which included mock exams, classroom observations and interviews, and think-aloud protocols and examined how the Exam satisfies Bachman and Palmer’s criteria for test usefulness.

1.4 Structure of the thesis

This thesis consists of 10 chapters. This chapter has introduced the background of this study, the rationales, the aims of the study and the research design.

Chapter 2 contains a literature review which addresses both background studies on the English curriculum in the Chinese education system and certain key works on theories of language ability, validity and test validation. Gaps in previous research were identified.

Chapter 3 provides a theoretical framework and methodology. It describes how the data was collected and analysed. In order to enrich the data, a multi-method design was adopted and a justification for this approach was provided.

Chapters 4 and 5 compare the content of the past 10 years' UEE papers against the English Curriculum Standards and Bachman and Palmer's model of language ability respectively. Both chapters demonstrate exactly what the UEE tested in the past 10 years (2005-2014).

Chapter 6 empirically validates the English language examination for university entry in Shanghai. Validity (construct validity, concurrent validity, and face validity) and reliability (internal consistency, stability, and parallel-form reliability) were investigated.

Chapter 7 examines the impact of the UEE on English language teaching and learning in a senior school in Shanghai according to the class observations, interviews and artefacts: the exercise is known as consequential validity.

Chapter 8 reports on the study into response validity of the UEE. Think-aloud protocols revealed whether the responses of the participants reflected UEE designer expectations. This chapter also evaluates the extent to which test-takers were involved in completing a test task.

Chapter 9 summarises and evaluates the research findings by using the six test qualities model (Bachman & Palmer, 1996), and discuss the major issues at a macro-level from the findings.

Chapter 10 concludes major findings from this study, presents implications for different stakeholders, discusses limitations, and makes recommendations for further research.

Chapter 2: Literature review

2.1 Introduction

This chapter reviews the literature associated with the main areas of interest in this study. These areas are first, education and examinations in China; second, English Curriculum Standards; third, theories of language ability; fourth, validity and test validation; and finally, empirical studies on validation.

The first section briefly reviews the literature on the education and examinations system in China. Shanghai has always been a pioneer in education to develop its own curricula and syllabuses; therefore, the latest curriculum and examination reform of Shanghai will be introduced.

The second section reviews the newly implemented English Curriculum Standards for Full-time Compulsory Education and Senior High Schools by the Chinese Ministry of Education (MOE). It also illustrates the framework of the objectives of the English Curriculum Standards (MOE, 2003) in detail.

The third section reviews the literature on different models of language proficiency from the 1960s to the 1990s. Also, two major comprehensive models of language ability – Canale and Swain (1980); revised by Canale (1983a) and Bachman and Palmer (1996, and later revised in 2010) – will be described.

The fourth section addresses the literature on the most important issue of all – validity. The traditional and current concepts of validity and test validation will be elaborated in this section. This section also focuses on the interpretations or uses argument for facilitating the validation process by using an argument-based approach. Kane's (2002) argument-based validation and its reformulation by Chapelle, Enright and Jamieson (2008a) will be introduced in this section.

In the final section, the differences between empirical studies on validation related to language testing in China and overseas will be reviewed. The overseas studies include research that has not been replicated in the Chinese context. This study aims to fill the gap.

2.2 Education and examinations in China

Examinations and education are inseparable; examination education has a lengthy history in China. In this section, some research findings related to the examination system in China and the latest curriculum reform in Shanghai are summarised.

2.2.1 Examination system in China

In contemporary China, the Compulsory Education Law stipulates that each child must receive at least nine years of free education, which includes primary education and junior secondary education. Every student has to sit the University Entrance Exam if they want to continue on to tertiary education; this well-known hurdle is called *Gaokao*, which means higher education examination. Dello-Iacovo (2009) notes that the University Entrance Exam is always a popular topic when talking about the education system in China as the exam is considered “the baton of education” (p.247). Most principals in senior high schools care a great deal about their rate of university enrolment, whilst parents are anxious about the scores their children will obtain in the exam.

Liu and Wu (2006) emphasise that the UEE is a high-stakes exam which is very competitive and significant. This has three main positive consequences (Liu and Wu, 2006). First, the UEE is comparatively fair and the best instrument to measure and identify qualified test-takers for higher education. Second, the UEE is an important tool to improve students’ intelligence and ability as it is strongly linked to primary and secondary education in China. Third, the UEE, being intended to have a positive impact on teaching and learning in secondary schools, should test what the students are taught in class.

Despite the positive education consequences, the UEE has had some negative consequences. The practice of “teaching to the test” is very common in the classroom, teachers teaching test subjects and content only (Liu & Wu, 2006, p.14). As with other public examinations, all of which are ascribed a vital role in Chinese society, this is also a strong obstacle to education reforms. Another consequence is that because examinations are considered the yardstick for measuring school performance, schools could be reluctant to accept the quality-oriented education reforms if their University Entrance Exam results were to be affected (Dello-Iacovo, 2009).

2.2.2 Curriculum reform in Shanghai

Shanghai is the largest city in China with a population of 24.15 million as of 2014. A report from the Organisation for Economic Co-operation and Development (OECD, 2010) indicates that, of the four first-tier cities in China, Shanghai was earliest to establish universal primary and secondary education. Tan (2012) stresses that Shanghai was the first city to implement nine-year compulsory education (five years of primary education and four years of junior secondary education) and students may continue their senior secondary education for another three years after which they can sit the University Entrance Exam. Hu (2002) states that when the Ministry of Education staged a new policy about curriculum development, Shanghai was always considered a pioneer in education in China in developing its own curricula and syllabuses. The OECD (2010) also reports that English was allotted more instruction time in class in Shanghai, and that more resources were provided for teaching and learning English in schools than in other provinces. The Shanghai curriculum reform follows the general framework of national curriculum reform. In advance of implementing national curriculum reform, Shanghai has often been a pioneer.

Shanghai has launched two phases of curriculum reform since 1988 (Hu, 2002). The first phase started in 1988. At that time Shanghai established its Curriculum and Teaching Materials Reform Commission (SCTMRC) and began curriculum reform (SCTMRC, 1998a, 1998b). Students were allowed to choose courses for personal reasons. A curriculum of three blocks was established: compulsory courses, elective courses, and extra-curricular activities. The second phase was implemented in 1998. The OECD (2010) also reports that the aim of the whole reform was to transform students from passive learners into active participants in acquiring knowledge in order to develop their creativity and potential. There are three components in the new curriculum: the basic curriculum, comprising standard compulsory subjects, to promote students' grasp of the essentials of literacy and numeracy; the enhanced curriculum, to cultivate students' interests and potentials; and the research curriculum, to develop independent learning and inspiration through extra-curricular activities and help them to conduct their research topics through their life experiences. The Shanghai Municipal Education Commission (SMEC, 2004) expected the students to 'learn to learn', and think creatively and critically through learning and exploring in their studies. At the same time, traditional courses were re-scheduled into eight learning fields: language and literature,

mathematics, social sciences, natural science, technology, physical education, arts, and integrated practicum. Twenty subjects were involved within the academic courses, extra-curricular activities and seminars in these eight fields. They were in the form of modules and themes. Schools were encouraged to develop their own activities in order to meet their individual needs. The structure of the new curriculum in Shanghai can be seen in Figure 1.

Figure 1: New Curriculum structure in Shanghai (SMEC, 2004, p.5)

The OECD report (2010) said that examination reform in Shanghai has been another controversial topic. In 1985, Shanghai obtained authorisation to set up an independent tertiary education entrance examination. Since then, much effort has been put into reforming assessments and examinations whilst exam reform has also had to be matched with changes in curriculum and pedagogy. The independent university examination, now separated from the national system, allowed for other changes: it created “a comprehensive platform for reforming the curriculum” (OECD, 2010, p.94). The entrance examination comprises three core subjects – Chinese language, English language, and mathematics – and optional subjects (physics, chemistry, biology, geography, history, and politics). Since 1998, the exam has adopted the form of ‘3 + X’ system: three core subjects plus one or more additional subject(s) from the list as required by the tertiary institutions (Davey et al., 2007). The content of the ‘X’ component may cover one or more disciplines and the universities will make the final decision on the weight of the core subjects and ‘X’ component.

2.3 English Curriculum Standards

This section reviews the newly implemented English Curriculum Standards for Full-time Compulsory Education and Senior High Schools in China by the Chinese Ministry of Education. The ECS was so important because the University Entrance Exam was required to adjust to meet the new Standards. In this part, the framework of the objectives of the ECS is described in detail.

2.3.1 Overall design of the ECS

English Curriculum Standards for Full-time Compulsory Education and Senior High Schools (Trial Version) were released in 2001 (MOE, 2001), and English Curriculum Standards for Senior High Schools (Trial) were introduced in 2003 by the Chinese Ministry of Education (MOE, 2003). The former Standards, which were more general, covered both nine-year compulsory and senior secondary education, whereas the latter focused more specifically on senior secondary schools. Gu (2012) mentions that MOE published English Curriculum Standards for Compulsory Education, which is considered as a final version of the 2001 trial version, in December 2011 (MOE, 2011), and the new curriculum standards were finally implemented in September 2012 in the whole country.

The design for Senior High School National English Curriculum Standards was founded on five basic concepts (MOE, 2003):

1. Emphasise common ground and build a platform for future development;
2. Provide a wide selection and adapt to individual needs;
3. Optimise learning methods and improve self-learning ability;
4. Care for students' affective needs and enhance human quality;
5. Improve an assessment system and promote students' continuous development.

The Senior High School English Curriculum is a continuation of the nine-year compulsory education curriculum. The learning targets are divided into nine different English language proficiency levels with different required sets of standards as illustrated in Figure 2. Both of them act as two inter-related stages. Wang and Chen (2012) explain that the first stage covers nine-year compulsory education from grade 3 to grade 9 with five different English language proficiency levels, whereas the second stage steps into senior secondary education with four

higher levels from level 6 to level 9. Level 7 is established for all senior high school graduates, level 8 for University Entrance Exam takers, and level 9 for foreign language school graduates. In order to reflect the ideas of the curriculum reform, the English Curriculum Standards have adopted a combination of required and optional courses. Required courses enable students to acquire basic language ability and a positive attitude towards learning, flexible learning strategies and cross-cultural understanding and ability, whereas the purposes of the optional courses are to satisfy students' needs and develop individual interests and potentials (MOE, 2003).

Figure 2: The overall design of the National English Curriculum based on English Curriculum Standards for Full-time Nine-Year Compulsory Education and Senior High School (Trial Version; MOE, 2001) and English Curriculum Standards for Senior High Schools (MOE, 2003)
(Source: Wang & Chen, 2012, p.93)

The English Curriculum Standards (ECS) for Senior High Schools introduced in 2003 (MOE, 2003) made substantial changes compared to the previous English Syllabus in 1993. The main features are (Wang & Chen, 2012; Wang & Lam, 2009):

1. The objectives of the ECS focus not only on the language knowledge and skills, but also affect and attitude, learning strategies and cultural understanding;

2. The achievement standards according to students' interests and ability become more flexible, consistent and unique;
3. Two modules (one compulsory and one elective) allow all students to underpin a stronger foundation in English language and develop their potential with diverse competences;
4. The learning and teaching goals are defined very specifically at each level and both language knowledge and skills are integrated by engaging students in the language use;
5. The assessment system, which appraises student achievement, covers both summative and formative assessments, in order to encourage students' learning autonomy.

2.3.2 Objectives of the ECS

Figure 3: Framework of the objectives of the ECS (Source: Wang, 2007, p.97)

The Chinese Ministry of Education (2003) claims that the overall goal of the ECS in basic education is to promote students' general ability in language use, based on the comprehensive development of their language skills, language knowledge, affect and attitude, learning strategies, and cultural understanding. Wang and Lam (2009) believe that the new curriculum would enhance students' critical thinking and analytical ability and language use ability. Figure 3 illustrates the framework of the objectives of the English Curriculum Standards (MOE, 2003; Wang, 2007).

Wang (2007) states that the overall structure of the ECS is “the most comprehensive ever designed” (p.96). Appropriate language skills play an important role in students’ language communicative ability. Language skills take into account ability in not only the four sections (listening, speaking, reading, and writing), but in the whole. Language knowledge in basic education, as an important part of language abilities and the foundation of language skills, consists of phonetics, vocabulary, grammar, functions, and topics. Affect and attitude refer to interest, motivation, confidence, perseverance, and team spirit, which influence the students’ learning process and effect. It also takes in students’ national awareness and international horizon. Learning strategies refer to various kinds of actions and steps students take to study and improve effectively. They include cognitive strategy, control strategy, communication strategy and resource strategy. Cultural understanding includes knowledge, understanding and awareness. Language contains abundant cultural connotations and cannot be separated from culture. Cheng (2011) expects students to explore the cultures, lifestyles, and values of English-speaking countries.

Among these objectives, language skills and language knowledge are the basis, while affect and attitude are critical factors. Furthermore, a well-developed cultural understanding ensures an appropriate use of language, and wise learning strategies allow students to improve their efficiency and initiative. Cheng (2011, p.137) agrees with the ECS that “language is not only a tool for communication”, but also an instrument for thinking, learning and social involvement. Students are able to develop positive attitudes and values, enhance their life experiences, broaden their world horizon and heighten their thinking skills through learning a foreign language (MOE, 2001, 2003).

One of the most significant characters in the ECS is “the creation of a coherent set of targets” within the whole school system (Gu, 2012, p.47). Can-do statements, representing concrete targets to be achieved, are specified in the English Curriculum Standards. The following objectives are listed for the skills of listening, speaking, reading, and writing in English language subject for university entry level (MOE, 2003):

Table 1: Objectives for Skills in English Language Subject for University Entry Level

Skills	Description of Objectives
Listening	<ol style="list-style-type: none"> 1. Can recognise different attitudes through different tones; 2. Can understand discussions or conversations on familiar topics, and remember the main points; 3. Can grasp the viewpoints in simple passages; 4. Can figure out the themes or main ideas in English broadcasting and television news; 5. Can understand euphemistic / implied meanings of suggestions or advice.
Speaking	<ol style="list-style-type: none"> 1. Can use proper intonations and rhythms; 2. Can discuss and then make plans in accordance with learning tasks; 3. Can report the processes and results of experiments or research; 4. Can make 3-minute prepared speeches on common topics; 5. Can make effective use of the English language to judge, criticise, complain, or express opinions in daily life communications; 6. Can do translations in such daily life activities as shopping and sightseeing.
Reading	<ol style="list-style-type: none"> 1. Can understand different opinions and attitudes in reading materials; 2. Can recognise the features of different types of writing; 3. Can understand long sentences and difficult sentences by analysing sentence structures; 4. Can enjoy basic literature with the help of teachers; 5. Can get and then process the information required by learning tasks from e-books or the Internet; 6. Should do as much reading as over 360,000 words outside textbooks.
Writing	<ol style="list-style-type: none"> 1. Can write coherent and structurally complete compositions; narrating events or expressing your own opinions and attitudes; 2. Can paraphrase or summarise articles after reading; 3. Can write in appropriate styles with natural and grammatical sentences; 4. Can write short essays or reports based on textual and graphical information provided.

The new senior high school ECS has had a strong impact on English teachers since it was unveiled. Many teachers may find that the teaching practices of the new curriculum do not match with their teaching beliefs. Wang and Chen (2012) think that teachers who are not confident and competent enough find it difficult to teach to new curriculum standards and requirements. In addition, there is a lack of communication between curriculum developers and teachers; some teachers neither comprehend the concepts from the curriculum standards nor know how to implement it in the classrooms. Gu (2012) believes that inadequate training and resources may frustrate teachers in their efforts to smoothly implement the new curriculum standards in schools.

The New English Curriculum Standards has led to change in the University Entrance Exam as the form and content of the exams are adjusted to influence classroom learning and teaching. However, students still expect to learn grammar and vocabulary in class in order to deal with the examinations. Wang and Chen (2012) comment that teachers and students are still interested in maintaining or increasing the number of students who can enter universities. As a result, both teachers and students are reluctant to try new learning methods. Gu (2012) indicates that formative assessment is promoted in the ECS in order to monitor and upgrade the learning and teaching process, but very little has been done to introduce such assessment. Summative assessment is still the main way of appraising language learning.

2.4 Theories of language ability

In this part of the review, the most important research findings about theories of language ability with regard to language proficiency, communicative competence and strategic competence are summarised. In addition, two major comprehensive models of language ability will be described in detail.

2.4.1 History of language ability theories

The first framework regarding the measurement of language proficiency was proposed in the early 1960s by Lado (1961) and Carroll (1961). The models mention language skills – listening, speaking, reading, and writing – and differentiate those skills from language knowledge, which includes vocabulary, syntax, morphology, phonology and graphology. However, the relationship between skills and knowledge was not revealed.

Language performance is referred to as “the actual manifestation of linguistic competence ... in behaviour” (Carroll, 1968, p.50). Carroll (1961) also considers those language skills as different kinds of performances of the language structure. However, Lado (1961) did not mention the definition of performance, only which elements and skills should be included in language tests. Chomsky (1965) presents the idea of *competence*, which involves linguistic knowledge, versus *performance*, which is the actual language-use unaffected by “grammatically irrelevant conditions” (p.3). Habermas (1970) retains Chomsky’s idea of competence and performance but comments on competence as a mono-logical capability, which means single dimensional communication.

Hymes's theory of communicative competence is believed to be "the most influential general discussion of language performance in applied linguistics" (as cited in McNamara, 1996, p.154). Hymes (1972) makes a very clear distinction between *competence*, which depends on both knowledge and ability for use, and *performance*. Hymes (1972) indicates that there is an ambiguity in Chomsky's concept of performance, noting that Chomsky fails to justify the sociocultural situation. He clarifies the usage of two contrasts: 1: (underlying) competence versus (actual) performance, and 2: (underlying) grammatical competence versus (underlying) models/rules of performance (p.280). According to Hymes (1972), contextual sociocultural factors should be considered as part of language use.

Similarly, Halliday's (1976) interpretation of language functions, van Dijk's (1977) investigation of the relationship between text and context, Morrow's (1977) operation of a situational syllabus through contextualisation, and Munby's (1978) idea of contextual appropriateness highlight the importance of the context beyond the sentence to the appropriate language use in both the surrounding discourse and in sociocultural situations.

The function of language is to communicate (Kramsch, 1986; Savignon, 1983; Widdowson, 1983). Savignon (1983) describes communication as "the negotiation of meaning between two or more persons" (p.8). It is a dynamic notion – competence can be described as an "interpersonal" rather an "intrapersonal" characteristic (p.8). She adds that one's performance in any particular situation reveals the interaction between one's competence, the competence of others, and that particular occasion's character. Kramsch (1986) presents a similar concept, noting that "interaction always entails negotiating intended meaning [and] anticipating the listener's response and possible misunderstandings, clarifying one's own and the other's intentions and arriving at the closest possible match between intended, perceived, and anticipated meanings" (p.367).

To summarise, from the early 1960s to the 1990s, literature about *language proficiency* was not just about knowledge and skills; it was further extended to knowledge of how language is communicated in order to achieve one's communicative target in a particular context. *Language proficiency* also involves interaction between different parties as a dynamic process. Vollmer and Sang (1983) point out that *language proficiency* is one's generalised ability to make use of a language as a tool of social interaction in similar future situations. For Cummins (1983), *language proficiency* can be measured along two ranges: "context-embedded and

context-reduced” (p.120). Context-embedded communication relies on hidden knowledge being shared and understood by the interlocutors, while context-reduced communication requires speakers to elaborate on messages and to use explicit knowledge in order to avoid misunderstanding.

2.4.2 Communicative competence

Canale (1983b) sets up three dimensions of *language proficiency* for assessing different levels of competence: basic language proficiency focused on biological elements of language; communicative language proficiency based on social, interpersonal uses of language through productive media, and autonomous language proficiency entailing “intrapersonal uses of language such as problem solving” (pp.339-340).

As Bachman (1990) writes, communicative language ability can be described as “consisting of both knowledge, or competence, and the capacity for implementing, or executing that competence in appropriate, contextualized communicative language use” (p.84). Bachman’s framework includes three components: language competence, strategic competence, and psychophysio-logical mechanisms.

Proficiency is a term that implies variability, and language proficiency has been “associated with language measurement and testing” (Llurda, 2000, p.88). McNamara (1996) claims that the most influential discussions about the model of language ability are Canale and Swain (1980); and revised by Canale (1983a) and Bachman and Palmer (1996, and later revised in 2010). All models of language ability, McNamara believes, have three dimensions – ability to know a language (knowledge), ability to use language (performance), and actual language use.

The first comprehensive model of language ability is that of Canale and Swain (1980), further elaborated by Canale (1983a), which posited four components for *communicative competence*:

- Grammatical competence: “knowledge of lexical items and of rules of morphology, syntax, sentence-grammar semantics, and phonology” (p.29);
- Sociolinguistic competence: knowledge of “sociocultural rules of use” (p.30);
- Strategic competence: the possession of “coping strategies” (p.31);
- Discourse competence: ability to deal with extended use of language in context.

Communicative competence is therefore made up of knowledge and the skill required to demonstrate this knowledge in actual language performance (actual communication). Fulcher

and Davidson (2007) comment that this model has three significant implications for language testing. First, the “distinction between communicative competence and actual performance” suggests that test items or tasks not only measure knowledge (competence) but also require test-takers to demonstrate their knowledge (performance) in a meaningful circumstance (p.39). Second, discrete-point tests can also be valuable tools for measuring communicative competence. Third, a more “fine-grained” model would facilitate with developing criteria for assessing language performance at different proficiency stages (p.39).

In comparison with Canale and Swain, Bachman and Palmer’s (1996) model defines language ability as involving two components: “language knowledge and strategic competence” (p.67). It focuses on interactions between language ability, personal attributes, topical knowledge, and affective schemata, and showing how these areas interact with each other in language use circumstances or test tasks (See Figure 4). Language knowledge includes two broad categories: organisational knowledge and pragmatic knowledge.

Organisational knowledge involves controlling the formal language structure for producing correct sentences while pragmatic knowledge allows people to decide the intended meaning of discourse by relating texts to their meanings in suitable contexts.

There are two areas of organisational knowledge (p.68):

- Grammatical knowledge: vocabulary, syntax, phonology, and graphology;
- Textual knowledge: cohesion and rhetorical or conversational organisation.

Similarly, there are two areas of pragmatic knowledge (p.69):

- Functional knowledge: ideational, manipulative, instrumental, and imaginative;
- Sociolinguistic knowledge: dialect or variety, registers, natural or idiomatic expressions, cultural references and figures of speech.

Strategic competence is considered to be “a set of metacognitive strategies” which integrate language and topical knowledge and affective schemata, and provide a cognitive management function in language use (p.70). Metacognitive components involve the areas of goal setting, appraising, and planning in executive process use.

Figure 4: Bachman and Palmer's model of language ability (1996, p.63)

2.4.3 Models of language ability

Canale and Swain (1980) point out that *a coping strategy* can be mastered only in “real-life communication situations not through classroom practice that involves no meaningful communication” (p.31). However, Canale (1983b) supplements the definition of strategic competence, arguing that it should comprise the compensatory trait of communication and the enhancement trait of expression, both to be attained by mastering “verbal and non-verbal communication strategies”, which is a substantial part of *strategic competence* in communication (p.339).

The planning process in strategic communication is described as an interaction of the communicative goal, individual communicative resources and the evaluation of the communicative condition while the process executes the plan through a “neurologic and physiological process, resulting in language use” (Færch & Kasper, 1983, p.27). However, this model is limited to the use of communication strategies in interlanguage communication.

Celce-Murcia, Dörnyei and Thurrell (1995) define *strategic competence* as “knowledge of communication strategies and how to use them” (p.26). They emphasise three strategic use functions from three distinct viewpoints: problem-solving in the planning and execution stages to achieve a communication goal from a psycholinguistic perspective; seeking help from an interactional perspective, and maintaining communication pathway from communication continuity and maintenance perspective. However, they do not mention situations where competence might be used, as Canale or Bachman do.

Bachman (1990) separates strategic competence from language knowledge. The former includes recognising the use of language in a dynamic process, assessing related information in the context and negotiating meaning in order to achieve a communicative goal. It is defined as “a general ability, which enables an individual to make the most effective use of available abilities in carrying out a given task” (p.106). He also perceives *strategic competence* as “an important part of all communicative language use” instead of interlanguage communication (p.100). Bachman and Palmer, in their modified model (2010), re-define *strategic competence* as a set of metacognitive strategies, which are executive processes that allow test-takers to engage in “goal-setting, appraising, and planning” (p.49).

Test-takers decide, by using strategic competence, how they are going to proceed with the test (goal-setting), assess what the situation and their resources are (appraising), and determine how to use the resources they have (planning) (Llurda, 2000). *Goal-setting* involves identifying different language-use tasks or test-tasks, choosing if options are given, and deciding whether or not to try to finish the tasks in the exam (Bachman & Palmer, 2010). The purpose of taking the exam is to complete as many of the question-items and test-tasks correctly as test-takers can, using the language – this is their goal. *Appraising* entails test-takers identifying what competence is needed to finish the tasks. It involves assessing the characteristics of the language-use or tests, assessing the individual’s own knowledge (topical and language), and assessing the accuracy of the answers to the test tasks which relate to the grammatical, textual, functional, and sociolinguistics knowledge of the response. *Planning* involves determining how to incorporate language, topic knowledge, and affective schemata in order to achieve the communicative goal or to finish the test tasks successfully.

Language use is described as “the dynamic and interactive negotiating of intended meanings between two or more individuals in a particular situation” (Bachman & Palmer, 2010, p.34).

They add that when more than one language participant is engaged directly in the activities (Widdowson, 1978), it is referred to as reciprocal language use; otherwise, it is non-reciprocal.

Compared to Canale and Swain's model, there are three significant features in Bachman and Palmer's model. First, affective schemata are involved in language use; test-takers' affective schemata in combining with other characteristics of the particular tasks may affect their performance in completing the test tasks. Second, strategic competence is not "part of language competence" (McNamara, 1996, p.69). It is reinforced as the central role in the model. Third, this strategic competence is the main component connecting other components inside the individual. Therefore, the model recognises the issue of the 'interaction' of the components of communicative competence. This model provides the cognitive link between the characteristics of the language use task.

In this research, Bachman and Palmer's model was adopted, for the following reasons:

1. McNamara (1996) claims that this model is a more coherent and consistent model compared to Canale and Swain's model, which did not identify the relationship between the four competence components;
2. As mentioned above, strategic competence is separated from language knowledge; McNamara (1996) describes strategic competence as "*being more properly part of ability for use, not knowledge*" (original italics; p.69). Bachman and Palmer's model is a more adequate model than that of Canale and Swain, in that it helps to clarify the concept of language performance in test tasks; hence, it helps to make inferences by appraising language ability;
3. This model is considered to be easier to understand and described as a "refinement and elaboration" of Canale and Swain's work (McNamara, 1996, p.66).

2.5 Validity and test validation

The literature about validity and validation of language testing dates back to Robert Lado (1961). Bachman and Palmer (1996) state that the main purpose of language testing is to make inferences about language ability, whereas McNamara (1996) describes those models of language ability as "theoretical rationales" for inferences about test-takers' abilities through their test performance (p.49).

2.5.1 Validity

Validity in language testing and assessment has generally been defined as being to discover whether a test measures accurately what it is claimed or intended to measure (Davies, 1990; Fulcher, 2010; Hughes, 1989; Lado, 1961; McNamara, 1996); and whether it unveils the “appropriateness of a given test or any of its component parts as a measure of what it is purported to measure” (Henning, 1987, p.89). Heaton (1988) comments that when one starts to write a test, one has an intention to know and find out if that test actually does measure what it is intended to test.

Lado (1961) defines validity as follows: “Does a test measure what it is supposed to measure? If it does, it is valid” (p.321). Chapelle (1999) describes Lado’s portrayal of the “all-or-nothing” character in a language test (p.255). However, some researchers disagree with that approach, suggesting instead that validity is a matter of degree. Henning’s (1987) definition, for example, allows for degrees of validity; tests are more or less valid for particular purposes – validity is not an “all-or-nothing matter” (p.89). Similarly, Messick (1989) emphasises that “it is important to note that validity is a matter of degree, not all-or-none” (p.13). Weir (2005) agrees that validity should be regarded as a relative concept.

In the traditional view, correlational methods were perceived as the central devices to validation (Chapelle, 1999). Most researchers have identified three main types of validity. *Content validity* (also called *rational validity*) depends on a theoretical and logical analysis of the test’s content to see if its content constitutes a representative sample of the relevant language skills and structures. *Criterion validity* (also called *empirical validity*) depends on empirical and statistical evidence so as to agree with those provided by some independent assessment of the students’ ability. Finally, *construct validity* refers to what the test scores actually mean and the tester’s theories and predictions about the test are reflected in the test scores (Alderson et al., 1995; Hughes, 1989; Kunnan, 2004; Morrow, 1979; Popham, 2013; Weir, 1990).

In the 1980s, language testing researchers tried to identify different types of validity by adding different elements or rearranging different definitions. This has been discussed by Carol Chapelle (1999). Henning (1987) lists five types of validity by including *response validity*, which is described as “the extent to which examinees responded in the manner expected by the test developers” (p.96), and *concurrent and predictive validity*, which was separated from criterion

validity. Hughes (1989) considers that another way to collect evidence about the construct validity of a test is to examine what test-takers think when they respond to question-items or test-tasks, another term for *response validity*; thinking aloud and retrospection. Heaton (1988) distinguishes *face validity* – a superficial inspection of the test items would be sufficient to reveal if the test is valid or not – from *content validity*; most testers make no difference between them (Henning, 1987). Madsen (1983) identifies affect – the degree to which unnecessary anxiety is caused by the test – as a third test feature of concern after validity and reliability. Instead of affect, Hughes (1989) presents *washback* – the impact of testing on learning and teaching – as the third test element, or the current understanding of test consequence as *consequential validity* (O’Sullivan & Weir, 2011). Canale (1987) includes the topic of *ethics* in his review of language testing on top of issues related to validity.

Reliability has been considered a quality of language tests distinct from validity. However, most language testing researchers believe that reliability is a prerequisite for validity (Oller, 1979). For Campbell and Fiske (1959) the difference between validity and reliability is that reliability is “the agreement between two efforts to measure the same trait through maximally similar methods” whereas validity is “represented in the agreement between two attempts to measure the same trait through maximally different methods” (p.83). If a test is valid, it should be reliable. However, the reverse is not necessarily true; that is, if a test is reliable, it may not be necessarily valid. Reliability is said to be a “necessary condition for validity, in the sense that test scores that are not reliable cannot provide a basis for valid interpretation and use” (Bachman, 1990, p.289).

Messick (1989, 1990) argues that the traditional conception of validity is fragmented and incomplete, and proposes a new approach to validity that regards it as a unified concept more concerned with judgement than statistics. Messick (1989) defines *validity* as follows:

Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment. (Emphases in original) (p.13)

Traditionally, the study of validity only focused on different types of evidence to support a particular meaning or use in the test. However, the examination of validity is no longer restricted to factual evidence to support a given test interpretation or use, or reliabilities and

correlations with other tests. The investigation of validity becomes more complex when considering both the educational and social consequences of test uses, because most of the testing occurs in an educational or social environment (Bachman, 1990; Messick, 1989). The consequential aspects of validity, including washback, impact, and social consequence, have been mentioned in different language testing literature (e.g., Alderson & Wall, 1993; Davies, 1997).

The AERA/APA/NCME Standard (1985, 1999) for educational and psychological testing replaces the previous definitions of three types of validity with a unitary concept of validity that sees construct as central and superordinate for test validities, with content and criterion analyses providing evidence for supporting and investigating construct validity. These are “not alternatives but complementary aspects of an evidential basis for test interpretation” (Weir, 2005, p.13). Chapelle (1999) describes Bachman’s (1990) concept of validity as the most “influential” milestone in the 1990s (p.257), which devises “a single unified view of validity” referring to test interpretation and use (p.256). Bachman (1990) emphasised that the inferences are based on the test scores, and their uses are the main purpose of validation. Table 2 summarises significant changes in the way validation has been interpreted.

Table 2: Summary of Contrasts Between Past and Present Conceptions of Validation
(Chapelle, 1999, p.258)

Past	Present
Validity was considered a characteristic of a test: the extent to which a test measures what it is supposed to measure.	Validity is considered an argument regarding test interpretation and use: the extent to which test interpretations and uses can be justified.
Reliability was seen as distinct from and a necessary condition for validity.	Reliability can be seen as one type of validity evidence.
Validity was often established through correlations of a test with other tests.	Validity is argued on the basis of a number of types of rationales and evidence, including the consequences of testing.
Construct validity was seen as one of three types of validity: content, criterion-related, and construct validity.	Validity is a unitary concept with construct validity as central while content and criterion-related evidence can be used as evidence for construct validity.
Establishing validity was considered within the purview of testing researchers responsible for developing large-scale and high-stakes tests.	Justifying the validity of test use is the responsibility of all test users.

Messick (1989) argues that it is insufficient to use individual traditional approaches to validation (including content, criterion-related, construct validity) as evidence to support test interpretations. He proposes “a unified validity framework” to reveal validity as a unitary but multi-faceted concept with “construct validity” but with an extra element in each facet (Messick, 1989, p.20). Rows represent *the source of justification of the testing*, which includes the consideration of either evidence or consequence, or both, and columns record *the function of outcome of the testing*, which includes either test interpretation or use, or both. When justifying a particular interpretation of a test score, evidence for construct validity and the value implications of this interpretation must be considered. If this test score is used for a particular purpose, the relevance or utility of the particular use and the social consequences should be taken into account, not only construct validity and value implications. Table 3 shows the progressive matrix for defining the facets of validity.

Table 3: Progressive Matrix for Defining the Facets of Validity (Messick, 1989, p.20)

<i>Source of justification of testing</i>	<i>Function of outcome of testing</i>	
	Test Interpretation	Test Use
Evidential Basis	Construct validity	Construct validity + Relevance / Utility
Consequential Basis	Construct validity + Value implications	Construct validity + Relevance / Utility + Value implications + Social consequences

2.5.2 Test validation

Bachman (1990) depicts validation as “a general process that consists of the marshalling of evidence to support a given interpretation or use, a process that is based on logical, empirical, and ethical considerations” (p.238). Validation is empirical evaluation of the meaning and consequences of measurement; the term “empirical evaluation is meant to convey that the validation process is scientific as well as rhetorical and requires both evidence and argument” (Messick, 1989, pp.253-254). Validation can be regarded as a form of evaluation, in which evidence is created to support inferences from test scores by using different quantitative and qualitative methodologies (Weir, 2005).

The main purpose of test validation is not to examine the validity of the test content or the test scores, but to observe the process by which the collected evidence is collected, used, and

interpreted. To establish if a test or test score is valid or not, we have to examine the specific abilities the test is intended to test; otherwise, the test is meaningless (Bachman, 1990). Therefore, the test validation is crucial because its aims in language testing are to provide evidence that the “underlying theoretical constructs being measured are themselves valid” (Henning, 1987, p.98) and to safeguard the “defensibility and fairness of interpretations based on test performance” (McNamara, 2000, p.48).

The most significant issue in designing, developing and evaluating a language test is its “usefulness” (Bachman & Palmer, 1996, p.17). Bachman and Palmer’s (1996) model of test usefulness comprises the following qualities: reliability, construct validity, authenticity, interactiveness, impact, and practicality (p.17). This model is regarded as the essential basis for quality control during test design, development and evaluation.

In the operationalisation of this model of test quality, Bachman and Palmer (1996) suggest three principles: 1: “overall usefulness” of the test should be maximised, over the individual qualities that influence usefulness; 2: the individual test qualities cannot be assessed independently, and 3: “test usefulness and the appropriate balance among the different qualities cannot be prescribed in general” but must be specific to each testing situation (p.18). In this research, Bachman and Palmer’s model will be used to check if evidence showing these test qualities can be found in the exam.

2.5.3 Interpretive and validity arguments

Recent research (Bachman, 2005; Bachman & Palmer, 2010; Chapelle et al., 2008b, 2010; Kane, 1992, 2001, 2013; Mislevy, Steinberg, & Almond, 2002, 2003) using interpretive and validity arguments are based on Toulmin’s (1958, 2003) argument structure. Toulmin (1958) developed a general structure to justify building claims by establishing observations on the basis of data. Figure 5 shows Toulmin’s model of the validity arguments.

Toulmin’s model is applicable to individual inferences from an argument. Each *inference* makes a conclusion, which is referred as a *claim*, by beginning from a datum. The inference depends on a *warrant*, which is a general rule for inferring the claims of the argument. Warrants are supported with *backing*, which is evidence. However, there are *exceptions* that in some situations may undermine the inference, as a result of disproving the interpretive argument; that is *rebuttal*. This rebuttal weakens the inferential link between the datum and the claim.

Figure 5: Toulmin's model of Inference (Source: Bachman, 2005, p.8)

Messick (1989) proposes a new approach to validity, which regards it as a unified concept. Such a unified concept of the construct validity is “elegant and conceptually rich and suggestive” as described by Kane (2012, p.8); however, it is difficult to implement it effectively for the validation of test score interpretations and uses. Kane (1992, 2006) states that validation is an overall plan to evaluate the interpretations and uses of test scores as proposed in the interpretive argument. He employs two kinds of argument for validation: an *interpretive argument* that specifies the proposed interpretations or uses by laying out pertinent claims, while the *validity argument* provides an evaluation of the claims in the interpretive argument (Kane, 2006). This proposed interpretations or uses argument is called an “interpretation/use argument” (or “IUA”) where the IUA covers all claims are the basis of the test scores (Kane, 2013, p.2).

Validity entails clarifying and justifying the “interpretations or uses of test scores” (Kane, 2001, p.328). By using an argument-based approach, this proposed interpretations or uses argument facilitates the process of validation. The development of the unified model of construct validation in the 1990s did not provide a specific function to operate validation (Kane, 2013); nevertheless, a more simple and straightforward argument-based approach helps validation (Bachman, 2005; Chapelle et al., 2008b, 2010) – the claims and inferences are stated in a proposed interpretations/uses argument, and then these claims and inferences are to be assessed in the argument (Kane, 2012, 2013).

As mentioned earlier, validity is considered as a matter of degree (Henning, 1987; Messick, 1989; Weir, 2005). It is possible that the validity varies over time as more evidence supporting the warrant in the interpretations/uses arguments grows (Kane, 2013). The more backing the argument has, the higher the validity of the argument; in contrast, validity may decrease when there are more rebuttals found on the proposed IUA. Therefore, in the process of validation, interpretations and uses of test scores adequately supported by appropriate and reasonable evidence are considered to have high validity; in other cases, the IUA is said to have low validity. Kane (2012) sets up three criteria for “evaluating interpretive arguments” (p.13). First, the argument should be clear – the claims should be stated in detail so as to make the justification for the proposed interpretations and uses clear. Second, the argument should be coherent – the coherence for argument is expected so that the reasoning guiding from the observed performances to conclusions and decisions becomes more convincing. Third, the inferences and assumptions should be plausible: the reasonable assumptions should be related to the stakeholders for the required empirical evidence. Kane (2012) comments on this validation process by using the argument-based approach as “simple in principle; but difficult in practice” (p.15). However, Kane emphasises that the interpretive argument performs three crucial functions: it provides assistance to develop the assessment on the condition that critical assumptions are fulfilled; it offers a solid structure for validation: the evidence is required for evaluating the specific inferences and assumptions in the interpretive argument; and it establishes a foundation for evaluating the validity of the proposed interpretations and uses as long as the criteria for the interpretive arguments as mentioned are satisfied.

Kane, Crooks and Cohen (1999) demonstrated an interpretive argument that measures performance. The interpretive argument consists of three inferential bridges: “scoring, generalization, and extrapolation” (p.15). *Scoring* inference links from the observed performance to an observed score. *Generalisation* inference links from the observed score based on actual performance to the claim about expected score. Finally, *extrapolation* inference involves relationships between the test performance and the performance in target domain. However, Chapelle et al. (2008b) argue that Kane’s “three-bridge framework” is insufficient to link to a theoretical construct because it does not include test use, which is a critical part in a language test, especially one with high-stakes results (p.12). They reformulated and expanded the three bridges with three additional inferences: *explanation* inference, which

links between observed test performance and a construct; *utilisation*, which links the target score to the decisions about the test-takers for which the score is used; and *domain definition*, which connects between the target language use domain and the observed performance (Chapelle, et al., 2008b, p.10).

2.6 Empirical studies on validation

This section reviews empirical research studies on validation in language testing. According to AERA/APA/NCME 1999 Standards, empirical validation of a test is the process of gathering evidence to provide “a sound scientific basis for interpreting the scores as proposed by the test developer and/or the test user” (p. 9). Empirical validation depends on “empirical and statistical evidence”, which is gathered by checking if test-takers’ marks on the test are similar to their marks on other appropriate measures of their ability (Alderson et al., 1995, p.171). Messick (1996) claims that validation is empirical evaluation of the meaning and consequences of measurement and that empirical evaluation is “the validation process is scientific as well as rhetorical and requires both evidence and argument” (p.254).

Most of the Chinese studies regarding the University Entrance Exam focus mainly on washback – an influence that a testing process has on teaching and learning within the classroom (Alderson & Wall, 1993). For example, Qi (2004, 2005) reports on studies of the intended washback effects of a high-stakes test – the National Matriculation English Test (NMET) – in Guangdong Province of China. Qi (2005) stresses that the primary selective function of the University Entrance Exam has imposed a great number of limitations on the test design and results in the failure of intended washback. Instead, negative washback occurs – teaching the content and format of the exam in class. Further, Qi reports on washback studies on the writing and proofreading tasks in the writing section of the NMET in 2007 and 2010 respectively. Similarly, Xiao, Sharpling, and Liu (2011) focus on the washback of the NMET on students’ learning in the Chinese secondary school context in the Hubei Province of China.

In China, there are two other high-stakes exams similar to the University Entrance Exam: the College English Test (CET) and the Test for English Majors (TEM). The aim of the CET is to examine undergraduate students’ English proficiency (non-English majors) in China, whereas that of the TEM is to measure the English proficiency of Chinese university undergraduates majoring in English language. Yang and Weir (1998) led a Sino-British joint research group carrying out a project called ‘The CET Validation Study’ in mid-1990s. The group conducted a

comprehensive study on validity evidence of the CET and found that strict quality control measures, objectivity of scoring, and detail and clear administrative procedures have been achieved. Yang (2000) also reported that the CET committee studied and set up detailed test specifications in order to accomplish content validity. Regarding the TEM, Jin and Fan (2011) reviewed the test by focusing on the test qualities, especially validity and reliability. In the review, they reported that the TEM Test Centre conducted a validation study in the mid-1990s and concluded that the tests were “reasonably reliable and valid tests” as the content tested were defined in the test specifications (p.593). Zou, Zhang, and Zhou (2002) also examined the response validity of the TEM tests and investigated how test-takers responded to the questions items. They found out that test-takers have used the reading strategies as specified in the test specifications.

It is interesting to review some international studies, because of the limited research on test validation in China. Ito (2005) conducted a validation study on the English language test in a Japanese Nationwide University Entrance Examination. He investigated the construct validity of the test by looking at the correlations between each pair of subtests, each subtest and the whole test, and each subtest and the whole test minus the subtest itself. The results revealed very low correlations, especially between the pronunciation test and the other subtests, indicating the low construct validity of the test. In another similar example validating an English language test, Dörnyei and Katona (1992) validated the C-test as a test of English language proficiency amongst Hungarian EFL learners. One hundred and two university English major students were invited to take four different language tests (Department Proficiency Test, TOEIC, Oral interview, and Cloze test) to form a General Language Proficiency measure against which the C-test was evaluated. The result was produced by factor-analysing the four different language tests by the Statistical Package for the Social Sciences (SPSS). The results of the study confirmed that the C-test was a reliable and valid instrument.

In Fulcher’s (1997) study validating an English language placement test in one of the universities in the United States, he examined the test in different aspects: construct validity by using correlation and a principle components analysis; concurrent validity by comparing and correlating the students’ TOEFL test scores with the placement test results; and content validity by interviewing the university faculty members as to whether test items in the placement test are testing students’ language ability. Fulcher’s methodology was based on

Wall, Clapham and Alderson (1994), which evaluated and validated an institutional placement test at Lancaster University. Wall, Clapham and Alderson's study was comparatively more comprehensive in terms of the test content and the sample size was bigger. The results showed that face and content validity, by comparing the test content and teaching course material, were reasonably high; construct validity, by correlating between the subtests and the total score, were medium; concurrent validity, by associating each subtest score with other assessments in class, was satisfactory; and the reliability was generally satisfactory.

As commented by Kane (2012), the uniform concept based on construct validity is not easy to instigate successfully because it does not have direction on how to proceed with the validation process. However, an argument-based approach to validation can tackle these drawbacks by building an interpretation and uses argument structure for validation (Cronbach, 1988). Kane (2002) validates high-stakes test programmes including high school graduation tests in the United States. The interpretations extended from "the scores to conclusions about achievement on the Test Standards, and to conclusions about achievement on the State Standards, and then to conclusions about achievement in high school"; eventually the high-school diplomas are awarded after evaluating the test scores (Kane, 2002, p.40). In Chapelle Enright and Jamieson's (2008a) study, in building a validity argument for the Test of English as a Foreign Language (TOEFL), they demonstrate the validation of TOEFL score interpretation as an indicator of academic English language proficiency and score use for admissions decisions at English-medium universities. Wang, Choi, Schmidgall, and Bachman (2012) reviewed the Pearson Test of English Academic (PTEA) by building an assessment use for making admission decisions at tertiary level institutions and organisations where English is used for communication. Such institutions and organisations gathered adequate evidence from the review documents and then analysed the evidence in order to support the relevant claims and warrants in the interpretations arguments.

There are very few studies on validity and test validation about the English Language Exam for university entry made public in China (Qi, 2010). Cheng and Qi (2006) reviewed the National Matriculation English Test (NMET) and looked at the content of the questions generally, but did not go into detail on what language ability each question was actually testing. Zeng (2010) comments on the construct validation of the Computerized Oral English Test of the NMET in

Guangdong Province; however, the oral test is designed only for students who are planning to study English language as their major at university.

2.7 Chapter summary

This chapter has reviewed literature concerned with five areas of critical importance to the research presented in this thesis. First, literature addressing education and examinations in China was described. Second, literature on newly implemented English Curriculum Standards and education reforms in Shanghai was introduced. Third, this section examined the literature on different models of language proficiency and two major models of language ability. Fourth, literature on the most important issue of all in language testing – validity – was evaluated. Finally, empirical studies that explored validity and test validation were reviewed. Yet it should be pointed out that, very little empirical research on validity and test validation regarding the English language examination for university entry has been made public in China. Until now, no researcher has undertaken a detailed investigation of validating the English language examination for university entry in Shanghai. This study is an attempt to fill these gaps by looking at the examination papers and different empirical studies in Shanghai. In the next chapter, two key research questions are presented, and the methodology and research design most suitable for conducting this research is introduced.

Chapter 3: Methodology

3.1 Research questions and methodological framework

The study was guided by the following research questions:

1. Does the English Language Exam for university entry in Shanghai test the language abilities specified in the Curriculum and Bachman and Palmer's model?
 - 1.1. What is tested in the UEE?
 - 1.2. How closely does the UEE content match the Curriculum content?
 - 1.3. How closely does the UEE content match Bachman and Palmer's model of language ability?
2. What evidence is there for the following test qualities (Bachman & Palmer, 1996)?
 - 2.1. Validity
 - 2.2. Reliability
 - 2.3. Impact
 - 2.4. Interactiveness
 - 2.5. Practicality
 - 2.6. Authenticity

The study comprised two phases: Phase 1 – a qualitative, content analysis to answer research question 1; and Phase 2 – an empirical study to answer research question 2.

In Phase 1 – a qualitative content analysis of the past 10 years' UEE papers (2005-2014) was conducted. By investigating all UEE questions and test tasks, the exact nature of what the UEE was testing was clearly identified. The UEE papers were then compared to ascertain how closely the UEE content matched the Curriculum content and Bachman and Palmer's model of language ability. This phase examined the content validity and construct validity of the UEE in comparison with the Curriculum and the model respectively.

In Phase 2 – empirical studies were carried out in Shanghai. The studies included mock exams, class observations and interviews, and think-aloud tasks. In China, nobody can obtain University Entrance Exam result data from the Chinese Ministry of Education. Those results are never revealed to the public. The best way to get similar data is to invite some students in Shanghai to take past English Language Exam papers. The purpose of the mock exam activity was to obtain empirical evidence for the qualities of the test. Sixty first-year university students

in Shanghai were invited to take two past year UEE (2008 and 2011) papers and an IELTS test over three months.

The class observations were meant to reveal what teachers were actually teaching in the language classroom. This was used to obtain evidence of consequential validity, if any. A class for the teaching of one unit of a textbook was observed over two weeks, audio-recorded and real-time field notes were taken. Post-observation interviews were carried out – asking the teachers why they taught particular materials and asking the students what they did after class. Teaching materials including testing syllabi and textbooks were collected for analysis, which determined how serious the impact of the test was on teaching and learning.

A group of eight senior II students were also asked to verbalise how they answered each task. Through this process, the researcher could collect information about the way students tackled the activities. Each participant took the test separately and sat in front of the researcher. As Mackey and Gass (2005) recommend, participants were asked what was going through their minds while they were taking the test. The process of taking the tests was audio-recorded. From these tasks, response validity of the UEE was determined. Table 4 shows the two phases and how the data was elicited and the research questions answered.

Table 4: Data Elicitation to the Research Questions

Phase	Research Questions	Data Elicitation
1	1. Does the English Language Exam for university entry in Shanghai test the language abilities specified in the Curriculum and Bachman and Palmer's model?	
	1.1 What is tested in the UEE?	By analysing the UEE questions and test tasks.
	1.2 How closely does the UEE content match the Curriculum content?	By comparing the UEE content and the Curriculum content.
	1.3 How closely does the UEE content match Bachman and Palmer's model of language ability?	By comparing the UEE content and Bachman and Palmer's model.
2	2. What evidence is there for the following test qualities?	
	2.1 Validity ➤ Construct validity	By correlation analysis, correlating each subtest (listening comprehension, vocabulary and grammar, cloze, reading comprehension) with other subtests, and each subtest with total test; By Rasch analysis, checking if the UEE papers measured the appropriate construct;

	➤ Concurrent validity	By correlating students' past UEE papers scores with their scores on other tests, such as IELTS;
	2.2 Reliability ➤ Internal consistency ➤ Stability and parallel-form reliability	By using Spearman-Brown, Cronbach alphas, and Kuder-Richardson formula 21 reliability coefficients, determined the internal consistency of the UEE; By correlating students' scores on past UEE papers.
	2.3 Impact / consequential validity	By observing classes and interviewing teachers and students after class.
	2.4 Interactiveness / response validity	By think-aloud tasks and subsequent interviews with the students.
	2.5 Practicality	By interviewing examination designer.
	2.6 Authenticity	By interviewing with students and analysing the UEE questions and test tasks.

As Bryman (1992) has claimed, "The integration of quantitative and qualitative research is most frequently encountered in terms of triangulation" (p.63). Using multiple measures of investigation is a form of triangulation, and the use of multiple methods enables the investigation of a phenomenon from different perspectives (Mackey & Gass, 2005). In this research, different research methods were adopted to answer the research questions to ensure that the validating of the English language examination for university entrance in Shanghai would be comprehensive.

Triangulation involves the application of various sources and different data collection methods in the investigation of research questions (Denscombe, 2007; Patton, 1990). In this research, interviews with the researcher were an opportunity to elaborate on patterns or questions or supplement the information missed resulting from classroom observations. Classroom observations were good for capturing what happened in class including the content the teacher taught and the students learned, their behaviour and beliefs in class. However, the classroom observations alone could not explain what exactly happened in the classroom. The subsequent interviews would help the researcher find out why the participants had such behaviour and beliefs in and outside the class.

3.2 Phase 1: Qualitative content analysis

3.2.1 Materials

Papers from the past 10 years (2005-2014) of the English Language Exam for university entry in Shanghai were used for analysis. There were two papers in the UEE. In Paper I, there were three parts: Listening Comprehension, Grammar and Vocabulary, and Reading Comprehension, whereas there were two parts in Paper II: Translation and Guided Writing. The total scores of the examination paper were 150 and time allowed for the exam was two hours.

3.2.2 Stage one: English Curriculum Standards

The examination designers had basic guidelines when they set up the English language examination papers. They were the English Curriculum Standards and Examination Specifications. Both documents were originally written in Chinese, and translated into English by the researcher. English Curriculum Standards for High School was established by the Ministry of Education of the People's Republic of China. The overall objective of the English curriculum in basic education was to promote students' general ability in language use, which was based on the comprehensive development of their language skills, language knowledge, affect and attitude, learning strategies, and cultural understanding. The Examination Specifications, which were formed by a group of professional under the Examination Centre based on the English Curriculum Standards, included the nature of the UEE, objectives of the UEE measurement, UEE rules, the contents and requirements of the UEE, and question examples.

Content analysis is considered as "a systematic reading of a body of texts, images, and symbolic matter" (Krippendorff, 2013, p.10). Content analysis also allows researchers to enrich understanding of the data by examining theoretical issues. Through content analysis, it is possible to extract and place keywords into content-related categories (Elo & Kyngäs, 2008). In phase one of this research, items in each section in the examination papers were checked against the objectives for the Examination Specifications and the English Curriculum Standards to see if the objectives were fulfilled. The analysis was validated by asking a colleague of the researcher to check the items of different sections independently. Details of the research method will be introduced in Chapter 4.

3.2.3 Stage two: Bachman and Palmer's model of language ability

Bachman and Palmer's (1996) model defines language ability as involving two components: language knowledge and strategic competence. It focuses on the interactions between language ability, personal attributes, topical knowledge, and affective schemata, and showing how these areas interact with each other in language use circumstances. Language knowledge includes two broad categories: organisational knowledge (grammatical and textual knowledge) and pragmatic knowledge (functional and sociolinguistic knowledge). Strategic competence is believed to be a set of metacognitive components, or strategies. These components function in three areas: goal-setting, appraising, and planning.

The past 10 years of UEE papers were analysed against the Bachman and Palmer's model of language ability theory. Question items and test tasks in the UEE were checked and analysed as to whether they tested the language abilities specified in Bachman and Palmer's model. If they did, examples would be listed and explained in detail. Regarding the inter-coder checking, a colleague of the researcher was invited to check independently the items against the area of functional knowledge only, due to limited time and resources.

Table 5: Components of Language Ability – Analysis Checklist

Component of Language Ability	Comments
Grammatical knowledge <i>Vocabulary</i> <i>Syntax</i> <i>Phonology and graphology</i>	
Textual knowledge <i>Cohesion</i> <i>Rhetorical or conversational organisation</i>	
Functional knowledge <i>Ideational functions</i> <i>Manipulative functions</i> <i>Heuristic functions</i> <i>Imaginative functions</i>	<i>Required</i> <i>Or</i> <i>Marginally required</i> <i>Or</i>
Sociolinguistic knowledge <i>Dialect and varieties</i> <i>Registers</i> <i>Natural or idiomatic expressions</i> <i>Cultural references and figures of speech</i> <i>Genres</i>	<i>Not Required</i>
Strategic competence <i>Goal-setting</i> <i>Appraising</i> <i>Planning</i>	

A checklist was formed by using the model of language ability (Bachman & Palmer, 1996, p.77). The checklist (See Table 5) was very useful for determining what the UEE was testing, that is, the construct. It helped measure the degree to which components of language ability were involved in the UEE. Based on the analysis of the prompt of the question items and test tasks, comments were made in the checklist. If the components were relevant, 'required' was filled in the list, whereas 'marginally required' or 'not required' was noted if those were irrelevant. Details of the research method will be introduced in Chapter 5.

3.3 Phase 2: Empirical studies

3.3.1 Mock exams

A total of 60 first-year university students were invited to participate in this study. These 60 students were native Chinese speakers and English majors. They took the University Entrance Exam in June 2013, nine months before this study took place. They were chosen because they were familiar with the examination format. Senior high school students were not so acceptable since senior III students were busy preparing for the examination, and senior I and II students had limited vocabulary to cope with past examination papers. A large number of participants is recommended to enhance the stability of the data distribution (Phakiti, 2010), although Hatch and Lazaraton (1991) suggest that a minimum of 30 participants is enough for generating a distribution of a range of scores for quantitative data analysis.

The tests were conducted in 90 minutes on three different afternoons over three months. These 60 first-year university students from the same university in Shanghai were invited to take UEE papers from the past two years (2008 and 2011) and an IELTS sample test. There were two sections – listening and reading. For the UEE, the listening section of 30 minutes included 10 short conversations, two passages and two long conversations, whereas the reading section of 60 minutes included grammar and vocabulary, cloze and reading comprehension. For the IELTS sample test, four long conversations or lecture seminars made up a 30-minute listening section, whereas a 60-minute reading section had three reading passages. At the end of all three tests, two participants were randomly chosen to attend an interview. The details of the data collection will be introduced in Chapter 6.

The mock exam activity was analysed quantitatively. Construct validity of the UEE was determined by correlating each sub-test with other sub-tests and each sub-test with the total

test in order to examine both convergent and discriminant validity, whereas concurrent validity was studied by correlating the UEE with the IELTS test sample. In this study, an IELTS sample test was used when investigating the concurrent validity of the UEE because the IELTS is a standard international language test and the time taken for the listening and reading sections is roughly the same as those in the UEE. Regarding reliability, internal consistency of the test was determined by using split-half method whereas stability and parallel-form reliability was obtained by correlating students' scores on past UEE papers. Furthermore, descriptive statistical analysis and classical item analysis were applied to provide information about the test samples and the measure. Finally, Rasch analysis was used to reveal the test item difficulty and fit, and to see if the two tests measured the appropriate construct in terms of relevance, representativeness and technical quality (Baghaei & Amrahi, 2011).

3.3.2 Classroom observations and interviews

The participants in this study were a class of 40 Chinese senior II high school students and a Chinese teacher from a government high school in Shanghai City. The students were over 16 years old and learned English as a foreign language. The teacher was a native Chinese speaker and at the time the study was conducted had been teaching for eight years, having graduated from a university in Shanghai majoring in teaching English. The English lessons of the teacher for one teaching unit (about two weeks) were observed. Each lesson lasted 40 minutes and it was audio-recorded. Notes were jotted down and particular attention was paid to what special language the teacher used about the UEE in class. In addition, an interview was conducted with one of the UEE designers in which questions were asked about the UEE test development process and what the designer thought about the UEE.

Data collection from classroom observations was supplemented with face-to-face interviews with both teacher and students in Chinese, which were semi-structured and structured respectively. The semi-structured interviews with the teacher were conducted immediately after every lesson; the questions were related to what she taught and the purpose of her teaching in each lesson. The students' interviews were arranged for the lunch hour after class in school. The same questions were asked to five different students and covered their learning beliefs and behaviour related to the University Entrance Exam. The advantage of using interviews as a tool for data collection was to discover information from participants that the researcher could not directly uncover during class observations, for example, their thoughts

and feelings (Mackey & Gass, 2005). The details of the data collection for the class observations and interviews will be described in Chapter 7.

Data analysis methods used in this research were determined by the research questions. Classroom observations were analysed both quantitatively by looking at the time allocation in the teacher's teaching content in class and qualitatively by focusing on the teacher's teaching belief and behaviour when teaching. The interview data was analysed qualitatively in order to unveil what students' learning beliefs and behaviour were in view of the University Entrance Exam. The audio-recordings of classroom observations and participants' interviews were transcribed and translated into English. The transcripts were then coded, analysed, and summarised qualitatively (Wagner, 2010).

3.3.3 *Think-aloud tasks*

Mackey and Gass (2005) describe introspective methods as techniques that inspire participants to convey their "internal processing and perspectives about language learning experiences" so that information missed by observational methods can be accessed by researchers (p.201). In the think-aloud tasks, eight students were randomly selected from a senior high school in Central Shanghai City and invited to take one of the two past year UEE (2008 and 2011) papers. They were in their senior II and all native Chinese speakers. These eight students were divided into four groups – two groups worked on the 2008 UEE and two on the 2011 Exam. Thinking-aloud protocols were used to capture participants' internal thinking processes of test tasks of the examination paper. It revealed if the test item or task engaged the test-taker's language ability the way in which it was designed (Nunan, 1992). These verbal reporting methods, as described by McDonough and McDonough (1997), were designed to help researchers reveal a comprehensive picture of the thinking process.

Two participants from each group took part in the task on Saturday morning in school: one participant worked on the grammar and vocabulary, cloze and reading comprehension sections whereas the other one worked on the listening and writing sections. The tasks were recorded on audio mp3. During recording, the researcher made observations including the participants' facial expressions and the general environment. Short interviews were conducted after they finished their tasks in school on the same day. Finally, all their verbal responses and interviews were transcribed and translated by the researcher. The procedures and recording details of the data collection for the think-aloud tasks will be described in detail in Chapter 8.

Kasper (1998) stresses that introspective data reveals what is going through a participant's mind when they are taking a task or question item, and the thinking processes need to be analysed and speculated on from the verbal response data; however, they cannot be considered as the "ultimate revelations" about the thinking processes (Dörnyei, 2007, p.150). Introspective data was analysed qualitatively so as to uncover whether the UEE question items or tasks engaged the test-taker's language ability. When analysing the data, the Examination Specifications were compared with their responses in the think-aloud tasks. Also the post-task interviews were then transcribed and translated in English, and the transcripts were coded, analysed and summarised qualitatively.

3.4 Ethical considerations

Research regarding people's lives must be mindful of ethical matters in the social world (Dörnyei, 2007). In accordance with the ethical guidelines issued by the Human Ethics Committee of Victoria University of Wellington, privacy and confidentiality are respected through the entire research process. This research formed the basis of a Victoria University of Wellington PhD studentship and has been reviewed and approved by the Human Ethics Committee of Victoria University of Wellington (Appendix A). No names or identifiable information has been used in any report (Gass, 2010).

The empirical studies included three groups of participants: a class of 40 senior high school students, the English teacher, and the UEE designer for class observation and interviews; a group of 60 first-year university students for the mock exams activities; and a group of eight senior high school students for the think-aloud tasks. The participants were gathered one day before the studies began, and the objectives of the research and the nature of the study were clearly explained and outlined to them. Requests were made for them to participate in the studies on a voluntary basis. Requests for class observation and individual interviews to be audio-recorded were also made at this point. The English teacher was informed that there would be no employment implications and the class observation and interviews were not related to her school teaching; more importantly, participation in this study (class observations and interviews) would not influence her teaching at school. On the other hand, the students understood that this research was not related to their school studies; participation in this study (the mock exams, think-aloud tasks, and interviews) would not influence their grades at school. They were provided with copies of Participant Information Sheets and Consent Forms

(Appendix B) on the same day, and were encouraged to think about their willingness to take part in the study.

Signed consent forms were collected from all participants prior to the commencement of the studies. In order to protect the anonymity of the participants and the school, the students in the mock exams were assigned a number whereas the participants in classroom observation and interviews, and think-aloud tasks were reported using pseudonyms. Information collected through the interviews was kept confidential. They were assured that no identifying information would be included in the study. Furthermore, recorded data was kept in a locked file on the researcher's computer and paper data was kept in a locker in their office. Only the researcher and their supervisors had the access to the data. The data will be kept for about two years after the completion of the PhD programme and will then be destroyed. Participation in this research was entirely voluntary and participants were informed of their right to withdraw from the study, including any information that they have provided, any time up until 1 March 2014 without giving a reason.

3.5 Chapter summary

This chapter has outlined the research design and described the research framework and procedures. The content of the research method for each study will be illustrated in detail in the following chapters. No researcher to date has undertaken a detailed investigation of the validation of the English language examination for university entry in Shanghai. A multi-method approach to data collection in two different phases was applied to ensure data was investigated from different perspectives, enhancing a comprehensive analysis of the research questions and fabricating validity and reliability. In phase one, a qualitative content analysis by analysing the past English language papers delivered information of content and construct validation of the University Entrance Exam. Empirical studies in the second phase – mock exams, classroom observations and interviews, and think-aloud tasks – provided comprehensive field data collection in Shanghai City. The schedule for research data collection and analysis is listed in Table 6. Last but not the least, the research process ensured the privacy and confidentiality of participants in this study in accordance with research ethics.

Table 6: Schedule for Research Data Collection and Analysis

Phase	Studies	Timeframe
1a	Compared and analysed the past 10 years (2005-2014) English language examination papers content and the Curriculum content.	June 2013 – August 2013
1b	Compared and analysed the past 10 years (2005-2014) English language examination papers content and Bachman and Palmer's model of language ability.	Sept 2013 – November 2013
2	Designed Phase 2 schedule; selected and prepared research site and participants.	December 2013 – February 2014
2a	Class observations and interviews	March 2014
2b	Mock exams: First mock exam (UEE 2008) Second mock exam (UEE 2011) Third mock exam (IELTS sample test)	(Tuesdays) 25 March 2014 22 April 2014 20 May 2014
2c	Think-aloud tasks: UEE 2008 – listening and writing sections UEE 2008 – grammar & vocabulary and reading sections UEE 2011 – listening and writing sections UEE 2011 – grammar & vocabulary and reading sections	(Saturdays) 19 April 2014 26 April 2014 10 May 2014 17 May 2014
2d	Data transcription and translation, and data analysis.	June 2014 – November 2014

Chapter 4: Content validity: English Curriculum Standards

4.1 Introduction

This chapter aims to answer the first research question, “Does the English Language Exam for university entry in Shanghai test the language abilities specified in the English Curriculum Standards?” It thus identifies what was tested in the UEE and how closely the UEE content matched the language ability specified in the English Curriculum Standards. A qualitative, content analysis of the past 10 years’ UEE papers is conducted and the research method for the analysis is introduced. By investigating all question-items and test tasks, the exact character of what the UEE tests (or has tested) is clearly identified. The UEE papers are then compared to ascertain how closely the UEE content matches language ability of the English Curriculum Standards. Finally, the findings and discussion are presented at the end of this chapter.

4.2 Research method

4.2.1 Materials

Papers from the past 10 years’ (2005-2014) English Language Exam for university entry in Shanghai were used for the analysis. There were two papers in the UEE (See Table 7). In Paper I, there were three parts: Listening Comprehension, Grammar and Vocabulary, and Reading Comprehension. In Paper II, there were two parts: Translation and Guided Writing. The total score of the UEE paper was 150 and time allowed for the exam was two hours.

Table 7: The Examination Format of the UEE English Language Exam

Paper I		
Part I	Listening Comprehension	<i>Section A – short conversations Section B – passages Section C – long conversations</i>
Part II	Grammar and Vocabulary	<i>Section A – grammar Section B – vocabulary</i>
Part III	Reading Comprehension	<i>Section A – cloze Section B – passages (multiple-choice) Section C – passage (short-answer)</i>
Paper II		
Part I	Translation	<i>Translate 5 Chinese sentences into English</i>
Part II	Guided Writing	<i>Write a composition of 120-150 words</i>

4.2.2 English Curriculum Standards

The English Curriculum Standards (ECS) for High School (Trial Version) was instituted by the Ministry of Education of the People's Republic of China (31 March 2003). The ECS was originally written in Chinese, and was translated into English by the researcher. The overall objective of the English curriculum in basic education was to promote students' general ability in language use, which was based on the comprehensive development of their language skills, language knowledge, affect and attitude, learning strategies, and cultural understanding (MOE, 2003). Among them, language skills and language knowledge were the basis, while affect and attitude was a critical factor, learning strategies allowed students to improve their efficiency and initiative, and a well-developed cultural understanding ensured appropriate use of language (See Figure 6).

Figure 6: Framework of the objectives of the ECS (Source: Wang, 2007, p.97)

The various levels of the curriculum objectives in basic education should all demonstrate the comprehensive development of students' language skills, language knowledge, affect and attitude, learning strategies, and cultural understanding. Table 8 provides a general description of the curriculum objectives for Level 8 (university entry level).

Table 8: Curriculum Objectives for Level 8

Level	A General Description
8	<ol style="list-style-type: none"> 1. Have confidence and initiative; 2. Communicate naturally with English-speaking people on familiar topics; make comments on oral or written materials; 3. Compose coherent and completed short essays; 4. Arrange, organise, and carry out language practices independently, such as plan making and result reporting; 5. Make effective use of a variety of educational resources to obtain and process information; 6. Develop productive learning strategies through self-assessments of the progress; 7. Be aware of and respect the cultural differences and backgrounds.

4.2.2.1 Language skills

Language skills play an important role in students' language communicative ability. The assessment of language skill tests the ability not only of the four sections, viz., listening, speaking, reading, and writing, but also of the whole (MOE, 2003). On the one hand, the four sections complement each other in content, as listening and reading mean understanding while speaking and writing measure expression. On the other hand, the four sectional and comprehensive practices should be sufficiently performed in order to help students gain a general ability in language use so as to prepare for real world situations. The objective for language skills here is mainly to focus on 'what students are able to do' at different levels (See Table 9).

Table 9: Objectives for Language Skills at Level 8

Skills	Description of Objectives
Listening	<ol style="list-style-type: none"> 1. Be able to recognise different attitudes through different tones; 2. Be able to understand discussions or conversations on familiar topics, and remember the main points; 3. Be able to grasp the viewpoints in simple passages; 4. Be able to figure out the themes or main ideas in English language broadcasting and television news; 5. Be able to understand euphemistic/implied meanings of suggestions or advice.
Reading	<ol style="list-style-type: none"> 1. Be able to understand different opinions and attitudes in reading materials; 2. Be able to recognise the features of different types of writing; 3. Be able to understand long sentences and difficult sentences by analysing sentence structures; 4. Be able to enjoy basic literature with the help of teachers; 5. Be able to get and then process the information required by learning tasks from e-books or the Internet;

	6. Should do as much reading as over 360,000 words outside textbooks.
Writing	<ol style="list-style-type: none"> 1. Be able to write coherent and structurally complete compositions; narrating events or expressing your own opinions and attitudes; 2. Be able to paraphrase or summarise articles after reading; 3. Be able to write in appropriate styles with natural and grammatical sentences; 4. Be able to write short essays or reports based on textual and graphical information provided.
Speaking	<ol style="list-style-type: none"> 1. Be able to use proper intonations and rhythms; 2. Be able to discuss and then make plans in accordance with learning tasks; 3. Be able to report the processes and results of experiments or research; 4. Be able to make 3-minute prepared speeches on common topics; 5. Be able to make effective uses of the English language to judge, criticise, complain, or express opinions in daily life communications; 6. Be able to do translations in such daily life activities as shopping and sightseeing.

4.2.2.2 Language knowledge

Language knowledge in education, an important part of language abilities and the foundation of language skills, consists of phonetics, vocabulary, grammar, functions, and topics (MOE, 2003; see Table 10).

Table 10: Objectives for Language Knowledge at Level 8

Knowledge	Description of Objectives
Phonetics	<ol style="list-style-type: none"> 1. Gradually acquire natural, appropriate, and fluent pronunciation and intonation through communication; 2. Be able to recognise and express implicit intents and attitudes with pronunciations and intonations; 3. Have some knowledge of the rhythms in poems; 4. Be able to recognise and spell unfamiliar words or simple sentences after hearing them pronounced.
Vocabulary	<ol style="list-style-type: none"> 1. Be able to understand and express different functions, attempts, and attitudes with words and phrases; 2. Use words and phrases to describe complicated objects, behaviour, features, and concepts; 3. Be able to use 3,000 words and 400-500 phrases or collocations.
Grammar	<ol style="list-style-type: none"> 1. Further grasp the expressions that are used to describe time, place, and direction; 2. Further understand and grasp expressions that are used to compare people, objects, and events; 3. Use proper language forms to describe or express opinions, attitudes, and emotions; 4. Learn and grasp basic discourse knowledge, and be able to effectively organise the information for certain purposes.

Functions	<ol style="list-style-type: none"> 1. Be able to understand and express communicative functions such as greetings, farewells, thanks, and introductions in larger contexts; 2. Be able to effectively and appropriately use the English language to judge, criticise, complain, and express opinions in communication in daily life; 3. Make flexible use of learned commonly used functions, and learn and grasp the new ones; 4. Make experienced use of the English language to communicate in real life situations.
Topics	<ol style="list-style-type: none"> 1. Become familiar with the topics on individuals, families, social interactions, and so forth; 2. Have further knowledge of the topics in daily life, hobbies and interests, manners and customs, science and cultures, and so forth; 3. Be knowledgeable about Chinese social and life topics such as professions, festivals, customs, and social etiquettes; 4. Have some knowledge of the topics on English-speaking countries' daily lives and habits.

4.2.2.3 Affect and attitude

The concepts of affect and attitude refer to the elements such as interest, motivation, confidence, perseverance, and team spirit, all of which will influence the learning process and affect the students (MOE, 2003). They include the students' national and international awareness, which will be cultivated gradually during the process of learning. The general target of English learning is to strengthen the student's national awareness and widen their international horizon (See Table 11).

Table 11: Objectives for Affect and Attitude at Level 8

Description of Objectives
<ol style="list-style-type: none"> 1. Maintain the interest and motivation to learn English and get involved in activities which can help to improve active use of English; 2. Have the appropriate English learning motivation and understand clearly that English learning is for communication and expression; 3. Keep confident in English learning and have the courage to communicate with others in English; 4. Have the ability to overcome all difficulties in English learning and willing to ask others for help; 5. Be able to understand and respect the feelings of others when communicating with them; 6. Have a highly cooperative spirit and willing to share various kinds of learning resources with others; 7. Be able to introduce China's national culture to others in English; 8. Be able to understand and respect foreign cultures and demonstrate an international cooperative spirit in your actions.

4.2.2.4 Learning strategies

Learning strategies refer to various kinds of actions students take to study and improve effectively (MOE, 2003). The strategies in English learning include cognitive strategies, metacognitive strategies, communicative strategies, and resourcing strategies. Teachers are expected to apply those strategies into the teaching process and help their students establish their own strategies. Cognitive strategies refer to the methods and steps students use to finish the learning tasks. Metacognitive strategies refer to students' plans, practices, reflections, evaluation and adjustment of learning. Communicative strategies refer to the strategies students use to find more opportunities to communicate, to maintain communication, and to improve their performance in communication. Resourcing strategies refer to the student's correct and effective use of multi-media devices to learn and practise English. Teachers ideally help their students cultivate the ability to adjust their strategies. In the learning process, helping students to use learning strategies effectively is not only to help them stay on course and improve the effectiveness of study, but also to help them establish their autonomous learning (See Table 12).

Table 12: Objectives for Learning Strategies at Level 8

Types of Strategy	Description of Objectives
Cognitive Strategy	<ol style="list-style-type: none">1. Establish relations amongst different kind of relevant knowledge with using your imagination;2. Analyse and solve problems with the help of logic by reasoning and summarising;3. Be good at summarising general linguistic principles from texts and be able to apply them to your own situations;4. Be capable of focusing on key points, writing down notes and summarising knowledge in the process of learning;5. Be able to guess or speculate on the meanings of new words from the context or in specific situations, during the process of reading and listening;6. Use diagrams and other non-verbal information to help express yourself or understand an idea as you are learning;
Metacognitive Strategy	<ol style="list-style-type: none">1. Make an English learning plan suited to your own purpose/need;2. Look for more ways to learn English actively;3. Be good at creating or grasping opportunities to learn English;4. Know how to get help when facing difficulties in learning;5. Communicate with teachers and/or classmates on your learning experience;6. Evaluate learning performance, summarise effective learning methods and improve the memorising effect according to the rules of memory.

Communicative Strategy	<ol style="list-style-type: none"> 1. Communicate actively with classmates in English both in curricular and extra-curricular activities; 2. Be good at using gestures, facial expressions and other non-verbal communications to improve the effect of communication; 3. Be able to overcome the linguistic difficulties and maintain the communication; 4. Be capable of grasping every opportunity to communicate with others in English in daily life; 5. Observe and be aware of basic etiquette in English communication.
Resourcing Strategy	<ol style="list-style-type: none"> 1. Acquire more information about English with the help of library resources, the internet, radio, television, etc.

4.2.2.5 Cultural understanding

Culture refers to the history, geography, customs and practices, traditions, lifestyle, accepted conduct, and values of English-speaking countries (MOE, 2003). Experiencing and understanding those cultures is good for the student's English learning and practice. Students should be taught to have a general idea about the culture of the English-speaking countries and the difference between them and Chinese culture. Table 13 shows the detailed targets of cultural understanding at level 8.

Table 13: Objectives for Cultural Understanding at Level 8

Description of Objectives
<ol style="list-style-type: none"> 1. Understand common idioms, proverbs and their hidden meanings in English; 2. Understand legends and allusions which are commonly used in communication; 3. Know the names of the major writers, artists, scientists and their experiences, achievements and contributions; 4. Have a general idea about the political and economic circumstances of English-speaking countries; 5. Know something about the major mass media in English-speaking countries; 6. Know about differences in lifestyles between China and English-speaking countries; 7. Know about the differences of behaviour and manners between Chinese and foreigners; 8. Know about religious traditions in English-speaking countries; 9. Understand various cultures and cultivate the awareness of the international society by learning English; 10. Deepen your understanding of Chinese culture by comparing Chinese culture with foreign cultures.

4.2.3 Examination Specifications

The Examination Specifications are based on the English Curriculum Standards. The objectives of the Examination Specifications were designed by a group of professionals under the Shanghai Examination Centre for each of the different sections listed on Table 14: listening comprehension, grammar and vocabulary, reading comprehension, writing and speaking (SMEEC, 2014b, pp.65-66). These Specifications included the nature of the UEE, objectives of the UEE measurement, UEE rules, the contents and requirements of the UEE, and question examples.

Table 14: Objectives of English Language for Examination Specifications

Section	Description of Objectives
Listening Comprehension	<ol style="list-style-type: none">1. Can acquire factual information from conversations;2. Can make simple judgement(s) from factual information;3. Can understand hidden or implied meaning(s) of discourse;4. Can summarise the main idea of a discourse.
Grammar and Vocabulary	<ol style="list-style-type: none">1. Can identify and understand different expressions of different tones and intonations correctly in a specific context;2. Can comprehend and use vocabulary correctly in a specific context;3. Can identify, comprehend and use language knowledge correctly in a specific context;4. Can comprehend and apply language functions correctly in a specific context.
Reading Comprehension	<ol style="list-style-type: none">1. Can understand the basic content of the article;2. Can understand words and sentences correctly in context;3. Can summarise the main idea of a paragraph or article;4. Can infer the implied meaning of the article.
Writing	<ol style="list-style-type: none">1. Can translate sentences correctly and fluently by using language knowledge;2. Can write correctly, coherently, and aptly according to the meaning of question items or topics.
Speaking	<ol style="list-style-type: none">1. Can grasp basic language knowledge and skills and can pronounce and intone clearly and correctly the read aloud articles;2. Can use proper language functions and concepts for questioning and answering, requests and requirements, providing explanations and expressing opinions;3. Can express views and opinions on familiar topics based on personal experience;4. Can narrate and describe some common life events and phenomena, and make simple comments.

4.2.4 Data analysis

Content analysis is very common as a research method for “making replicable and valid inferences from data to the contexts of their use”, with the aim of providing knowledge, new insights, factual information and rational guidelines (Krippendorff, 2013, p.24). By using content analysis, it is possible to extract and categorise different data into content-related classifications (Elo & Kyngäs, 2008). In order to identify what was tested in the UEE, the test items or tasks for the past 10 years’ UEE papers were examined and investigated one by one against the objectives of the Examination Specifications, which were described in Section 4.2.3. The UEE items were checked against different topics or themes by using a form of content analysis. Subsequently, this data helped to generate and refine ideas, and identify and group it into different categories (Hyland, 2010). In this study, the analysis was validated by asking a colleague to independently check the items of the listening and reading comprehension sections against the Examination Specifications, and the inter-coder reliability checks reached 91% agreement with the coding of the researcher.

The test items or tasks for the past 10 years’ UEE papers were then examined and investigated one by one against the objectives of the English Curriculum Standards, which were described in Section 4.2.2. A checklist was used (see Table 15). The first column listed the five different categories in the ECS: language skills, language knowledge, affect and attitude, learning strategies, and cultural understanding. In the middle column, a tick (✓) or cross (x) was inserted depending on whether or not the objective was achieved. The final column listed which test tasks were involved and some examples if the category was achieved or listed reasons if it was not achieved. The aim was to ascertain how closely the UEE content matched the language ability required by the English Curriculum Standards.

Table 15: Checklist Sample

Objectives for different categories at Level 8 of Curriculum Standards:	Fulfilled or not	Test tasks and examples:
1. Objective one;	✓	Test task type and examples given;
2. Objective two.	x	Reason why the item was not achieved.

4.3 Findings and discussion: Part 1 Examination Specifications

4.3.1 *Listening Comprehension*

In this part, there were three sections:

- Section A – 10 short conversations with 10 multiple-choice questions. The 10 short conversations were about events in daily life in, for example, the office, a restaurant, and a shopping mall;
- Section B – two passages of six multiple-choice questions, with each passage containing about 250 words; the information was about news reports or narrative stories;
- Section C – two longer conversations of eight gap fillings; test-takers were required to fill in the tables from what they listened from the conversations.

There were 24 items in this part, which tested test-takers on whether they understood and obtained information from dialogues and passages in easy texts; they were required to conduct selection, judgement, speculation and generalisation from the information.

Below are some of the examples in Listening Comprehension from the 2014 UEE. The question items tested different abilities compared with the objectives of listening comprehension in the Examination Specifications.

1. Acquiring factual information from conversations

2014:5 W: Would you please help me clean the back yard today?

M: Sure. I am picking Jane up from the railway station at 3, but I will help you as soon as I get back.

Q: What will the man do first?

Answer: A. Catch the train; B. Meet Jane; C. Get some stationery; D. Clean the backyard.

From the man's conversation, the first thing he would do was to pick Jane up from the railway station. The answer can be deduced very directly.

2. Making a simple judgement from factual information

2014:1 W: What can I do for you, sir?

M: I want to report a theft. My briefcase was stolen.

Q: Who most probably is the woman?

Answer: A. A policewoman; B. A judge; C. A reporter; D. A waitress.

The man wanted to report a case; it can be judged that the occupation of the woman was a policewoman when the man talked to her.

3. Inferring implied meaning(s) of discourse

2014:8 M: You are not planning to transfer to a different university next year, are you?

W: If I were, you'd be the first one to know.

Q: What does the woman imply?

Answer: A. She doesn't plan to continue studying next year.

B. She has already told the man about her plan.

C. She isn't planning to leave her university.

D. She recently visited a different university.

If the woman had such a plan, the man would be the first one to know. It is implied that she was not planning to leave her university.

4. Summarise the main idea of a discourse

2014:13 Question for the passage: What is the passage mainly about?

Answer: A. A skiing;

B. A special community;

C. A splendid mountain;

D. A successful businesswoman.

From the whole passage, it could be concluded that Paradise Ridge was a very special residential community. This item asked test-takers to summarise the passage's main idea.

Table 16: Listening Comprehension Items versus Objectives of Examination Specifications

Listening Comprehension	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Average (items)	By %
1. Acquiring factual information;	12	10	11	13	11	10	10	9	11	10	10.7	45%
2. Making Judgement by comprehending details;	3	6	5	6	6	7	6	6	6	7	5.8	24%
3. Inferring/speculating on the implied meaning;	6	5	6	3	5	5	6	6	5	5	5.2	21%
4. Summarising the main idea.	3	3	2	2	2	2	2	3	2	2	2.3	10%

Table 16 shows the number of items in this part each year matches with the objectives set by the Examination Specifications. There were 24 items in this part. It was found that about 45 percent of the questions (about 11 items) required test-takers to gain basic information from the conversations or passages, and 24 percent (six items) and 20 percent (five items) of the

questions tested test-takers' judgement by comprehending details from factual information and inferring implied meaning from the conversations respectively. The remaining 10 percent of the questions asked test-takers to work out the main idea from the conversations or passages.

By analysing the questions in Listening Comprehension, the question items were found to match with the specified objectives in the Examination Specifications: can acquire basic information from conversations, can make simple judgments from factual information, can understand hidden meanings of discourse, and can summarise the main idea of a discourse. The primary use of language tests is to draw inferences about the language ability of test-takers. The more the exam items covered the Examination Specifications, the more evidence there was to determine whether or not the test score reflected the area of language ability to be measured.

4.3.2 Grammar and Vocabulary

From 2005 to 2007, there were 20 multiple-choice questions each year in this part – 16 sentences on grammar and four sentences on vocabulary. From 2008 to 2013, this part was divided into two sections: Section A – grammar, which comprised 16 multiple-choice questions; and Section B – vocabulary, in which a passage had nine blanks and test-takers were required to complete the passage by using the words provided. Since 2014, multiple-choice sentences in Section A have been changed to two short passages with 16 blanks – test-takers were required to fill in the blanks to make the passages coherent and grammatically correct; the vocabulary section was unchanged. It can be noticed that, for the past 10 years, the patterns of the test items were moved from the discrete items to more integrated skills. Test-takers were required to understand not only individual sentences but also the meaning of the context when answering the question items. The UEE began testing integrated language use more than it tested discrete item points.

Section A – Grammar took in the following grammatical items: nouns, pronouns, adjectives, adverbs, modifiers, tenses and voices, modal verbs, non-predicates, prepositions, conjunctions, question tags, subject agreement, noun clauses, attributive clauses, adverbial clauses, inversion, exclamation, comparison, multiples, modifiers and determiners. This section examined if test-takers were able to use the meaning according to syntactic functions,

and also comprehend and use different parts of speech in the functions and meanings of a sentence.

In Section B – vocabulary, from 2005 to 2007, four different sentences examined test-takers' understanding of vocabulary in the context of each sentence. There were four options for each question item, and the options each of the same part of speech, for instance adverb, adjective, noun, verb. From 2008 to 2014, vocabulary was tested separately by using a passage with blanks. Some words were provided in the box; they were in a variety of forms: noun, verb, adjective or adverb. The grammar and vocabulary sections from the 2014 UEE are presented here.

2014: Grammar and Vocabulary – Section A

Over time, the high cost of living became a little burden on my already (28) _____ (exhaust) shoulders. On the other hand, my search for a respectable job had not met with much success. As I had studied literature at university, I found it quite difficult to secure a suitable job in big companies. Mother had just said that (29) _____ I wanted to have a better career advancement, I had to find work in the city. Perhaps (30) _____ my mother had told me was deeply rooted in my mind. I did just as she had expected.

<u>Item number</u>	<u>Answer</u>	<u>Grammatical feature to be tested</u>
28	exhausted	non-predicate – past participle
29	if	adverbial clause of conditional
30	what	noun clause as a subject

The above paragraph tested test-takers on identifying and comprehending grammatical knowledge, such as non-predicates, adverbial clauses of time, and noun clauses, in the specific context in the passage.

2014: Grammar and Vocabulary – Section B

Let's say you've decided you want to eat more healthfully. However, you don't have time to carefully plan menus for meals or read food 41 at the supermarket. Since you really 42 yourself to a healthier lifestyle, a little help would come in handy, wouldn't it? This is where a "choice architect" can help 43 some of the burden of doing it all yourself. Choice architects are people who organize the contexts in which customers make decisions.

<u>Item number</u>	<u>Answer</u>	<u>Part of speech to be tested</u>
41	labels	noun
42	commit	verb
43	relieve	verb

The above paragraph tested test-takers' language knowledge on vocabulary and how to use the vocabulary correctly in a specific context with the correct parts of speech.

Table 17: Grammatical Features Tested in the Past 10 years' UEE Papers

Types / number of item	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
noun	1	1	1	3	3	2	3	4	2	3
verb	1	1	1	3	2	4	3	2	2	4
adjective	1	1	1	2	2	2	2	3	3	2
adverb	1	1	1	1	1	0	1	0	3	1
pronoun	1	1	1	1	1	1	1	1	1	1
preposition	1	1	1	1	1	1	1	1	1	1
modal verb	1	1	1	1	1	1	1	1	0	0
tenses / voices	2	2	2	2	2	2	2	2	2	2
question tag	1	0	0	0	1	0	1	0	0	0
subject agreement	1	0	0	0	0	0	0	0	1	0
non-predicate	4	5	3	4	5	5	4	4	5	4
nominal / noun clause	1	2	2	2	2	2	2	3	1	1
attributive clause	1	1	1	1	1	1	1	1	2	2
adverbial clause	1	2	2	2	1	2	2	2	2	2
inversion	1	0	1	1	1	0	0	1	1	1
exclamation	0	0	1	0	0	0	0	0	0	0
comparison / multiple	1	1	1	1	0	1	0	0	0	1
modifier / determiner	0	0	0	0	1	0	1	0	0	1

From Table 17, the past 10 years of UEE papers were analysed to show that different grammatical features and different parts of speech have been tested in the UEE. Those question items from the past 10 years of UEE papers fulfilled the UEE content requirements listed in the Shanghai Examination Handbook (SMEEC, 2014b). Generally speaking, the objectives were achieved for the grammar and vocabulary section. The objectives included: can identify, comprehend, and apply grammatical knowledge and vocabulary correctly in a specific context.

4.3.3 Reading Comprehension

From 2005 to 2007, the cloze and reading comprehension sections were in separate parts of the exam; there were two passages of 20 blanks in the cloze part and five passages in the reading comprehension part. From 2008, the cloze section became one passage of 15 blanks

and was included in Reading Comprehension (Section A). From 2005 to 2009, there were four passages of reading comprehension with 15 multiple-choice questions (Section B) and a passage of five paragraphs in which test-takers were required to choose an appropriate heading for each paragraph from six options (Section C). From 2010 to 2013, there were four sections in this part: A) cloze – a passage of 15 blanks that each with four multiple-choices; B) three passages of 11 multiple-choice questions; C) a passage of five paragraphs with blank headings; and D) a passage with 4 short-answer questions. Since 2014, there have been three sections in Reading Comprehension: Section A – a cloze passage of 15 blanks with four multiple options; Section B – three passages of 12 multiple-choice questions; and Section C – a passage with four short-answer questions.

Section A – Cloze Passage – was analysed with reference to the Examination Specifications. The items were analysed by referring to five different categories: 1: speculating on a meaning by understanding the context and word repetition phenomena; 2: finding answers by speculating on the passage according to the context; 3: finding answers by understanding details within the sentence; 4: obtaining answers by referring to similar words or meanings shown in the passage, and 5: analysing the article by using language knowledge such as collocations or idioms. Below is a paragraph from the passage of Section A of Reading Comprehension in the 2014 UEE, which shows how the question items were analysed with reference to the Examination Specifications.

2014: Reading Comprehension – Section A

Research has shown that two-thirds of human conversation is taken up not with discussion of the cultural or political problems of the day, not heated debates about films we've just watched or books we've just finished reading, but plain and simple 51. Language is our greatest treasure as a species, and what do we 52 do with it? We gossip. About others' behaviour and private lives, such as who's doing what with whom, who's in and who's out-and why; how to deal with difficult 53 situations involving children, lovers, and colleagues.

- | | | | |
|----------------------|----------------------|------------------|---------------|
| 51. A. claim | B. description | <u>C. gossip</u> | D. language |
| 52. A. occasionally | <u>B. habitually</u> | C. independently | D. originally |
| 53. <u>A. social</u> | B. political | C. historical | D. cultural |

For item 51, the answer is C. gossip; the word gossip appeared in the passage several times and according to the context it was talking about plain and simple gossip.

For item 52, the answer is B. habitually; the passage talked about conversation being taken up with recurrent topics in people's lives. From the first paragraph, B should be chosen by speculating about the context.

For item 53, the answer is A. social; the sentence talked about a conversation involving family, friends, and work colleagues. Therefore, A is the most appropriate option. The answer can be obtained by understanding details within the sentence.

Table 18 shows the number of items in the cloze passages which were tested according to different requirements set by the Examination Specifications. By analysing past examination papers, it can be found that over 60 percent of the blanks asked test-takers to find out the answers by speculating about the content according to the context, or understanding the details within the sentence. It means that test-takers were required to understand the details and the content throughout the article; by understanding the context, they were then able to find out the correct answers in the cloze passage.

Table 18: Cloze Items versus Objectives of Examination Specifications

Cloze (number of items)	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
1. Speculating on meaning by understanding the context and word repetition phenomena;	2	4	4	2	3	2	1	4	2	1
2. Finding answers by speculating on the passage (according to the context);	7	3	6	5	5	3	4	3	5	6
3. Finding answers by understanding details within the sentence;	8	9	7	4	5	6	4	5	7	5
4. Obtaining answers by referring to similar words or meanings from the passage;	0	2	2	2	0	3	3	3	0	3
5. Analysing the passage by using language knowledge.	3	3	1	2	2	1	3	0	1	0

For Section B and Section C of the Reading Comprehension, the question items were analysed in terms of four different categories in the Examination Specifications: 1: finding out the answers by comprehending details; 2: speculating on meaning by understanding words or sentences in context; 3: summarising or concluding the main idea from the paragraph or passage, and 4: inferring or speculating on the implied meaning from the context. The question items below were extracted from passage C in Section B. The tested items covered the requirements of the Examination Specifications.

2014: Reading Comprehension – Section B (Passage C)

74. The word 'swapping' (paragraph 1) is closest in meaning to_____.

- A. building; B. exchanging; C. controlling; D. transplanting

For item 74, the answer is B; from the context, test-takers were tested to speculate on the meaning by understanding the word 'swapping'.

75. We can infer from the experiment at the Be Another lab that_____.

- A. our feelings are related to our bodily experience
B. we can learn to take control of other people's bodies
C. participants will live more passionately after the experiment
D. 'The Machine to Be Another' can help people change their sexes

For item 75, the answer is A; test-takers were tested to infer the implied meaning from the passage.

76. In the Implicit Association Test, before the participants used virtual reality glasses to control a dark skinned digital character, _____.

- A. they fought strongly against racism
B. they scored lower on the test for racism
C. they changed their behaviour dramatically
D. they were more biased against those unlike them

For item 76, the answer is D; test-takers were tested to find out the answer by comprehending the details on the fourth paragraph of the passage about the implicit Association Test.

77. It can be concluded from the passage that_____.

- A. technology helps people realize their dreams
B. our biases could be eliminated through experiments
C. virtual reality helps promote understanding among people
D. our points of view about others need changing constantly

For item 77, the answer is C; test-takers were required to summarise the main idea from the passage for this item.

Table 19: Reading Comprehension Items versus Objectives of Examination Specifications

Reading Comprehension (number of items)	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
1. Finding answers by comprehending details in the passage;	6	5	5	5	5	6	5	5	6	6
2. Speculating on the meaning by understanding words and/or sentences correctly in context;	3	3	4	4	2	2	2	4	3	4
3. Concluding the main idea from the passage/paragraph;	4	3	3	3	4	3	4	3	4	4
4. Inferring the implied meaning from the context.	2	4	3	3	4	4	4	3	2	2

Table 19 shows the number of items in Reading Comprehension was tested according to different requirements set by the Examination Specifications. It can be seen that 40 percent of the items tested test-takers' comprehending details on the passage. The other three criteria were quite evenly distributed. In Reading Comprehension, test-takers were required to find out the answers by reading and understanding the details in the passages. Therefore, understanding the basic content of the article was one of the essential criteria in Reading Comprehension. Test-takers also were asked to find out meanings of some words or phrases or sentences from the articles by referring to the context. Some questions in the UEE tested test-takers if they were able to summarise or concluding the main idea of the passage and infer the hidden meaning of the article.

4.3.4 Writing

In Paper Two – Writing, there were two parts: Translation and Guided Writing.

In Part I – Translation, there were five Chinese sentences (six sentences from 2005 to 2009). Test-takers were required to translate the Chinese sentences into English by using the English word(s) given in the brackets. This part tested test-takers in three ways: 1: use of vocabulary; 2: sentence structure, especially compound sentences, and 3: grammatical knowledge and features, such as tenses, modal verbs, relative clauses, or non-predicative verbs. The example below is from the translation section of the 2014 UEE.

2014: Translation – Item 3

The Chinese sentence: 没有什么比获准参加太空旅行项目更令人兴奋的了。(than)

Suggested answer: There is nothing more exciting *than* being allowed to take part in the space travel programme.

The item tested test-takers' in the following criteria:

1. Use of vocabulary: more exciting than, take part in, space travel programme;
2. Sentence structure: there is nothing ... ;
3. Grammatical knowledge: more than, being allowed.

In Part II – Guided Writing, test-takers were required to write an English composition in about 120-150 words according to the Chinese instructions. This part tested test-takers for: 1: content and coherence; 2: use of vocabulary and grammar, and 3: composition structure and organisation. The instructions were in Chinese because the writing section was testing test-takers' writing ability but not their ability to read instructions. The test task below is from the guided writing section of the 2014 UEE.

2014: Guided Writing

The School English newspaper is preparing a makeover. It intends to remove one of the existing columns (health, entertainment, and culture), and add a new one (fashion, career planning, readers' feedback) to replace the removed option. You are Cheng Fei, a student from the school. You are writing an email to the editor to express your opinions.

Your email must include the following:

1. The existing column you suggest removing and your reasons;
2. The new column you propose to add and your reasons.

For the last 10 years, the questions covered different varieties of topics and different types of writing have been tested in the guided writing section; for example, comparing and contrasting two different favourite universities (2006); expressing your opinions on participating in a group dancing competition (2008); writing a letter to apply for funding to help poor children (2011); or writing an email to the editor of school newspaper (2014). Table 20 shows the topics and types of writing tested in the past 10 years of UEE papers.

Table 20: Topic and Type of Writing Tested in the Past 10 years' UEE Papers

Year	Topic	Type of writing
2005	A use for my talent	Narrative
2006	Making a choice between two favourite universities	Compare and contrast
2007	A gift	Descriptive
2008	Group dancing competition	Expressing opinions
2009	Summer camp	Letter writing
2010	Our primary school life	Describe a picture
2011	Fundraising	Letter writing
2012	Painting lessons	Diary writing
2013	An art exhibition	Expressing opinions
2014	School newspaper	Email to the editor

From the past 10 years of UEE papers, it is found that the materials testing test-takers in the writing section (both translation and guided writing) have covered what the objectives stated in the Examination Specifications. The objectives included translating fluent sentences by using language knowledge, and writing correctly, coherently, and aptly according to the meaning of question items or topics.

4.3.5 Speaking

There were five parts in the Oral English Test.

Part 1 required test-takers to read aloud a short essay – they had one minute to prepare and half a minute to read. Part 1 tested whether test-takers could grasp basic language knowledge and skills, and pronounce and intone the read aloud articles clearly and correctly. The aim was to test test-takers' ability to tell English stories by using clear and correct pronunciation.

Part 2 required test-takers to make quick responses to the sentences they heard. The quick responding task in Part 2 tested whether test-takers could use proper language functions and concepts for questioning and answering, requests and requirements, providing explanations and expressing opinions in their daily lives. The aim was to test test-takers' ability to communicate with people in English in real-world situations, to demonstrate that they could understand what people say by using their language functions.

Part 3 asked test-takers two questions each in two different situations. Similarly to Part 2, test-takers were tested on whether they were able to use appropriate language functions and raise suitable questions by using proper interrogatives such as who, what, which, when, where, why or how. It was advisable for test-takers to pay particular attention to the relevance of the contents and completeness of the sentence structures.

Part 4 gave test-takers a topic or question and they then had one minute to prepare and another minute to talk on the topic in at least six sentences. The topic talking task tested whether test-takers could express views on familiar topics based on personal experience in their daily lives. Test-takers were required to make their own comments logically and reasonably according to their experience and finally provide a conclusion of the stance they had taken. The objective tested whether test-takers were able to express personal opinions or ideas, and share thoughts and feelings with others.

Part 5 gave test-takers four pictures to describe and make a story from. They had one minute to prepare and another minute to talk about the pictures in at least seven sentences. The picture describing tested whether test-takers could narrate and describe some common life events and phenomena and make simple comments by telling or making stories. Test-takers were required to narrate and describe the stories clearly and coherently by using correct language and pronunciation. The objective of this part was to see if test-takers were able to give clear descriptions of general and familiar topics.

The content of the Oral English Test matched with the objectives in the Examination Specifications. The objectives included: 1: can grasp basic language knowledge and skills and can pronounce and intone clearly and correctly the read aloud articles; 2: can use proper language functions and concepts for questioning and answering, requests and requirements, providing explanations and expressing opinions; 3: can express views and opinions on familiar topics based on personal experience, and 4: can narrate and describe some common life events and phenomena, and make simple comments.

4.4 Findings and discussion: Part 2 English Curriculum Standards

4.4.1 Language skills

4.4.1.1 Listening skill

Table 21: Checklist for Listening Skill of ECS

Objectives for Listening Skill at Level 8 of Curriculum Standards:	Fulfilled or not	Test tasks and examples:
1. Be able to recognise different attitudes through different tones;	✓	Short conversations with multiple-choice items, for example, UEE 2006:7: M: Would you like to go to the ballet next Friday? I've got two tickets. W: Oh, that sounds like fun. Q: How does the woman feel?
2. Be able to understand discussions or conversations on familiar topics, and remember the main points;	✓	Long conversations with table-filling items, for example, shopping (2007), refugees (2008), school system in England (2009), skateboarding/sports (2013), or critical thinking (2014).
3. Be able to grasp the viewpoints in simple passages;	✓	Passages with multiple-choice items, for example, examination system in Britain (2006), regulations in National Parks (2008), daily life of a hotel manager (2010), or company management (2012).
4. Be able to figure out the themes or main ideas in English language broadcasting and television news;	✓	Passages with multiple-choice items, for example, a news report on the radio (2005), or a documentary programme about Canada (2012).
5. Be able to understand euphemistic or implied meanings of suggestions or advice.	✓	Short conversations with multiple-choice items, for example, UEE 2013:9: W: What's your advice for someone to become a fashion designer? M: Go to school. I mean it. Find a good school and learn as much as you can. Q: What does the man mean? Suggested Answer: School learning is a must for fashion designer.

In the Listening Comprehension part of the UEE, test-takers were required to recognise or notice different tones or intonations of the conversations, so that they could then infer the meanings. Topics covered in the UEE were related to daily life and included shopping, travelling, school, and friendship. In respect of any topic, test-takers were required to understand the details of the conversations.

Further, in Section A, some of the short conversation items contained euphemistic and implied meanings. Using idioms or proverbs to express oneself is very common in English-speaking countries; each country has many of its own idiomatic or proverbial expressions; understanding of such kinds of expression is very important and crucial to minimising any misunderstanding between speakers.

In Section B of the listening comprehension of the UEE, some passages were like news reports. Test-takers had to identify the information and figure out the main themes from the reports, which were similar to the programmes from English language broadcasting and television news. Some question items also required test-takers to summarise the main idea of the passage or conversation.

From Table 21, it can be observed that all objectives in the English Curriculum Standards for listening skills were fulfilled by the question items of the past 10 years of UEE papers.

4.4.1.2 Reading skill

Table 22: Checklist for Reading Skill of ECS

Objectives for Reading Skill at Level 8 of Curriculum Standards:	Fulfilled or not	Test tasks and examples:
1. Be able to understand different opinions and attitudes in reading materials;	✓	There was at least one argumentative passage in Reading Comprehension testing test-takers' understanding of the writer's opinions in the passage each year.
2. Be able to recognise the features of different types of writing;	✓	Different types of writing appeared in the passages, for example, narrative, expository, argumentative, journal article, or interview.
3. Be able to understand long sentences and difficult sentences by analysing sentence structures;	✓	Questions like "Which of the following statements is true according to the passage?" tested test-takers' comprehension of long and difficult sentences from the passage.
4. Be able to enjoy basic literature with the help of teachers;	x	No literature was tested in the past 10 years' UEE Papers.
5. Be able to get and then process the information required by learning tasks from e-books or the Internet;	x	Students should learn in the classroom how to process information from the Internet.
6. Should do as much reading as over 360,000 words outside textbooks.	x	Students should set up a reading plan in school and evaluate the plan from time to time with their teacher.

In the Reading Comprehension part, the passages contained different types of writing, such as narrative, descriptive, argumentative, and expository. The question items in those passages examined test-takers' ability to comprehend different ideas and meanings by analysing sentence structures. These tested items covered half of the objectives in the English Curriculum Standards (See Table 22). However, for item 4 – being able to enjoy plain literature, it was very difficult for teachers to quantify how much students enjoyed reading literatures or novels, as reading enjoyment is quite personal. Item 5 – being able to access and process information through the Internet, could be only assessed by classroom practice rather than the examinations; students needed to be encouraged to spend more time on learning and practising searching information by using English language on the internet. Regarding item 6 – reading outside textbooks, students were encouraged to read outside textbooks and share with other schoolmates in class. Students were expected to set up a reading plan of how many books should be read through every year, and suitably graded readers should be recommended to the students in school. Students were expected evaluate their reading plan with their teacher intermittently. Students could also be assessed by writing book reports and making presentations in class after they had finished reading the graded readers.

4.4.1.3 Writing skill

Table 23: Checklist for Writing Skill of ECS

Objectives for Writing Skill at Level 8 of Curriculum Standards:	Fulfilled or not	Test tasks and examples:
1. Be able to write coherent and structurally complete compositions; narrating events or expressing your own opinions and attitudes;	✓	Guided writing test tasks, for example, UEE 2013: Shanghai Museum was organising an Art Exhibition. Test-takers were required to write a letter to the Museum expressing their opinions.
2. Be able to paraphrase or summarise articles after reading;	x	No item tested test-takers' ability to summarise articles from reading passages in the UEE papers.
3. Be able to write in appropriate styles with natural and grammatical sentences;	✓	Translation and guided writing test tasks: grammar, language use, and sentence structure were the marking criteria in the writing section.
4. Be able to write short essays or reports based on textual and graphical information provided.	✓	Guided writing test tasks, for example, UEE 2006: Text and figures about two universities were provided; test-takers were required to compare the two and make a choice.

For the writing section, both translation and guided writing tasks tested test-takers' ability to translate fluent sentences by using language knowledge, and to express their ideas by writing correctly, coherently, and aptly according to the meaning of topics. They were required to write on different types of topics, such as narrative story, picture description, or giving opinions on some issues.

The UEE papers never tested students on paraphrasing or summarising from reading texts. Students were only tested separately in the reading and writing sections. In Reading Comprehension, only short answers were required; in the writing sections, a clear instruction was given and test-takers were asked to write an essay on a given topic. Test-takers were unfamiliar with this kind of integrated reading and writing. In summary, most objectives in the English Curriculum Standards were fulfilled, except item 2 – testing test-takers' ability to paraphrase or summarise articles after reading them (See Table 23).

4.4.1.4 Speaking skill

Table 24: Checklist for Speaking Skill of ECS

Objectives for Speaking Skill at Level 8 of Curriculum Standards:	Fulfilled or not
1. Be able to use proper intonations and rhythms;	✓
2. Be able to discuss and then make plans in accordance with learning tasks;	x
3. Be able to report the processes and results of experiments or research;	x
4. Be able to make three-minute prepared speeches on common topics;	x
5. Be able to make effective uses of the English language to judge, criticise, complain, or express opinions in daily life communications;	✓
6. Be able to do translations in such daily life activities as shopping and sightseeing.	x

Since the Oral English Test questions and test tasks were not available, this section was analysed according to the samples of the Oral English Test provided by the Shanghai Examination Centre. As mentioned in Section 4.3.5, regarding the Oral English Test format, test-takers were required to read aloud a short essay, make responses and ask questions about some situations, give a one-minute talk, and describe some pictures.

The Oral English Test was conducted using a computer; the questions were shown on the computer screen and test-takers were required to answer by speaking to the microphone speaker. This kind of speaking test was considered inauthentic because in real life, conversations between speakers should be interactive. By comparing the test with the objectives of the Curriculum Standards for speaking skills, it can be found that most of the objectives were not fulfilled (See Table 24).

In the Oral English Test, there was no opportunity for test-takers to discuss any topics with other students. They did not have any opportunities to express their own ideas related to daily life issues. In the Oral English Test, making quick responses to the questions (Part 2) and asking questions (Part 3) under different situations tested only part of the conversational practice. If people would like to have deep conversations or discussions, other skills are essential, such as being able to question and answer, express opinions or ideas, agree or disagree, or paraphrase other's ideas. For the spoken presentation, test-takers were only required to prepare a talk on a familiar topic for one minute (not the preferred three minutes). Also, no interpretation was required in the Oral English Test; therefore, test-takers did not have any opportunities to practise their interpretations along the lines they had learned in school.

One of the criteria in communication is non-verbal communication, which includes eye contact, facial expressions and gestures. Since test-takers were required to talk to the computer, but not the examiners, they did not listen to the questions from the machine but read the questions off the computer monitor. Therefore, it was not a real communicative process and authentic interaction.

The Oral English Test in Shanghai was not compulsory and the result scores were not shown on the final result record. Only those students applying for the English Major subjects in the university were required to show their Oral English Test scores to the university. Therefore, senior high schools did not expend any resources on speaking lessons in their syllabus and teachers seldom had any discussion or presentation activities in class for their students to practise speaking English.

4.4.2 Language knowledge

4.4.2.1 Phonetics

Table 25: Checklist for Phonetics of ECS

Objectives for Language Knowledge (Phonetics) at Level 8 of Curriculum Standards:	Fulfilled or not	Test tasks and examples:						
1. Gradually acquire natural, appropriate, and fluent pronunciation and intonation through communication;	✓	Oral English Test: Part 1 – Reading Aloud; reading aloud a short passage in half a minute – this test task focused on pronunciation and intonation.						
2. Be able to recognise and express implicit intents and attitudes with pronunciations and intonations;	✓	Listening Comprehension, included short and longer conversations and passages, tested test-taker’s understanding and comprehension of the meaning of conversations and passages by recognising pronunciations and intonation.						
3. Have some knowledge of the rhythms in poems;	x	No questions or test tasks tested student’s knowledge about poems in the UEE.						
4. Be able to recognise and spell unfamiliar words or simple sentences after hearing them pronounced.	✓	In Listening Comprehension – Longer Conversations: <table border="1"><tr><td>Class Registration Form</td></tr><tr><td>Name: <u>Andrew Smith</u></td></tr><tr><td>Department: The <u>17</u> Department</td></tr><tr><td>Student ID: <u>18</u></td></tr><tr><td>Class: The <u>19</u> Class</td></tr><tr><td>Time: <u>20</u>; 2:00 – 4:00 p.m.</td></tr></table> Test-takers were required to fill in the form from the conversation. (UEE 2012:17-20)	Class Registration Form	Name: <u>Andrew Smith</u>	Department: The <u>17</u> Department	Student ID: <u>18</u>	Class: The <u>19</u> Class	Time: <u>20</u> ; 2:00 – 4:00 p.m.
Class Registration Form								
Name: <u>Andrew Smith</u>								
Department: The <u>17</u> Department								
Student ID: <u>18</u>								
Class: The <u>19</u> Class								
Time: <u>20</u> ; 2:00 – 4:00 p.m.								

Test-tasks in Listening Comprehension tested test-takers' recognition and understanding of pronunciation, intonation and the meanings of their conversations. Test-takers were required to recognise and write down words, and fill in tables (spelling) when they listened to the longer conversations. Therefore, objectives 2 and 4 for language knowledge (phonetics) as shown above were fulfilled (See Table 25).

Meanwhile, pronouncing accurately and clearly was one of the most important parts in the Oral English Test. In Part 1 – Reading Aloud, test-takers were given a short passage and they had one minute to prepare and half a minute to read. They were expected to be able to use their basic language knowledge and skills, make clear and correct pronunciation and intonation to read aloud articles, and use proper intonations and rhythms to read the passages smoothly and fluently. However, test-takers were speaking to the computer; it was a one-way communication so objective 1 was only marginally achieved.

Regarding objective 3, there were no test tasks testing any knowledge about the poems in the UEE. Also in the Oral English Test, test-takers were required to read aloud a short passage but not a poem by using proper intonations and rhythms. It seemed that test-takers were not required to learn any knowledge about the poems in school and objective 3 was not achieved. Other than that, objectives for language knowledge (phonetics) in the English Curriculum Standards were accomplished.

4.4.2.2 Vocabulary

Table 26: Checklist for Vocabulary of ECS

Objectives for Language Knowledge (Vocabulary) at Level 8 of Curriculum Standards:	Fulfilled or not	Test tasks and examples:
1. Be able to understand and express different functions, attempts, and attitudes with words and phrases;	✓	In Listening Comprehension – Short Conversation, W: You will take care, won't you? The roads are very icy. M: I will drive very slowly. I promise. Q: How does the woman feel? (UEE 2014:2) Answers: A. confident; B. puzzled; C. satisfied; D. worried. This question tested test-taker's understanding of vocabulary to describe the woman's feeling.
2. Use words and phrases to describe complicated objects, behaviour, features, and concepts;	✓	Translation and guided writing tasks; writing section tested test-taker's ability to use appropriate vocabulary to describe and translate sentences.
3. Be able to use 3,000 words and 400-500 phrases or collocations.	✓	Students were required to learn a certain number of words listed in the Vocabulary Handbook in Senior High School before taking the UEE.

The University Entrance Exam Vocabulary Handbook for English Language was issued by the Examination Centre. This handbook contained a recommended vocabulary (about 500 words) that senior high school students were required to learn and to understand before sitting the Exam. Test-takers were expected understand and know how to use these words and phrases in the Exam.

In the listening comprehension section, test-takers were expected to have a certain level of vocabulary in order to understand the conversations and passages in the audio and complete the tasks in the Exam. On the other hand, in the Oral English Test, test-takers were required to understand and express different functions, attempts and attitudes with appropriate use of words and phrases in different test-tasks. They were also required to describe complicated

objects, features and concepts by using appropriate words and phrases in different tasks such as topic talking in Part 4 and picture describing in Part 5.

Test-takers were required to understand and comprehend vocabulary in the passages in the reading comprehension section. The more vocabulary the test-takers understood, the easier they comprehended the passages in the reading comprehension section.

In the writing section, Part I – Translation, tested test-takers’ ability to use vocabulary when test-takers translated the sentences from Chinese into English, whereas in Part II – Guided Writing, test-takers were prompted to use words and phrases to describe objects or concepts, or narrate stories or events, or express their opinions and views on particular topics.

In summary, the UEE content achieved all three objectives for language knowledge (vocabulary) in the English Curriculum Standards (See Table 26).

4.4.2.3 Grammar

Table 27: Checklist for Grammar of ECS

Objectives for Language Knowledge (Grammar) at Level 8 of Curriculum Standards:	Fulfilled or not	Test tasks and examples:
1. Further grasp the expressions that are used to describe time, place, and direction;	✓	Grammar section with multiple-choice items; Time: You can’t borrow books from the school library <u>before</u> you get your student card. (UEE 2009:32) Place: At a rough estimate, Nigeria is <u>three times the size of</u> Great Britain. (UEE 2005:28)
2. Further understand and grasp the expressions that are used to compare people, objects, and events;	✓	Grammar section with multiple-choice items; People: In ancient times, people rarely travelled long distance and most farmers only travelled <u>as far as</u> the local market. (UEE 2010:30) Objects: Yet I think the term “automatic shop” is far <u>more appropriate</u> . (UEE 2014:37) Events: London’s not as expensive in price as Tokyo but Tokyo is <u>more organized</u> in traffic. (UEE 2008:27)
3. Use proper language forms to describe or express opinions, attitudes, and emotions;	✓	Guided writing tested test-takers’ ability to use proper language and grammar to describe or express opinions or views on different topics.
4. Learn and grasp basic discourse knowledge, and be able to effectively organise the information for certain purposes.	✓	Discourse knowledge is the basic foundation for test-takers in language skills.

Test-takers were required to have basic and minimal discourse knowledge so that they were able to effectively organise the information from the audio for particular purposes in the listening comprehension section. For example, one important variable is cohesion, which is a semantic relation between one element in a text and another element that has to be understood. Cohesive devices like conjunctions and pronouns indicate important semantic relations between the ideas they connect.

In the second longer conversation of Listening Comprehension, test-takers were required to complete the form with no more than three words for each answer. Therefore they were required to paraphrase from the conversation and fill in the form with correct grammar in order to complete the sentences in the table. Grammatical knowledge was thus tested in the listening section.

On the other hand, test-takers were required to use proper language forms to describe pictures or narrate stories or express opinions on particular topics. Being able to use correct grammar affects not only the accuracy of their expressions but also fluency with which questions were answered or ideas expressed on topics in the Oral English Test.

In Reading Comprehension, test-takers were required to have basic discourse knowledge so that they were able to effectively organise the information from reading materials. Being able to understand different parts of speech in the functions and meanings of a sentence helped test-takers understand long sentences and difficult sentences by analysing sentence structures.

One of the marking criteria in the writing section was the use of grammar. Test-takers were expected to use proper grammar and vocabulary in their writing – both translation and guided writing. They should be able to write normative, fluent, accurate and logical compositions by using their grammatical knowledge.

Basically, grammatical knowledge was very important in all listening comprehension, reading comprehension, writing, and speaking sections in the UEE. The content included in the UEE accomplished the objectives for language knowledge (grammar) in the English Curriculum Standards (See Table 27).

4.4.2.4 Functions

Table 28: Checklist for Functions of ECS

Objectives for Language Knowledge (Functions) at Level 8 of Curriculum Standards:	Fulfilled or not	Test tasks and examples:
1. Be able to understand and express communicative functions such as greetings, farewells, thanks, and introductions in larger contexts;	✓	In Oral English Test: Part 2 – Quick Responding, test-takers were required to give quick responses on the basis of given situations, for example, 1. Hello, this is Mary speaking. Is Peter there? 2. Would you mind not talking so loudly in class? 3. Is there anything I can do for you?
2. Be able to effectively and appropriately use the English language to judge, criticise, complain, and express opinions in communication in daily life;	✓	In Oral English Test: Part 4 – Topic Talking, test-takers were asked to provide their own opinions on the given topic, which was life-related, for example, 1. Do you think online games are good for middle school students? Why or why not? 2. What is the best way to keep healthy? Why?
3. Make flexible use of learned commonly used functions, and learn and grasp the new ones;	✓	In Oral English Test: Part 3 – Question Raising, test-takers were required to raise questions for the situation given, for example, It is your mother's birthday tomorrow and you have no idea about what to buy as her birthday gift. Ask your friend for advice.
4. Make experienced use of the English language to communicate in real life situations.	✓	In the writing section – Letter writing, an overseas school is organising a summer camp, four programmes – gardening, cooking, nursing, and self-defence – are provided for students to choose from, test-takers were required to write an application letter for one of the programmes in the camp. (UEE 2009: Guided-writing)

Most functional knowledge was tested explicitly in Listening Comprehension, most items tested test-takers' understanding of hidden meanings in the discourse, and required test-takers to grasp the main idea of the conversations and passages and realise the purpose of information communication in real lives.

The quick responding task in Part 2 and the question raising task in Part 3 tested whether test-takers could use proper language functions and concepts for questions and answers, requests and requirements, giving explanations, and expressing attitudes in daily life.

Similarly in Reading Comprehension, a few question items tested implied meaning in the passage – test-takers were required to guess at the hidden message in the passage. Also, basic conveying of ideas, indirect meanings or text-based inferences were tested in Reading Comprehension of the UEE.

Test-takers were required to write an English composition of about 120-150 words according to the instructions given in the UEE. The instructions included the contents, background, information and outlines. Test-takers were expected to make use of their language knowledge and learned common functions to express their opinions or views of their writing. They were expected to be able to make an experienced use of the language to realise communicate information in real life situations.

Functional knowledge in language is about how to use language properly to accomplish various functions in communication (Weigle, 2002). The UEE content covered all the objectives for language knowledge (functions) in the Curriculum Standards (See Table 28).

4.4.2.5 Topics

Table 29: Checklist for Topics of ECS

Objectives for Language Knowledge (Topics) at Level 8 of Curriculum Standards:	Fulfilled or not	Test tasks and examples:
1. Get familiar with the topics on individuals, families, social interactions, and so forth;	✓	In Section B and C of Listening Comprehension, many familiar topics were covered, such as the internet (2007), national parks (2008), the environment (2009), tourism (2010), or medical and health (2014).
2. Have further knowledge of the topics in daily life, hobbies and interests, manners and customs, science and cultures, and so forth;	✓	In Oral English Test: Part 5 – Picture Describing, test-takers were required to describe the pictures on the computer screen; the pictures were related to daily life, for example, sports, food and health, or the environment.
3. Be knowledgeable about Chinese social and life topics such as professions, festivals, customs, and social etiquettes;	✓	In the writing section, some topics were related to Chinese social life and culture, for example, test-takers were asked to write what they felt about primary school life in China. (UEE 2010)
4. Have some knowledge of the topics on English-speaking countries' daily lives and habits.	✓	The passages in Listening and Reading Comprehension gave examples of topics related to English-speaking countries, for example, the British examination system (2006), Canadian English (2012), Physical Education in America (2010), or Sydney transportation (2012).

The topics in the passages and conversations in Listening Comprehension were related to different topics, such as education (2006), the internet (2007), news reports (2008), or narrative stories (2013). If test-takers had such topical knowledge, it would help them to be more familiar with the listening topical backgrounds. World knowledge can influence listening comprehension; specific knowledge can be used to fill in details that are not explicitly mentioned in the passage (Buck, 2001).

One of the tasks in the Oral English Test was to talk on a topic for one minute. The range of the topics was wide; it covered education, Internet, health and life, social activities in school, friendship, relationship, transportation, and cultures. Also in Part 5, test-takers were given four pictures and required to make a story.

Passages in Reading Comprehension covered different types of texts, such as narrative, expository, argument, and journal articles. They covered different kinds of topics: environmental issues, nature, narrative stories, school and education, an interview with a musician, the world economy, health and food, computers and the internet, science, and cultures.

Also, different topics have been covered in the guided writing section for the past 10 years' UEE papers. For example, students could narrate their lives by explaining an old Chinese saying (2005), make a decision to choose an appropriate university to study by comparing information given (2006), write a letter to express their interests in a summer camp (2009), write a letter to the Shanghai Museum to express their opinions on an art exhibition (2013). These topics were familiar, arising as they did out of the circumstances of test-takers' daily lives.

The UEE content achieved all objectives for language knowledge (topics) in the English Curriculum Standards (See Table 29).

4.4.3 Affect and attitude

Table 30: Checklist for Affect and Attitude of ECS

Objectives for Affect and Attitude at Level 8 of Curriculum Standards:	Fulfilled or not
1. Maintain the interest and motivation to learn English and get involved in activities which can help to improve active use of English;	x
2. Have the appropriate English learning motivation and understand clearly that English learning is for communication and expression;	x
3. Keep confident in English learning and have the courage to communicate with others in English;	x
4. Have the ability to overcome all difficulties in English learning and be willing to ask others for help;	x
5. Be able to understand and respect the feelings of others when communicating with them;	x
6. Have a highly cooperative spirit and be willing to share various kinds of learning resources with others;	x
7. Be able to introduce China's national culture to others in English;	x
8. Be able to understand and respect foreign cultures and demonstrate an international cooperative spirit in your actions.	x

Affect and attitude influence the learning process by students. They also take in the gradual cultivation during the learning process of student national awareness of both the nation and the international community. The key to mastering English is to always keep a positive learning attitude. Teachers should constantly stimulate and reinforce student interest in learning, guiding them to holding on to a stable learning motivation, which will instil confidence, make students willing to cooperate with others, and to be positive about study. However, these aspects of English language learning were not easily observed and assessed; therefore, none of the UEE items or test tasks involved affect and attitude as stipulated in the English Curriculum Standards. It is probable that because these items were not tested, teachers did not think they were very important for students' English learning and therefore never introduced ideas of affect and attitude to their students in the classroom, nor did they mention any elements related to the motivation for and attitudes to learning English (See Table 30).

4.4.4 Learning strategies

Table 31: Checklist for Learning Strategies of ECS

Type	Objectives for Learning Strategies at Level 8 of Curriculum Standards:	Fulfilled or not
Cognitive Strategy	1. Establish relations amongst different kind of relevant knowledge with using your imagination;	✓
	2. Analyse and solve problems with the help of logically by reasoning and summarising;	✓
	3. Be good at summarising general linguistic principles from texts and be able to apply them to your own situations;	✓
	4. Be capable of focusing on key points, writing down notes and summarising knowledge in the process of learning;	✓
	5. Be able to guess or speculate on the meanings of new words from the context or in specific situations, during the process of reading and listening;	✓
	6. Use diagrams and other non-verbal information to help express yourself or understand an idea as you are learning.	✓
Metacognitive Strategy	1. Make an English learning plan suited to your purpose/need;	x
	2. Look for more ways to learn English actively;	x
	3. Be good at creating or grasping opportunities to learn English;	x
	4. Know how to get help when facing difficulties in learning;	x
	5. Communicate with teachers and/or classmates on your learning experience;	x
	6. Evaluate learning performance, summarise effective learning methods and improve the memorising effect according to the rules of memory.	x
Communicative Strategy	1. Communicate actively with classmates in English both in curricular and extra-curricular activities;	x
	2. Be good at using gestures, facial expressions and other non-verbal communications to improve the effect of communication;	x
	3. Be able to overcome linguistic difficulties and maintain the communication;	x
	4. Be capable of grasping every opportunity to communicate with others in English in daily life;	x
	5. Observe and be aware of basic etiquette in English communication.	x
Resourcing Strategy	1. Acquire more information about English with the help of library resources, the internet, radio, television, etc.	x

Cognitive strategies in learning English are the methods used by learners to solve problems in language-use tasks or test-tasks; they take into account such skills as reasoning, speculating, processing, and summarising. From the past 10 years of UEE papers, most of the question items involved assessing test-takers' strategies for mastering the English language. For example, speculating on the implied meaning of conversations or passages, focusing on key points and jotting notes in Listening Comprehension. Test-takers were also required to guess

the meanings of new words or phrases in passages by understanding the context and by summarising the main idea of the passages in Reading Comprehension. In the writing section, they were asked to describe pictures and write letters or emails to express their ideas on particular situations.

The objectives in *metacognitive strategy* involved strategies for creating opportunities to learn English language. Students and teachers were encouraged to evaluate their strategic plan from time to time in order to check how effective they were in learning English. In cases where students had not achieved what they had planned, they were encouraged to discuss the difficulties with their teachers and schoolmates and readjust their plan if necessary.

Communication strategy is “a systematic technique employed by a speaker to express his meaning when faced by some difficulty” (Corder, 1983, as cited in Douglas, 2000, p.79). Communication between a sender and a receiver is a purposeful activity involving an exchange of information and meaning. Students are encouraged to communicate or discuss actively with classmates both in class and out-of-class activities in order to practise speaking and listening. However, in the past 10 years of UEE papers, all question items or test-tasks were a ‘one-way’ communication; for example, in the listening comprehension section, test-takers were required to answer questions by listening to the audio. Even for the Oral English Test, test-takers were expected to talk to the computer when they received the questions or see the pictures on the screen. These test-tasks did not encourage students to practise English communication in schools.

Resourcing strategy aimed to encourage students to use resources from the media to expand their knowledge of English. It only worked when teachers taught their students in school how and what kind of resources or information they should acquire; it was difficult to measure this objective through tests or examinations.

In short, only the objectives in cognitive strategy were achieved. None of the objectives in other three strategies (metacognitive strategy, communicative strategy and resourcing strategy) were achieved, even in part. Teachers would ideally discuss learning strategies with students in the classroom and students needed their plans evaluated to keep them on the right track (See Table 31).

4.4.5 Cultural understanding

Table 32: Checklist for Cultural Understanding of ECS

Objectives for Cultural Understanding at Level 8 of Curriculum Standards:	Fulfilled or not
1. Understand common idioms, proverbs and their hidden meanings in English;	x
2. Understand legends and allusions which are commonly used in communication;	x
3. Know the names of the major writers, artists, scientists and their experiences, achievements and contributions;	x
4. Have a general idea about the political and economic circumstance of English-speaking countries;	x
5. Know something about the major mass media in English-speaking countries;	x
6. Know about differences in lifestyles between China and English-speaking countries;	x
7. Know about the differences of behaviour and manners between Chinese and foreigners;	x
8. Know about religious traditions in English-speaking countries;	x
9. Understand various cultures and cultivate an awareness of the international society by learning English;	x
10. Deepen your understanding of Chinese culture by comparing Chinese culture with foreign cultures.	x

Language contains abundant cultural connotations. Knowledge of foreign cultures involved in the process of teaching and learning should be relevant to the daily life of the students and stimulate their interest in learning English. In fact, no foreign cultures were taught or introduced in the classroom. Also, the textbook (*The Oxford English for Senior II*) the school used did not mention anything related to cultures of Chinese and English-speaking countries. From the past 10 years of UEE papers, no foreign cultural knowledge was taught or assessed in either school or the UEE (See Table 32).

Possibly because the UEE did not include intercultural issues, teachers did not think intercultural communication was important; notwithstanding, it was one of the important ingredients in the curriculum. It is true that the objectives were not tested in the UEE, but that should not mean that intercultural issues should not be addressed in class. Teachers should bring intercultural communication awareness and knowledge into the classroom by using different channels, for example showing documentary programmes or other media items about the culture of English-speaking countries; asking students to conduct research on the

internet and report in class, doing role-plays to act as different people from different cultures, inviting foreigners to talk to students about foreign cultures, and so on. Only when the students have had real interactions with foreigners, will they be able to deepen their understanding of foreign cultures. Irrespective, although cultural understanding was not easily observed in class, students could still be assessed on the topics related to cultural understanding in the classroom.

4.5 Conclusion

Among 76 objective items, only just half of the items in the English Curriculum Standards were achieved (See Table 33). In respect of *language skills*, all objective items in the listening section were fulfilled, whereas only half of those items in the reading section were fulfilled: reading planning and resource access were tested in the UEE. In the writing section, most items were achieved, except the one which required integrated writing by summarising or paraphrasing articles after reading. Only one-third of the items were achieved in the speaking section; none except test-takers' pronunciation, intonations and rhythms. The reasons for this poor result in the Oral English Test were: 1: speaking test tasks were 'one-way' communication, in which test-takers talked to the computer, no discussion and communication with other people was involved; 2: no speaking interpretation test items were involved in the Oral English Test, and 3: no presentation for projects or experiment or long prepared speech was required in the UEE.

Table 33: Summary of Objective Items Achieved in the English Curriculum Standards

Curriculum Standards	Total number of objective items	Total number of items achieved	By percentage
Learning skills	21	14	67%
Learning knowledge	19	18	95%
Affect and attitude	8	0	0%
Learning strategies	18	6	33%
Cultural understanding	10	0	0%
Total	76	38	50%

In respect of *language knowledge*, most of the objective items were achieved, except the objective of having some knowledge of the rhythms in poems in phonetic type: nothing related to poems in the Oral English Test was tested. For the two categories *affect and attitude* and *cultural understanding*, none of the items were achieved. Finally, for *learning strategies*, only those objectives in the cognitive strategy type of strategy were achieved. The other three types of strategy, which related to learning English language, communication and how to make use of resources to enhance English learning, were not achieved. These objectives could be evaluated in the classroom rather than being tested in the UEE.

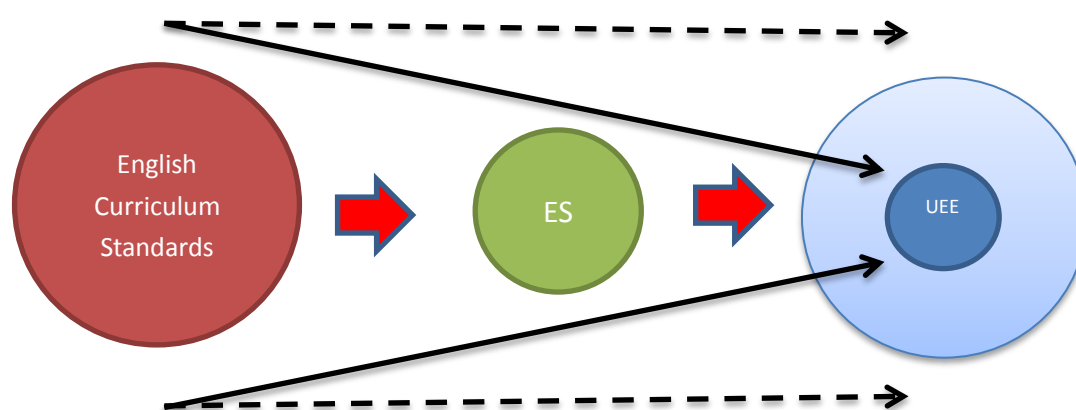


Figure 7: Narrowing down the English Curriculum Standards

The past 10 years' UEE papers have been checked against both the Examination Specifications (ES) and English Curriculum Standards (ECS). It was found that all objectives of the ES matched well with the UEE, but only half of the objectives were achieved in the ECS. The main reason is that the ES was designed by a group of professionals under the Shanghai Examination Centre, and they had to read their ES before they constructed the UEE papers. Although the ES was based on the ECS, they were designed by two different groups of people. The researcher doubted that the UEE designers had read the details of the ECS when they constructed the UEE papers. Ideally, the ECS should be directly related to the UEE (refer to the dash line in Figure 7), where the test items were supposed to reflect all the objectives listed in the ECS; however, the UEE was narrowed down by the objectives of the ES (refer to the solid line in Figure 7). As a result, the UEE papers were very far removed from the ECS.

This chapter has summarised what was tested in the UEE and how closely UEE content matched the English Curriculum Standards. The UEE has covered most of the objectives in the area of *language skills* and *language knowledge* but very few of *learning strategies*. However, nothing was tested in the area of *affect and attitude* and *cultural understanding*. In short, this study has shown that the UEE bears a very imbalanced relationship to the English Curriculum Standards.

Chapter 5: Content validity: Bachman and Palmer's model of language ability

5.1 Introduction

This chapter investigates whether the English Language Exam for university entry in Shanghai tested the language abilities specified in Bachman and Palmer's (1996) model. This is done through a qualitative, content analysis of the past 10 years' UEE papers, which are then matched against Bachman and Palmer's model of language ability.

5.2 Research method

5.2.1 Materials

Papers from the past 10 years' (2005-2014) English Language Exam for university entry in Shanghai were used for the analysis. There were two papers in the UEE (See Table 34). In Paper I, there were three parts: Listening Comprehension, Grammar and Vocabulary, and Reading Comprehension. In Paper II, there were two parts: Translation and Guided Writing. The total score of the examination paper was 150 and time allowed for the exam was two hours.

Table 34: The Examination Format of the UEE English Language Exam

Paper I		
Part I	Listening Comprehension	<i>Section A – short conversations Section B – passages Section C – long conversations</i>
Part II	Grammar and Vocabulary	<i>Section A – grammar Section B – vocabulary</i>
Part III	Reading Comprehension	<i>Section A – cloze Section B – passages (multiple-choice) Section C – passage (short-answer)</i>
Paper II		
Part I	Translation	<i>Translate 5 Chinese sentences into English</i>
Part II	Guided Writing	<i>Write a composition of 120-150 words</i>

5.2.2 Bachman and Palmer's model

Bachman and Palmer's (1996) model defines language ability as involving two components: language knowledge and strategic competence. It focuses on the way language ability is

influenced by personal attributes, topical knowledge, and affective schemata, and showing how these areas interact with each other in language use circumstances (see Figure 8).

Language knowledge includes two broad categories: organisational knowledge and pragmatic knowledge.

There are two areas of organisational knowledge:

- Grammatical knowledge: vocabulary, syntax, phonology, and graphology;
- Textual knowledge: cohesion and rhetorical or conversational organisation.

Similarly, there are two areas of pragmatic knowledge:

- Functional knowledge: ideational, manipulative, instrumental, and imaginative;
- Sociolinguistic knowledge: dialect or variety, registers, natural or idiomatic expressions, cultural references and figures of speech, and genre.

Strategic competence is understood as a set of metacognitive components or strategies, which provide a management function in language use. These components function in three areas: goal-setting, appraising and planning.

Figure 8: Bachman and Palmer's model of language ability (1996, p.63)

5.2.3 Data analysis

As in Chapter 4, content analysis has been used to analyse the data in the study. Test items or test tasks in different sections in the past 10 years' UEE papers – listening comprehension, grammar and vocabulary, reading comprehension, writing, and speaking sections – were checked, analysed and identified against Bachman and Palmer's model of language ability. If they were identified, examples would be listed and explained in detail. In this study, the analysis was validated by asking a colleague to independently check all items against the area of functional knowledge, and the inter-coder reliability checks reached 99% agreement with the coding of the researcher.

A checklist was derived from Bachman and Palmer's model (1996, p.77; see Table 35). The checklist was very useful for determining what the UEE was testing, i.e., the construct. It helped measure the degree to which components of language ability were tested in the UEE. Based on analysis of the prompt of the question items and test tasks, comments were made in the checklist. If the components were relevant, 'required' was entered, otherwise 'not required' or 'marginally required' were entered.

Table 35: Components of Language Ability – A Test Analysis Checklist

Components of Language Ability	Comments
Grammatical knowledge <i>Vocabulary</i> <i>Syntax</i> <i>Phonology and graphology</i>	
Textual knowledge <i>Cohesion</i> <i>Rhetorical or conversational organisation</i>	
Functional knowledge <i>Ideational functions</i> <i>Manipulative functions</i> <i>Heuristic functions</i> <i>Imaginative functions</i>	<i>Required</i> <i>or</i> <i>Marginally required</i> <i>or</i> <i>Not Required</i>
Sociolinguistic knowledge <i>Dialects and varieties</i> <i>Registers</i> <i>Natural or idiomatic expressions</i> <i>Cultural references and figures of speech</i> <i>Genres</i>	
Strategic competence <i>Goal-setting</i> <i>Appraising</i> <i>Planning</i>	

5.3 Findings and discussion

5.3.1 Grammatical knowledge

Grammatical ability is present to the degree that an individual can demonstrate “the capacity to realise grammatical knowledge accurately and meaningfully in test-taking or other language-use contexts” (Purpura, 2004, p.89). Grammatical knowledge embraces the understanding or creating of formally accurate utterances or sentences (Bachman & Palmer, 2010). It includes knowledge of vocabulary, syntax, phonology, and graphology. This section reviews which kinds of grammatical knowledge were tested in the UEE.

5.3.1.1 Vocabulary

Assessing *knowledge of vocabulary* of second language learners is considered to be “both necessary and reasonably straightforward” (Read, 2000, p.1). Learners should be assessed on the sufficiency of their *knowledge of vocabulary* for meeting their communication needs. Vocabulary should be assessed to take in a broader range of “lexical assessment procedures” (p.8). When analysing the UEE papers, three dimensions – discrete/embedded, selective/comprehensive, and context independent/dependent – were reviewed for vocabulary test items.

In the listening comprehension section, test-takers had to understand the meaning of particular words from the conversations or passages. The vocabulary measured in most of the items and test tasks of this section was embedded, comprehensive, and context-dependent. However, there were a few exceptions – some items were tested explicitly; that is, selectively. Test-takers were required to understand the main idea of the passage, and also understand which option provided the most suitable word for the answer.

2012: Listening Comprehension – Section B

2012:13 The passage was about an open management style run by a man;

Q: “What does the speaker consider important in running a small company?”

Answer: A. reward; B. safety; C. trust; D. honesty.

In the grammar and vocabulary section, Section A (grammar) mainly focused on syntax. Section B (vocabulary) focused on vocabulary itself. In this latter section from 2005 to 2007, the vocabulary tested was discrete, selective and context-dependent. It may seem self-evident that tests of this sort were measures of vocabulary knowledge.

2005: Grammar and Vocabulary – Section A

2005:41 There was such a long queue for coffee at the interval that we _____ gave up. Answer: A. eventually; B. unfortunately; C. generously; D. purposefully.

From the sentence, the phrases – ‘a long queue’ and ‘gave up’ provided test-takers with the context. Also by looking at the meanings of different options, only option A matched the context and meaning of the sentence. This item tested test-takers’ understanding of vocabulary knowledge rather than other abilities.

From 2008, the vocabulary section became a gap-filling passage with some words provided to choose from. The patterns of the test items were moved from the discrete items to more integrated skills and the test items became embedded, selective and context-dependent. Test-takers were required not only to understand individual sentences but also the meaning of the context when they filled in the gap in the passage.

2012: Grammar and Vocabulary – Section B

2012:47-49 ... Government 47 suggest that two thirds of adults and a third of children are overweight. If trends are not 48, this could rise to almost nine in ten adults and two thirds of children by 2050, putting them at 49 risk of heart disease, cancer and other diseases. ...

Answer: A. maintained; B. serious; C. indications; D. figures; E. anxious; F. concern; G. crisis; H. decided; I. available; J. reversed.

In this section, test-takers were tested on vocabulary in particular. Moreover, they were required to understand not only the meaning of the options, but also the context of the passage.

In the reading comprehension section, tests of vocabulary were embedded in reading tasks involving a written text with a set of comprehension questions. Such test items were a very common way of assessing the learners’ understanding of particular words or phrases in the text; that is, a few question items focused on testing vocabulary explicitly and students were required to find out the meaning of the terms or phrases by guessing the meanings in the context.

2012: Reading Comprehension – Section B (Passage A)

2012:66 What does the word “epic” in paragraph one most probably mean?

Answer: A. broke the world record; B. collected money for Oxfam;
C. destroyed several bikes; D. travelled about 1300 hours.

Test-takers were required to understand the vocabulary to comprehend the reading test tasks. In sum, the test of vocabulary was embedded, comprehensive and context-dependent. However, there were a few items in the UEE measuring specific items of vocabulary explicitly. There were two parts in the writing section: translation and guided writing. For Part I – Translation, five Chinese sentences had to be translated into English sentences, with key words provided. Vocabulary knowledge was measured as an independent construct in the translation section; test-takers needed adequate vocabulary to translate appropriately. One of the scoring criteria in this part was the use of vocabulary. The vocabulary assessment was discrete, selective, and context-dependent.

2012: Translation – Item 3

The Chinese sentence: 每位设计师都希望自己的作品能经受时间的考验。(stand)

Suggested answer: Every designer hopes that his work can *stand* the test of time.

For the translation section in the past 10 years, there was no change in terms of the patterns and format. The main purpose of this section is to test the use of vocabulary and grammatical knowledge, and the correct sentence structure. The probable reason for this lack of change is that translation tasks have been considered to be an effective tool to test the application of using vocabulary.

The test-tasks in Part II – Guided Writing required test-takers to write a diary (2012), or a letter (2013) or an email (2014). This part tested test-takers in three ways: 1: content and coherence; 2: use of vocabulary and grammar, and 3: composition structure and organisation. They had to use appropriate vocabulary to express their ideas. One of the significant marking criteria for this part was language use including grammatical structures and use of vocabulary. The vocabulary measure was embedded, comprehensive and context-dependent.

2013: Guided Writing

Write an English composition in 120-150 words according to the instructions given in Chinese: Students were required to write a diary after having observed an elementary school class from the perspective of a little girl, Amy, who had had the two painting lessons on the preceding Monday (two pictures were given). The content of the diary was to include: 1: a description of two painting lessons, and 2: how they had been inspired by the lessons.

5.3.1.2 Syntax

Syntax is defined as the grammatical arrangement of words in a sentence. An assessment of *knowledge of syntax* was embedded in most of the multiple-choice question items (listening comprehension, vocabulary, and reading comprehension sections) while in the grammar and writing sections syntactical knowledge was tested more explicitly.

In Listening Comprehension, short conversations in Section A and passages in Section B were multiple-choice items while longer conversations in Section C were in a table-filling format; test-takers were offered a limited choice of words from which to complete the table.

2007: Listening Comprehension – Section C

2007:20 The conversation was about booking a table for dinner at a Garden restaurant.

M: We'll be pretty busy on Friday night. I have a table for nine o'clock, but not at eight thirty.

W: Nine o'clock will be fine. I have a friend from Italy. Have you got any Italian food?

Garden Restaurant Reservation Form
Special Request: <u>20</u> dishes.

The woman had a special request for food – Italian dishes. Syntax knowledge was assessed in that an adjective form was required. If test-takers wrote Italy, no mark would be given.

2011: Listening Comprehension – Section C

2011:24 The conversation was about three types of memories. When the speakers talked about long-term memory, the listener supposed that permanent memory was meant; however, in the table, it was a sentence with the given word with 'It' – It is permanent.

Test-takers were required to complete the sentence in the table. It tested not only listening skill, but also the grammatical structure of a sentence.

How is long term memory different from the others?	It <u>24</u> _____.
----------------------------------------------------	---------------------

In the grammar section, most of the questions tested knowledge of syntax only. Most of the questions were in only one sentence. Test-takers were required to understand the grammatical arrangement of words in the sentences so that the correct answer could be chosen from the multiple-choice options. A couple of examples are given below:

2007: Grammar and Vocabulary – Section A

2007:25 Leaves are found on all kinds of tress, but they differ greatly ____ size and shape.

Answer: A. on; B. from; C. by; D. in.

In this item, test-takers were tested on preposition use.

2009: Grammar and Vocabulary – Section A

2009:29 Mary went to the box office at lunch time, but all the tickets ____ out.

Answer: A. would sell; B. had sold; C. have sold; D. was selling.

In this item, test-takers were tested on tense use.

Since 2010, the last passage in Reading Comprehension became short-answer format from multiple-choice format; test-takers were restricted to a set limit of words in answering the questions or completing the statements. These kinds of items involved writing skill, as the words from the passage had to be rephrased or rearranged in order to fit within the limit. If there were any grammatical errors in the written answers, marks would be deducted.

In Paper II – Writing (translation and guided writing), test-takers were required to write the sentences in correct grammar; use of grammar (syntax) was one of the scoring criteria.

2012: Translation – Item 1

The Chinese sentence: 她五年前开始拉小提琴。(play)

Suggested answer: She began to *play* the violin five years ago.

For this item, grammar was tested explicitly; for example, an ‘s’ was needed for five years because of the plural form and ‘began’ had to be used instead of ‘begin’ because the sentence was in the past tense.

Grammar is about the language use in the test tasks. It is one of the major criteria of marking standards in both translation and guided writing sections (See Appendix D for marking standards). If any grammatical errors are found in the answers, then marks will be deducted.

5.3.1.3 Phonology or graphology

Phonology is the study of sound and pronunciations in a particular language and *graphology* is the study of written and printed symbols (Buck, 2001). *Knowledge of phonology and graphology* facilitates the recognition and generation of the features of the sound or of a writing system so as to be able to convey meaning, and it also includes sound-spelling correspondences (Purpura, 2004).

Listening comprehension is partially a process of recognising language sounds (Lado, 1961). The basic testing technique is to check if students can understand a complete utterance or crucial parts of it. In the listening comprehension section, test-takers were required to understand the pronunciations of the spoken texts; all 24 question-items in the listening section were related to the knowledge of phonology.

In the UEE, words were required to complete the forms or tables in longer conversations; test-takers had to recognise the sound (the pronunciation) from the audio and then write down the correct spelling of the words in the table.

2013: Listening Comprehension – Section C

2013:17-20

Latest Conference Information	
Date:	8 th <u>17</u>
Place:	Palace <u>18</u> , Shanghai
Registration fee:	\$ <u>19</u>
Speaker:	Carla Marisco from Milan University
Speech Topic:	Opportunities and Risks in the <u>20</u> Market

For short-answer items in the reading comprehension, and both translation and guided writing in the writing section, correct spelling was required; these items assessed the graphological knowledge as well.

5.3.2 Textual knowledge

Textual knowledge is involved in comprehending or producing “the sequence of units of information in texts” (Bachman & Palmer, 2010, p.45). Those units are units of language – spoken and written – that consist of two or more utterances or sentences. Textual knowledge tells the student how to use grammatical knowledge to develop texts coherently (Weigle, 2002). There are two areas: knowledge of cohesion and knowledge of rhetorical or conversational organisation. This section reviews which of the textual knowledge was tested in the UEE.

5.3.2.1 Cohesion

Knowledge of cohesion is engaged in comprehending or producing a relationship among sentences in written texts or utterances in conversations (Bachman & Palmer, 2010). Cohesive texts are sequences of sentences or utterances which seem to “hang together” (McCarthy, 1991, p.26). These words and phrases allow the writer or speaker to establish connections across sentence or utterance limitations and help to link the sentences in a text or utterances in conversation together.

Nunan (1993, p.21) states that “the most comprehensive description and analysis” of cohesive devices are to be found in Halliday and Hasan (1976), which identified five (refined to four in Halliday, 1985) different types of cohesion: reference, substitution and ellipsis, conjunction, and lexical cohesion. Knowledge of cohesion was tested in most sections in the UEE papers except the grammar section (in which the items were discrete and mainly assessed grammatical knowledge). Test-takers without any cohesive knowledge were still able to answer the grammar section.

Table 36: Cohesion Tested in the Past 10 years’ UEE Papers

Cohesion	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Average
Reference	19.4%	9.7%	13.8%	23.5%	14.3%	13.2%	26.1%	9.0%	14.1%	6.3%	14.9%
Substitution and Ellipsis	4.8%	9.7%	4.6%	1.5%	3.2%	4.4%	2.9%	7.5%	1.6%	0.0%	4.0%
Conjunction	17.7%	12.9%	13.8%	8.8%	14.3%	16.2%	8.7%	11.9%	14.1%	14.3%	13.3%
Lexical cohesion	58.1%	67.7%	67.7%	66.2%	68.3%	66.2%	62.3%	71.6%	70.3%	79.4%	67.8%

By using Halliday and Hasan's (1976) model of cohesion, the question-items were analysed and checked to establish which cohesive ties were involved. From Table 36, on average, for the past 10 years' UEE papers, reference, substitution and ellipsis, conjunction, and lexical cohesion accounted for 14.9 percent, 4.0 percent, 13.3 percent and 67.8 percent of the paper respectively. Lexical cohesion plays a significant role among the question-items in terms of testing knowledge of cohesion, which matches with Halliday and Hasan's (1976) claim, "Cohesion ... is achieved through the association of lexical items that regularly co-occur" (p.284). Lexical cohesion includes reiteration and collocation – reiteration involves repetition, synonym, or near synonym, superordinate, and general words, whereas collocation involves those lexical items that are semantically related (Nunan, 1993).

From Table 36, the general trend of using substitution and ellipsis was decreasing whilst that of lexical cohesion was increasing for the cohesion tested in the past 10 years' UEE papers. Lexical cohesion deals with the meaning in text in which the cohesive effect is achieved by the usage of vocabulary (Halliday & Hasan, 1976). From the grammar and vocabulary section in the past 10 years' UEE papers, it can be noticed that the test items were changed from discrete items to more integrated passage items and most of the cohesive ties are lexical. Lexical cohesion became more important, especially when test-takers were required to find the answers by speculating on the passages and understanding the context, because lexical cohesion concerns the situation in which lexical items relate to each other so that textual continuity is formed. In the reading comprehension section, lexical cohesion (including reiteration and collocation) plays a significant role in creating cohesion and coherence in any given text. When test-takers would like to comprehend the whole passage, they were required to link up the relationship between the words in the passage and then they would be able to understand the idea of the paragraph or passage when they read through the passage.

The following examples are listed to show the ways in which the different cohesive ties (reference, substitution and ellipsis, conjunction, and lexical cohesion) were involved and assessed in different question items.

2013: Listening Comprehension – Section A

2013:3 M: Hi, Grace. Tell me something about your hometown.

W: It's so beautiful and peaceful. But it's really far away from everything.

Q: What does the woman think of her hometown?

Answer: A. Promising; B. Isolated; C. Crowded; D. Modern.

In this sentence, test-takers were required to understand what 'it' represents; it means 'the woman's hometown', so that they knew that the interlocutors were talking about the woman's hometown. This item tested knowledge of cohesion – *reference (personal)*.

2008: Grammar and Vocabulary – Section A

2008:26 Ann: Do you want tea or coffee?

Ali: _____. I really don't mind.

Answer: A. None; B. Neither; C. Either; D. All.

The answer is either, which means tea or coffee were both fine for him. This item tested knowledge of cohesion – *substitution (nominal)*;

2014: Reading Comprehension – Section A

2014:57 ... We don't spend two-thirds of our time gossiping just because we can talk, argues Dunbar – 57, he goes on to say, language evolved specifically to allow us to gossip. ...

Answer: A. for instance; B. in addition; C. on the contrary; D. as a result.

The answer is C. on the contrary. This item tested knowledge of cohesion – *conjunction (adversative)*.

2014: Grammar and Vocabulary – Section B

2014:48 ... A green light 48 that the amounts of the three nutrients are healthy; yellow indicates that the customer should be alert; and red means that the food is high in at least one of the three nutrients.

Answer: A. alert; B. classify; C. commit; D. delicately; E. gentle; F. impose; G. labels; H. moderation; relieve; J. signals; K. simply.

The answer 'signals' has a similar meaning (synonym) as 'indicates and means', and these three words have the same form of verb. This item tested knowledge of cohesion – *lexical cohesion (reiteration)*. Meanwhile, the words 'light' and 'signal' are related. 'Signal' is a movement or sound or wave that gives information or a message, whereas 'light' is brightness that comes from electrical devices. Therefore, this item also tested knowledge of cohesion (collocation).

5.3.2.2 Rhetorical or conversational organisation

Knowledge of rhetorical organisation involves being aware of organisational development of information in written texts, for example narrative, descriptive, and argumentative strategies, whereas *knowledge of conversational organisation* involves being aware of the techniques that interlocutors use in conversation, for example, adjacency pairs, turn-taking, and repair (Bachman & Palmer, 2010).

In the UEE, guided writing in the writing section involves assessing knowledge of rhetorical organisation. Test-takers were required to write an English composition in 120-150 words according to the instructions given in the UEE. The different topics and types of writing required for guided writing in the past 10 years' UEE papers are listed in Table 37.

Table 37: Topics and Text Types of Writing Tested in Guided Writing

Year	Topic	Text type of writing
2005	A use for my talent	Narrative
2006	Making a choice between two favourite universities	Compare and contrast
2007	A gift	Descriptive
2008	Group dancing competition	Expressing opinions
2009	Summer camp	Letter writing – expressing interests
2010	Our primary school life	Describe a picture – expressing feelings
2011	Fundraising	Letter writing – expressing ideas
2012	Painting lessons	Diary writing
2013	An art exhibition	Expressing opinions
2014	School newspaper	Email to the editor

Organisation and structure was one of the marking standards for guided writing. The total mark in the guided writing section was 25: 10 marks for content, 10 for vocabulary and grammar, and 5 for organisation and structure (See Appendix D). In analysing the guided writing of past papers, test-takers were required to express clearly what information or content they would like to write in rhetorical style in the guided writing section. In this instance (as elsewhere) the particular constructs for testing writing ability as intended by the UEE test designers were accurately reflected in the marking standards. Thus the marking standards for organisation and structure reflected what the test designers wanted to measure in the UEE; that is construct validity.

From the past 10 years' UEE papers, it can be seen that the text type of the writing topics has been changed. For the first few years, the topics were focused on narrative and descriptive types of writing. However, for the recent years, test-takers were asked to express their opinions or ideas on some issues in the letter or email writing. The tasks became more communicative between test-takers and some organisations.

Table 38: Text Types Tested in Reading Comprehension

Text type / nos. of passage	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Narrative	2	1	1	1	2	1	1	1	1	1
Expository	2	1	-	1	2	-	1	1	2	1
Argumentative	1	2	2	2	1	3	3	3	2	2
Journal article	-	-	-	-	-	1	-	-	-	-
Picture description	-	1	1	-	-	-	-	-	-	-
Interview / Q & A	-	-	1	1	-	-	-	-	-	-

Although knowledge of rhetorical organisation was not directly tested in the reading comprehension section, test-takers were required to comprehend different types of text in different passages. Table 38 shows different types of text passages tested each year in Reading Comprehension for the past 10 years' UEE papers.

When looking at the text type of the passages, it is interesting that the passages were more on narrative and expository but not argumentative. However, from 2010, the passages had more emphasis on argumentative text type. Narrative, descriptive and expository types of writing are the most fundamental for English learners (Morin, 2015). Students will read more argumentative articles when they study at university. It can explain why more argumentative passages were the focus in the UEE.

Table 39: Text Types Tested in Listening Comprehension (Section B)

Text type / nos. of passage	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Narrative	1	1	1	1	1	1	1	1	1	1
Expository	-	1	1	1	1	1	1	1	1	-
News report	1	-	-	-	-	-	-	-	-	1

Similarly, knowledge of rhetorical organisation was not directly tested in Section B of Listening Comprehension. In Section B, there were two passages – one was a narrative passage and the other was either expository or a news report (See Table 39). Test-takers were required to comprehend different types of text in different passages.

5.3.3 Functional knowledge

Testing of second language pragmatic knowledge, which includes functional and sociolinguistic knowledge, is still under exploration and is a growing area of second language assessment (Roever, 2011). Functional knowledge allows people to understand the relationships between utterances or sentences and texts and the meanings of language users (Bachman & Palmer, 2010). It is about how to use language properly to accomplish various communication functions (Weigle, 2002). It comprises knowledge of four functional sets: ideational, manipulative, instrumental, and imaginative. This section reviews what kind of functional knowledge was tested in the UEE.

5.3.3.1 Ideational functions

Knowledge of ideational functions pertains to expression or interpretation in terms of real-world experience (Bachman & Palmer, 2010). These functions include using language to express meaning, or exchange information about concepts or emotions. Utterances, such as descriptions and explanations of emotion or anger, perform ideational functions. Basically, all items in the UEE papers tested knowledge of ideational functions. Test-takers were required to use language to understand the information and convey the meaning, and they were then able to answer the question items by expressing their knowledge and ideas.

All items in the listening comprehension section involved the knowledge of ideational functions. Test-takers were required to speculate on the implied meaning from the conversations in addition to the literal meaning. Such question-items not only tested test-takers' listening ability, but also their skill at deducing implied meaning or ideational function knowledge from the conversations.

2005: Listening Comprehension – Section A

2005:6 M: During the summer vacation, I'll earn some money. How about you?

W: I am going to take an online course so I can graduate sooner.

Q: What are the two speakers mainly talking about?

Answer: A. job hunting; B. an online course; C. earlier graduation; D. summer vacation plans.

This item shows the speakers exchanging information about their summer vacation plans.

2007: Listening Comprehension – Section A

2007:3 M: I just read my blood test report, and everything seems OK.

W: I see. Then there is nothing to worry about.

Q: How does the woman probably feel now?

Answer: A. Relieved; B. Worried; C. Confused; D. Depressed.

This item shows that the woman was expressing her feelings about the man's health.

Similar to the listening comprehension section, items in the UEE involving knowledge of ideational functions were those questions in the reading comprehension section that tested the implied or inferred meaning passages which contained the information about ideas, opinions, and knowledge. Test-takers were required to understand those kinds of functional knowledge hidden in the passage.

All tasks in the writing section, including both translation and guided writing, also tested knowledge of ideational functions. For example, the translation tasks required test-takers to understand the meaning of the Chinese language and then, by using correct language use and sentence structure, to translate those sentences into English. Another example is that the guided writing task in 2013 asked test-takers to write a letter to the Shanghai Museum, which was organising an Art Exhibition, to express their opinions as to whether the exhibition should be held at the museum or at a community library. Knowledge of ideational functions was involved in such writing tasks.

5.3.3.2 Manipulative functions

Knowledge of manipulative functions allows language users to express and use language to carry out some actions and influence the world around them (Bachman & Palmer, 2010).

This includes knowledge of:

1. Instructional functions, which are performed to get actions completed (including requests, suggestions, commands, and warnings);
2. Regulatory functions, which are used to determine others' behaviour (including rules, regulations, and laws);
3. Interpersonal functions, which are used for starting, continuing, and altering interpersonal relationships (including greetings, compliments, and apologies).

In the listening section, some items tested the knowledge of manipulative functions. Test-takers were tested on whether they understood how language use would affect situations. These kinds of language functions always happen in daily life. The following examples show two items testing knowledge of manipulative functions in the UEE.

2005: Listening Comprehension – Section A

2005:9 M: Every time I call the number I get a busy signal!

W: Wait an hour and try again.

Q: What does the woman suggest the man do?

Answer: A. Phone later; B. Try harder; C. Wait for a signal; D. Check the number.

In this item, the woman suggested the man make the phone call later, which was an example of instructional function – giving suggestions.

2012: Listening Comprehension – Section A

2012:5 M: Shall I come and take you to the railway station?

W: No, thanks, I'll manage. It's not far anyway.

Q: What can we learn about the woman?

Answer: A. She lives close to the man. B. She changes her mind at last.

C. She will turn to her manager. D. She declines the man's offer.

In this conversation, the woman declined the man's offer, explaining that she could get to the railway station and avoid inconveniencing her friend. Her action was appropriate: it was polite to reject his offer but at the same time to maintain their interpersonal relationship.

The test task in the guided writing section in 2014 asked test-takers to write to the editor to express their suggestions on the school English newspaper, which was going to be made over. It was intended to remove one of the existing columns (health, entertainment, culture), and add one of the options (fashion, career plan, readers' feedback) to replace the removed option. Test-takers were required to write an email to the editor to express their ideas. Similarly, test tasks in the guided writing section in 2009, 2011, and 2012 also required test-takers to write a letter to apply for a programme in a summer camp, to apply for some funding to help poor children in remote areas, and to suggest the exhibition should be at the Shanghai Museum respectively, which involved testing test-takers' knowledge of manipulative functions.

5.3.3.3 Heuristic functions

Knowledge of heuristic functions is related to the use of language to broaden people's horizons in the world; it is using language as a tool for instruction and studying, for problem-solving, and for memorising purposes (Bachman & Palmer, 2010). This kind of functional knowledge was involved in the UEE in only a very limited way. Only two items appeared for the past 10 years' UEE papers using language for teaching and learning and problem solving; they were all in the listening comprehension section.

2006: Listening Comprehension – Section A

2006:1 W: Were you here on March 5th?

M: Mm, not really. In fact I arrived three days later.

Q: When did the woman arrive?

Answer: A. On March 2. B. On March 3. C. On March 5. D. On March 8.

2011: Listening Comprehension – Section A

2011:2 W: I wonder if there is a service charge for our meal.

M: I think so. The menu said the service charge is 10 percent.

Q: How much is the service charge if the food costs 50 dollars?

Answer: A. \$5. B. \$10. C. \$15. D. \$50.

From the above items, before the problem or calculation was solved, test-takers were required to understand and use the language as a tool to solve the problem, in which these items contain knowledge of heuristic functions. Indeed, the question-items were quite sophisticated and interesting in that simple mathematics calculations were involved. Test-takers might not be able to get a right answer if they were very poor at mathematics.

5.3.3.4 Imaginative functions

Knowledge of imaginative functions allows people to use their language to create an environment for “humorous or aesthetic purposes”; examples include telling jokes, making metaphors or figurative use of language, and attending theatres (Bachman & Palmer, 2010, p.47). No question-item was found in the listening comprehension, grammar and vocabulary, or reading comprehension sections related to knowledge of imaginative functions. In the writing section, no topics involved the knowledge of imaginative functions for the past 10 years' UEE papers as well.

In sum, out of these four functions, ideational function was tested in all items and manipulative functions were tested only in the listening comprehension and writing sections for the past 10 years' UEE papers.

5.3.4 Sociolinguistic knowledge

Language users having sociolinguistic knowledge are able to control or interpret language in an appropriate language-use context (Bachman & Palmer, 2010). Sociolinguistic knowledge also takes in the knowledge of how to apply language properly in different social contexts (Weigle, 2002). Sociolinguistic knowledge includes knowing the appropriate use of dialects and varieties, registers, natural or idiomatic expressions, cultural references, figures of speech, and genres. This section investigates whether any sociolinguistic knowledge was tested in the UEE.

5.3.4.1 Dialects and varieties

Language variations are linked with language users in different geographic regions. *Knowledge of dialects and varieties* involves being aware the characteristics of such varieties of language (Bachman & Palmer, 2010).

The short conversation in the 2010 UEE's Listening Comprehension section involved knowledge of varieties. In the utterance, one of the speakers used the term **football**, which is British English, whereas the option's answer **soccer** is American English. The question tested test-takers' understanding the meanings of football and soccer, called 'varieties' in sociolinguistics. This was the only question involving 'varieties' in the UEE papers.

2010: Listening Comprehension – Section A

2010:2 M: Be quiet! I am trying to watch our **football** team in the World Cup!

 W: So what's going on? What's the score?

 Q: What are the speakers talking about?

Answer: A. The exam score; B. The world news; C. A **soccer** match; D. A basketball match.

In the researcher's opinion, from this item, test-takers were tested on vocabulary rather than so-called knowledge of dialect and varieties.

5.3.4.2 Registers

In linguistics, *registers* denote the varieties of language used for a particular purpose or in a particular social setting. *Knowledge of registers* includes the characteristics of different levels

of formality in language use (Bachman & Palmer, 2010). When speaking in a formal setting contrary to an informal setting, a speaker may be more likely to use features of prescribed grammar with formal words rather than informal words. Variations in register occur in both speaking and writing.

This knowledge of registers was hardly tested in any sections of the past 10 years' UEE papers. It was tested in the writing section when test-takers were asked to write a letter to a school about a summer camp programme (2009), or to an international children's fund about fund-raising programme (2011), or send an email to the editorial page of a school newspaper (2014). An appropriate register for the letter or email was essential; test-takers had to pay attention to when test-takers wrote the letter.

5.3.4.3 Natural or idiomatic expressions

Knowledge of natural or idiomatic expressions permits language users to not only speak accurately or write correct text linguistically, but also do so as a native speaker (Bachman & Palmer, 2010). Test-takers in China were learning English as a foreign language. Testing knowledge of natural or idiomatic expressions would be difficult for non-English native speakers. For the past 10 years' UEE papers, only two question-items involved this kind of expression, which appeared only in the reading comprehension section.

2006: Reading Comprehension – Section B

2006:69 The expression 'tip the balance' in paragraph 1 (From the passage: If they are close to confessing a crime, the blue on the wall might tip the balance.) probably indicates that the blue might _____.

Answer: A. let suspects keep their balance.

- B. help suspects to confess their crimes.
- C. make suspects cold and unfriendly in law court.
- D. enable suspects to change their attitudes to colours.

2008: Reading Comprehension – Section B

2008:71 Which of the following is closest to the main idea of the passage?

- Answer: A. A friend in need is a friend indeed. B. Where there is a will, there is a way.
- C. A misfortune may turn out a blessing. D. Kill two birds with one stone.

In the first item, test-takers were asked to describe the meaning of the expression ‘tip the balance’ while the second one used an idiomatic expression to summarise the main idea of the passage after it had been read. These two question-items tested test-takers knowledge of idiomatic expressions in addition to comprehending the passages by using grammatical knowledge. Nevertheless, assessing this kind of knowledge was very rarely done in the UEE.

5.3.4.4 Cultural references and figures of speech

Knowledge of cultural references allows users to use and interpret particular events, places, or people referred to whereas *knowledge of figures of speech* includes figurative language such as metaphors and similes (Bachman & Palmer, 2010). Language users have to have particular cultural background knowledge to comprehend when particular events are mentioned in their utterance or texts. For the past 10 years’ UEE papers, this kind of knowledge was not tested.

5.3.4.5 Genres

Knowledge of genres involves knowledge of the conventions that shape communicative actions for particular social purposes (Bachman & Palmer, 2010). Genre provides information related to writing and helps with better writing of different kinds of texts. By having knowledge of genres, which include writing purpose, text organisation and structure, and language features, test-takers will understand how to write an essay well.

The knowledge of genres is tested in two ways: receptive and productive. In respect of receptive knowledge, listening to the passages in Listening Comprehension and reading different articles in Reading Comprehension involve the knowledge of genres. Test-takers were required to understand the meaning of the passages through the text type and its organisation and structure.

In respect of productive knowledge, test-takers were required to write different kinds of text types in the writing section (See Table 37). They were also required to understand the writing purpose of the essay and know how to express their essay in an appropriate text structure and use correct language features. Similarly, in speaking, appropriate genres are required when test-takers express their opinions and describe the pictures in the Oral English Test.

5.3.5 Strategic Competence

Strategic competence consists of *metacognitive strategies*, which are executive processes that allow language users to engage in *goal-setting*, *appraising*, and *planning* (Bachman & Palmer, 2010). Strategic competence involves deciding what test-takers are going to perform (goal-setting), appraising the situation and their resources to deal with it (appraising), and “deciding how to use the resources” they have (planning) (Luoma, 2004, p.99).

- *Goal-setting* involves identifying different language-use tasks or test tasks, and also choosing if options are given (that is, one or more tasks from a set of possible tasks), and deciding whether or not to attempt to complete the tasks in the UEE. The goal of taking this UEE was to finish the question-items and test tasks correctly.
- *Appraising* involves test-takers selecting what competence is needed to finish the tasks. It involves assessing the characteristics of the language-use or tests, assessing the individual's own knowledge (topical and language), and assessing the accuracy of the answers to the test tasks, which is related to the grammatical, textual, functional, and sociolinguistic knowledge of the response.
- *Planning* involves determining how to incorporate language, topic knowledge, and affective schemata in order to finish the test tasks successfully.

In the following, examples have been given to show how strategic competence was involved in different sections of the UEE.

Strategic competence includes the “cognitive strategies” and “metacognitive strategies” that fulfil the cognitive management function in listening (Buck, 2001, p.103). This is the ability to use language competence and metacognitive strategies in the management of language performance. In the listening section, test-takers were required to listen to longer conversations and then fill in a table. Test-takers assessed what sort of knowledge including topical and language knowledge they needed, and assessed what sort of language-use task they needed to finish. In this case, it was a table-completion task; test-takers were required to have accurate grammar and correct spelling of the vocabulary used in the conversations. They also had to decide how much topical knowledge and language knowledge was needed to complete the tasks. Once test-takers identified all these related elements, they formulated strategic plans and selected the best one mentally to respond and finish the task.

Metacognitive strategies in communication situations are essential for language learners because they need to overcome limitations in their vocabulary (Read, 2000). In the grammar and vocabulary section, there was a passage with 10 blanks and some words were provided to choose from for gap filling. The goal of this section was to fill in the blanks using the words provided in the box. Test-takers assessed whether or not they had topical knowledge about the passage and understood the meaning of the words. By using grammatical and textual knowledge, test-takers assessed the characteristics of the language use and tried to complete the task. Test-takers executed the plan and selection of appropriate topical and language knowledge and implemented these elements to the task. If test-takers were not sure about the answers, they might have to adopt other strategies, for example, using the grammatical structure of the sentence to determine the form of the blank (adverb, adjective, verb, or noun) in order to increase the chance of picking up the correct word from the box.

Strategic competence in reading allows learners to adapt the style and speed of reading to different texts and purposes and also to read flexibly by using appropriate reference sources selectively (Alderson, 2000). In the reading comprehension section, there were some questions about the meaning of words. The goal of this kind of item was to find the meaning of the highlighted word in the passage. Test-takers probably assessed how much language knowledge or topical knowledge they had, and they would have to acquire this kind of knowledge before they could understand the passage. When test-takers were not sure about the answers, they would try to guess the words by using the context in addition to their topic knowledge, and affective schemata in order to finish the test tasks successfully.

A writing task involves not just language competence, but also strategic competence. Potential metacognitive strategies involved in the writing task would be: 1: goal-setting, including what one tries to accomplish with this writing; 2: appraisal of the task's various facets, including the rhetorical situation, one's own linguistic resources for completing the task, and one's success in completing the task, and 3: planning how to complete the task. Test-takers set up a goal of writing and they had to assess various aspects of the task, including the main theme of the essay, rhetorical organisation, and the nature of the reader. The task also assessed test-takers' knowledge of topic and language, and the accuracy of answers to the test tasks, which related to the grammatical, textual, functional, and sociolinguistics knowledge. Finally test-takers would need to plan how to complete the task.

As mentioned in Section 4.4.4, the metacognitive strategy objectives in the English Curriculum Standards cover mainly strategies needed for the proactive management of learning in general terms. These metacognitive strategies listed in the ECS do not include task-specific metacognitive strategies such as planning, monitoring, and evaluating in the process of task completion. For example, in taking the UEE paper, the time limit was two hours, and test-takers would need to allocate their time for different sections and make a judgement about how much time they had to spend in those sections. These metacognitive strategies, or the managerial processes that permit test-takers to engage in goal-setting, appraising, and planning, were very much part and parcel of the strategic competence being tested, even if they are not listed in the English Curriculum Standards.

5.3.6 Oral English Test

The Shanghai Senior Secondary School Oral English Test was conducted separately from the written exam and the score was not included in the total score for the University Entrance Exam. The test was conducted in a laboratory with the aid of computers. Test-takers had to answer the questions communicated via the screen and earphones through the microphone. The test was divided into five parts.

1. Part 1: Short essay reading – being presented with a short passage on the screen, test-takers had one minute for preparation and 30 seconds for reading. This part tested test-takers' pronunciation of a text.
2. Part 2: Quick responding – Test-takers were required to give quick responses on the basis of given situations. The students listened to five recordings and had to answer each question appropriately within four seconds. This part measured test-takers' ability to comprehend and use functional language.
3. Part 3: Question raising – test-takers had to raise at least two questions for each of two given situations. Each situation was presented in 20 seconds and each question was allocated 10 seconds. This part tested test-takers' ability in raising questions based on different information in given situations.
4. Part 4: Topic talking – test-takers were asked to provide their own points of view for a given topic. The topic may be on life-related or topical fields such as society, family, school, environmental protection, and the Internet. There was one minute for preparation and another minute for talking on the given topic.

5. Part 5: Picture describing – test-takers had one minute for preparation and another minute for describing the picture on the screen. The description should contain at least six sentences. Other than grammatical knowledge, textual knowledge including cohesion and conversational organisation was involved in this part.

Phonology, which is about sound and pronunciation, was the main element in the Oral English Test. Knowledge of phonological form enables interlocutors to understand and produce features of the sound system to convey meaning in conversational situations. Knowledge of vocabulary and syntax, that is, grammatical knowledge, was also involved when test-takers discussed topics and described pictures. In addition, the cohesive devices and coherent conversational organisation, that is textual knowledge, used in the test tasks form part of the test task assessment.

Functional knowledge, which was involved in the Oral English Test, enables interlocutors to interpret relationships between utterances. Functional knowledge includes ideational and manipulative functions, which were tested in Part 4, and heuristic and imaginative functions which were tested in Part 5. On the other hand, language users having sociolinguistic knowledge are able to understand language in different social contexts. However, only knowledge of genres was tested in the Oral English Test. Test-takers were asked to give their opinions on some topics (in Part 4) and descriptive text type was used when they were required to describe the picture (in Part 5) on the computer screen.

Language knowledge is more “componential and static”, whereas strategic competence is more “active and dynamic” (Luoma, 2004, p.99). Three metacognitive components – goal-setting, appraising, and planning – are recognised under strategic competence. These components were covered in the Oral English Test. In all five parts the goal was to accomplish different test tasks (*goal-setting*). Test-takers were required to evaluate the situation and what language knowledge and topical knowledge they needed to deal with the test tasks (*appraising*). For example, for picture description, test-takers had to organise the conversational situations by considering the language knowledge related to the grammatical, textual, functional, and sociolinguistics knowledge of the response in addition to their topical knowledge. Finally, test-takers had to decide on the language and how they would use it in order to complete language-use tasks successfully (*planning*).

5.4 Conclusion

Many of the language assessments that have been developed assess only one or a few specific areas of language knowledge (Bachman & Palmer, 2010). Table 40 shows the summary of the findings in the past 10 years' UEE papers. The results confirm Bachman and Palmer: organisational knowledge (grammatical and textual) is the focal element tested.

Table 40: Summary of the Checklist – Items Tested in the UEE

Component of Language Ability	Listening Comprehension	Grammar & Vocabulary	Reading Comprehension	Writing	Speaking
Grammatical knowledge					
<i>Vocabulary</i>	required	required	required	required	required
<i>Syntax</i>	required	required	required	required	required
<i>Phonology and graphology</i>	required	required	required	required	required
Textual knowledge					
<i>Cohesion</i>	required	required	required	required	required
<i>Rhetorical or conversational organisation</i>	required	not required	required	required	required
Functional knowledge					
<i>Ideational functions</i>	required	required	required	required	required
<i>Manipulative functions</i>	required	not required	not required	required	required
<i>Heuristic functions</i>	marginally	not required	not required	not required	required
<i>Imaginative functions</i>	not required	not required	not required	not required	required
Sociolinguistic knowledge					
<i>Dialects and varieties</i>	marginally	not required	not required	not required	not required
<i>Registers</i>	not required	not required	not required	required	not required
<i>Natural or idiomatic expressions</i>	not required	not required	marginally	not required	not required
<i>Cultural references and figures of speech</i>	not required	not required	not required	not required	not required
<i>Genres</i>	required	not required	required	required	required
Strategic competence					
<i>Goal-setting</i>	required	required	required	required	required
<i>Appraising</i>	required	required	required	required	required
<i>Planning</i>	required	required	required	required	required

In respect of grammatical knowledge, knowledge of vocabulary, syntax, phonology and graphology were all required in the different sections in the UEE, viz., listening comprehension, grammar and vocabulary, reading comprehension, writing, and speaking. Grammatical knowledge was considered the main ingredient tested in the UEE English language examination.

In respect of textual knowledge, knowledge of cohesion, and rhetorical or conversational organisation were required in nearly all sections, except knowledge of rhetorical organisation in the grammar and vocabulary section. Since this section mainly tested test-takers' understanding of knowledge of syntax and vocabulary, no knowledge of rhetorical organisation was necessary.

In respect of functional knowledge, knowledge of ideational and knowledge of manipulative functions were the major elements in listening comprehension and writing. In the reading comprehension section, only knowledge of the ideational function was involved. However, all functional knowledge, including ideational, manipulative, heuristic and imaginative knowledge, was tested in the Oral English Test.

In respect of sociolinguistic knowledge, knowledge of registers was involved in the writing section. When asked to write a letter, test-takers were required to understand the characteristics of different levels of formality in language-use. Moreover, knowledge of genres was tested not only in listening and reading (receptive), but also in speaking and writing (productive). Other than that, there were very few items testing sociolinguistic knowledge.

Strategic competence involves test-takers in deciding what they are going to do (goal-setting), appraising the situation and their resources to cope with it (appraising), and deciding how to use the resources they have (planning). In the UEE, the goals were deciding whether or not to attempt the test tasks in the UEE; the assessment involved selecting what competence was needed to complete the tasks; and planning involved deciding how to incorporate language, topic knowledge, and affective schemata.

When checking the test items against Bachman and Palmer's model, the patterns of change for the past 10 years' UEE papers were noticeable and can be summarised as follows:

- The grammar and vocabulary section became passages requiring gap-filling from multiple-choice discrete items. More integrated skills are required when test-takers fill in the gaps in the passages;
- Since more integrated passage items were being used in UEE, the trend of using knowledge of lexical cohesion became more prevalent. Test-takers were required to find the answers by speculating on the passages and understanding the context;
- The passages in the reading comprehension section became more focused on argumentative text type than narrative and descriptive types of writing because students are expected to read more argumentative articles at university;
- From 2005 to 2009, all passages in the reading comprehension were multiple-choice format. However, since 2010, the last passage in this section has been in a short-answer

format. Test-takers have been required to rephrase from the passage to answer the questions, which involved some writing skill;

- The text type of the writing topics has been changed from focusing on narrative and descriptive types to the communicative type; test-takers were required to express their opinions or ideas on some issues in the writing section.

This chapter has summarised what was tested in the UEE and how closely UEE content matched with Bachman and Palmer's model of language theory. The UEE has covered many components of language ability in *grammatical knowledge*, *textual knowledge*, and *strategic competence* but very few in *functional knowledge* and *sociolinguistic knowledge*. In short, this study has shown that the UEE bears an uneven relationship to Bachman and Palmer's model of language theory.

Chapter 6: Construct validity, concurrent validity, face validity and reliability of the English Language Exam

6.1 Introduction

Most researchers recognise three main types of validity in language testing: *construct validity* refers to the actual meaning of the test scores and involves establishing the tester's theories and predictions about the test; *criterion validity* relies on empirical and statistical evidence that agrees with that provided by some independent assessment of the students' ability; and *content validity* depends on a theoretical and logical analysis of the test's content to see if it constitutes a representative sample of the relevant language skills and structures (Alderson et al., 1995; Hughes, 1989; Kunnan, 2004; Morrow, 1979; Popham, 2013; Weir, 1990).

This chapter attempts to validate the English Language Exam papers for university entry in Shanghai by obtaining empirical evidence for all three types of validity information. Correlation methods were adopted as the central device for validation. In this study, three research questions were asked: 1: Construct validity – do the different subtests in the UEE test different skills? 2: Concurrent validity – do the UEE scores match other measures of the students' abilities? 3: Face validity – do the students think the UEE is testing their language ability?

The classic approach to designing correlational studies for construct validation, depicted by Campbell and Fiske (1959), is the multitrait-multimethod matrix. Construct validity is often understood to consist of two other forms of validity: convergent and divergent (Litwin, 1995). In this study, patterns of both convergence and discrimination were examined. Convergence is the extent to which different measures of the same trait tend to agree, whereas discrimination is the extent to which measures of different traits tend to generate different outcomes. Therefore, high positive correlations between the different measures of the same traits indicate convergence – convergent validity, while relatively low correlations between measures of different traits using different approaches reveal discrimination – discriminant validity (Bachman, 1990).

Concurrent validity is a kind of empirical, criterion-related validity (Henning, 1987). The reason why it is empirical is that correlation coefficients are produced by the collection of data and the application of formulas. It is criterion-related because the generated coefficient signifies how strong the relationship is with some external criterion measure. Concurrent validation

compares the scores of a test with some other measures for the same test-takers taken at approximately the same period as the test (Alderson et al., 1995). The higher the correlation coefficient between two tests, the more closely related the two tests are.

Face validity refers to the “surface credibility or public acceptability” of a test (Ingram, 1977, p.18); unlike content validity, which considers “content relevance and content coverage” (Bachman, 1990, p.243), face validity is decided in an impressionistic way. What do the test-takers think of the test? Do they think the test results reflect their language ability?

The reliability of a test lies in its consistency and dependability (Abeywickrama & Brown, 2010). Bachman and Palmer (1996) define “reliability as consistency of measurement” (p.19). In this study, three areas for reliability were considered: 1: internal consistency is concerned with sources of error from within the test; 2: stability indicates how consistent test scores are over time, and 3: parallel-form estimates the reliability of a test to examine of two scores obtained from alternate forms of a test.

Reliability is said to be a “necessary condition for validity, in the sense that test scores that are not reliable cannot provide a basis for valid interpretation and use” (Bachman, 1990, p.289). Therefore, reliability is necessary and important for validity; however, reliability by itself for a test is not sufficient (Alderson et al., 1995).

The purpose of this chapter is to validate the English Language Exam for university entry in Shanghai by empirical evidence. The study was based on mock exams by 60 first-year university students in Shanghai. A description of the study and its research method is introduced. Descriptive statistical analysis and classical item analysis are used to describe the basic features of the data in the study. Correlation analyses are also employed to investigate validity and reliability. Rasch analyses are applied to provide evidence for the content aspect of construct validity. After finishing the mock exams, two participants were interviewed to comment on how they felt about the University Entrance Exam and IELTS sample test.

6.2 Research method

6.2.1 Participants

A total of 60 first-year university students participated in this study. These 60 students were English majors in one of the technical universities in Shanghai. All of them were native Chinese speakers studying English as their second language in the university. They had taken the University Entrance Exam in June 2013, nine months before the study took place. Consent from the participants was obtained on a signed form.

First-year university students were chosen to participate in the mock exams for the following reasons:

1. Senior III students were busy preparing for the University Entrance Exam which was to be held in three months; schools would not let any researchers disturb and affect their study;
2. First-year university students had taken the UEE just nine months ago, and they were still familiar with the UEE format, which was similar to the format of Test of English Major (TEM) – each English-major student was required to take TEM before graduation. Also their vocabulary level was higher than that required in senior high school;
3. It was thought that high school senior I and II students' English proficiency was too far below the UEE requirement.

6.2.2 Instruments

These 60 students were asked to take two past-year papers and one IELTS test sample, a total of three tests. The tests were held on three successive Tuesdays, viz., 25 March, 22 April, and 20 May 2014.

In this study, two years of UEE papers were chosen, 2008 and 2011 (See Appendix C). The UEE comprised two papers: Paper I included Listening Comprehension, Grammar and Vocabulary, and Cloze and Reading Comprehension, whereas Paper II included Translation and Guided Writing. In this study only Paper I was used. For the IELTS test sample (Cambridge ESOL, 2014), there were four sections in each test: listening, reading, writing, and speaking. In this study, only reading and listening sections were used. The details of the UEE format for the 2008 and 2011 UEEs, and the IELTS test sample are shown below (see Table 41):

Table 41: Details of the UEE Formats

Format	UEE 2008	UEE 2011	IELTS Test Sample
Part I Listening Comprehension	Section A 10 Short Conversations 10 MC Questions Section B 2 Passages 6 MC Questions Section C 2 Long Conversations 2 Tables (8 blanks)	Section A 10 Short Conversations 10 MC Questions Section B 2 Passages 6 MC Questions Section C 2 Long Conversations 2 Tables (8 blanks)	Listening Section A total of 4 sections: 10-question items for each section
Time allowed	30 minutes	30 minutes	30 minutes
Part II Grammar and Vocabulary Part III Cloze and Reading Comprehension	Section A 16 MC Questions Section B 1 passage 9 blanks (10 words provided) Section A 1 Passage 15 MC blanks Section B 4 Passages 15 MC Questions Section C 1 Passage with 5 blanks (6 headings provided)	Section A 16 MC Questions Section B 1 passage 9 blanks (10 words provided) Section A 1 Passage 15 MC blanks Section B 3 Passages 11 MC Questions Section C 1 Passage with 5 blanks (6 headings provided) Section D 1 Passage 4 short questions	Reading Section 3 reading passages with a total of 40-question items
Time allowed	60 minutes	60 minutes	60 minutes

The reasons why writing and speaking were not tested in this study are explained as follows:

1. The formats of the UEE and IELTS are very different, especially in respect of their writing sections. For the UEE, there are two parts: translation and guided writing. There are five Chinese sentences that are required to be translated into English, and a short essay of 120 to 150 words of guided writing. The writing section of the UEE takes about 30 minutes. For the Writing Section in IELTS, there are two passages: one requires about 150 words and the other requires about 250 words. The time allowed for this section is one hour;
2. Marking writing papers requires a massive amount of time and professional markers are needed. Because of the limitations of both time and resources, a writing section was not considered for this exercise;
3. Testing speaking requires a huge amount of time (15 minutes for each test-taker). Scoring speaking requires professional training and the researcher is not a qualified examiner. Also it is difficult to find qualified speaking examiners and it is very expensive to run a speaking test;

4. Marking writing and speaking sections is more subjective compared to marking listening and reading sections, which mainly consist of multiple-choice items. Reliability becomes an issue for marking those subjective sections.

6.2.3 Data collection

6.2.3.1 Mock exams procedure

The tests were conducted in 90 minutes on three different Tuesday afternoons over three months. These 60 first-year students from the same university in Shanghai were invited to take two past-year UEE papers (2008 and 2011) and an IELTS sample test. There were two sections: listening and reading. For the 2008 and 2011 UEEs, the listening section of 30 minutes included 10 short conversations, two passages and two long conversations, whereas the reading section of 60 minutes includes grammar and vocabulary, cloze and reading comprehension. For IELTS sample test, four long conversations or lecture seminars with 40-question items were in a 30-minute listening section, whereas three reading passages with 40-question items were in a 60-minute reading section.

6.2.3.2 Scoring

The UEE papers and IELTS sample test were scored differently. The IELTS sample test was scored dichotomously; a correct answer was awarded one point, and a wrong answer 0 points. The scoring criteria for the sample test were given on the IELTS sample tests book. For some answers requiring word-filling, wrong spelling was considered a wrong answer. The 2008 UEE was also scored dichotomously. A correct answer was awarded either one point or two points, and a wrong answer 0. Short and long conversations in listening, grammar and vocabulary, cloze, and heading-matching in reading comprehension were scored one point while passages in listening and passages in reading comprehension were scored two points for each correct answer. The scoring for the 2011 UEE was similar to that of 2008 except for the last short-answer section in the reading comprehension. That section was short-answer items, which accounted for two points each: two points for correct answers; one point if partially correct, and 0 for a wrong answer (SMEEC, 2014b).

6.2.3.3 Post mock exam interviews

Two students, Daisy and Danny (pseudonyms) were randomly selected for an interview after they finished the three tests. They were questioned about how they felt about the UEE and

IELTS sample test and whether the tests had tested their language ability. In other words, face validity was investigated. They were also asked to comment on the differences between the two tests in both listening and reading sections. These short interviews were conducted after the tests in school. Only one student was interviewed at a time. The interviews with students were in Chinese and structured, and they were about five minutes in duration.

6.2.4 Data analysis

6.2.4.1 Descriptive statistical analysis

Descriptive statistical analysis is useful to report on the overall performance of a test paper in that the statistics provide simple summaries about the sample and the measures. Based on the results of 60 participants in the two past years' UEE papers and an IELTS sample test, some of the important statistics for each of the papers were computed, such as means, mode, median, standard deviations, range, maximum, minimum, Skewness, and Kurtosis.

6.2.4.2 Classical item analysis

Traditionally, there are two indexes, which are calculated for each objective test item. They are the facility value and the discrimination index. The facility value (F.V.) measures the level of difficulty of an item, whereas the discrimination index (D.I.) is the degree to which test-takers with high overall test scores also get an individual item correct. An item's facility value is the percentage of test-takers giving the right answer. The value is from 0 to 100 percent. The higher the facility value, the easier the item.

As well as finding out how difficult an item is in the UEE, it is also vital to differentiate test-takers at different levels of language ability. This is measured by the discrimination index. In this study, the 60 participants were divided into three groups according to the total score they obtained from the mock exams. The discrimination index compared the proportion of correct answers in the top third (the strongest group) with those in the bottom third (the weakest group). The discrimination index ranges from -1.00 to +1.00. More high-scoring test-takers are expected to know the answer than low-scoring ones. If the more capable test-takers get an item wrong, while the less capable ones get it right, this means that there must be a problem with the item and that the item needs to be investigated again.

6.2.4.3 Validity

The main purpose of the mock exam study was to check the UEE paper's empirical validity. The construct validity of the UEE was determined by correlating each subtest (listening comprehension, grammar and vocabulary, cloze, and reading comprehension) with other subtests and each subtest with the total test. By correlating students' past UEE paper scores with their scores on other tests, we can gauge the concurrent validity of the UEE. In this study, an IELTS sample test was used as an external test for correlation. Face validity was investigated by interviewing two of the participants to see if they thought the UEE tested their language ability. Interview data were transcribed and translated into English. The transcripts were then coded, analysed, and summarised qualitatively.

6.2.4.4 Reliability

Internal consistency is concerned with sources of error from within the test and any factor other than the competence in a language test that influences the test score is considered as a potential error source (Bachman et al., 1995). By using three different reliability coefficients – the Spearman-Brown reliability coefficients, Cronbach coefficient alphas, and Kuder-Richardson formula 21 reliability coefficients – the internal consistency of the UEE was examined.

In this study, the UEE papers from both 2008 and 2011 were seen as parallel forms of the same test. They could also be regarded as the same test taken twice over a period of time; that is, stability. Basically, the examination format each year is very similar and the UEE papers are supposed to aim for the same construct. If the UEE papers are reliable, the participants will get similar scores on these two UEE papers. Correlating scores from these two exams could thus be regarded as evidence of stability and parallel-form reliability.

6.2.4.5 Rasch analysis

In the Rasch model, the probability of a specified response (e.g. right/wrong answer) is modelled as a function of person and item parameters. The Rasch model routinely sets at 50 percent the probability of success for any person on an item located at the same level on the item-person logit scale. By inputting data collected from the mock exams, question items from the UEE were evaluated in terms of difficulty with respect to the sample of test-takers. Each question in the UEE papers was also checked as to how it fits in the tests, since Rasch analysis

provides indicators of how well each item fits within the underlying construct. Analyses of the UEE items based on the Rasch model were focused on the content aspect of construct validity by referring to content relevance, representativeness, and technical quality.

6.3 Findings and discussion

6.3.1 Descriptive statistical analysis

Table 42: Figures of Descriptive Statistical Analysis for the 2008 and 2011 UEEs

	UEE 2008	UEE 2011
Mean (out of total score 105)	72.5	68.7
Mean (by percentage)	69.0%	65.4%
Standard Deviation (S.D)	14.4	17.1
Variance	206.9	293.9
Maximum score	100	99
Minimum score	41	33
Range	59	66
Median	74.5	71.5
Mode	82	83
Skewness	-0.27	-0.28
Kurtosis	-0.73	-1.07

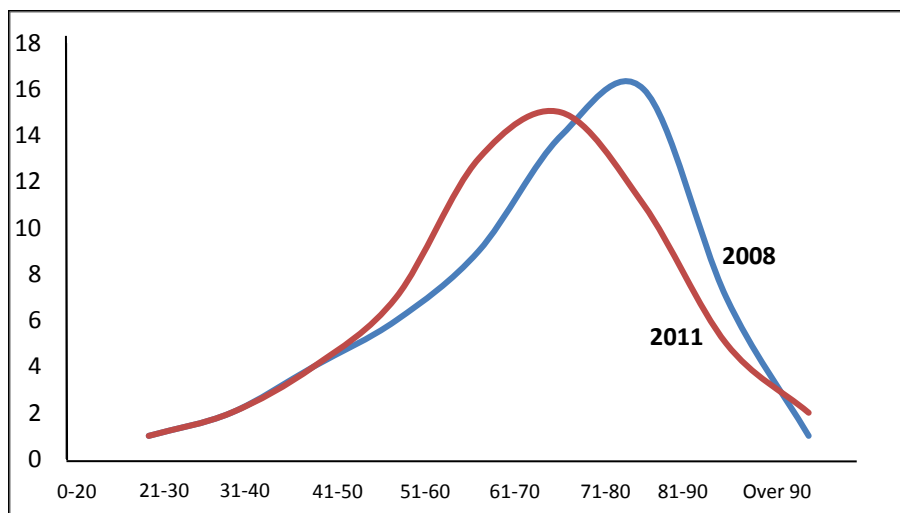


Figure 9: Normal distribution curves for the 2008 and 2011 UEEs

The most crucial statistics are the mean, the median and the mode, which indicate how the scores distribute themselves in the tests. Sixty participants took the 2008 and 2011 UEEs, and the mean, the median and the mode are 72.5, 74.5, 82 for the 2008 UEE, and 68.7, 71.5, 83

for the 2011 UEE out of the total score of 105 (See Table 42). Another important set of the statistics are the standard deviation, the variance, and the range, which exhibit how widely the scores are spread out in the class (Alderson et al., 1995). In this study, the standard deviation and the range are 14.4 and 59 for the 2008 UEE, and 17.1 and 66 for the 2011 UEE.

The mean scores are very interesting. The mean score in 2014 for the University Entrance Exam in Shanghai was 101 (out of a total score of 150), that is, around 67.3 percent (SMEEC, 2014a). It can be seen from the above descriptive statistics that the mean scores by percentage for the 2008 and 2011 UEEs are 69.0 percent and 65.4 percent respectively, which are very close to the factual data.

The distribution for both the 2008 and 2011 UEEs is negatively skewed, in which low scores have the smallest frequencies and high scores have the highest frequencies, and the mean is lower than the median and the mode in each year. In negatively skewed distributions, the tails are at the low score end, and the means for both the 2008 and 2011 UEEs are lower than the median and the mode, as shown in the table above.

Skewness is a measure of symmetry, or more precisely, lack of symmetry. The skewness figures for the 2008 and 2011 UEEs respectively are -0.27 and -0.28, which are very close to each other. The figures match with the statistics shown in the distribution. The distribution of scores is “negatively skewed” because the “scores tail off towards the left end of the graph”; it means that the UEEs were comparatively easy for the participants (Alderson et al., 1995, p.93).

Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. Data sets with high kurtosis (positive value) tend to have a distinct peak near the mean, decline rather rapidly, whereas those with low kurtosis (negative value) tend to have a flat top near the mean. The kurtosis figures are -0.73 and -1.07 for the 2008 and 2011 UEEs respectively, which indicate that the data are inclined to have a flat top near the mean.

As a rule of thumb, skewness and kurtosis figures of “between -2 and +2” display a reasonably normal distribution (Bachman, 2004, p.74). The figures for both skewness and kurtosis in this study are within this range, which mean that scores in both UEEs are normally distributed (See Figure 9).

6.3.2 Classical item analysis

6.3.2.1 Facility value

An item's facility value is the percentage of test-takers who answer the item correctly in the test. The lower the percentage value, the more difficult the item. The average facility values of the items for the 2008 and 2011 UEEs are 71 percent and 67 percent respectively (See Table 43). It shows that these two values are very close to each other. In this study, 60 participants took the mock exams. On average, 43 and 40 participants out of 60 answered each item correctly. These values are reasonably high, which indicates that the UEEs were particularly easy for the participants.

Table 43: Classical Item Analysis – Facility Value for the 2008 and 2011 UEEs

	UEE 2008	UEE 2011
Average facility value of Total Score	71%	67%
Highest F.V.	100%	95%
Lowest F.V.	28%	30%
Average F.V. of each section:		
Listening	57%	68%
Grammar & Vocabulary	82%	77%
Cloze	76%	57%
Reading	69%	61%

By looking at the average facility values of individual sections, it is apparent that the highest average values for the 2008 and 2011 UEEs are from the grammar and vocabulary section (82 percent and 77 percent respectively). This can be explained by the fact that test-takers were very familiar with grammatical and vocabulary items tested in the UEE and that teachers in senior high school spent quite a proportion of time in teaching grammar and vocabulary in class.

The highest average facility values of the total score are 100 percent (item 36) and 95 percent (items 25, 35, 36) for the 2008 and 2011 UEEs respectively. For item 36 in the 2008 UEE, every participant was able to answer this item correctly. This item was not a well-designed item because it was too easy and could not distinguish between test-takers at different levels of the grammatical competence.

The lowest average facility values of the total score are 28 percent (item 24) and 30 percent (item 84) for the 2008 and 2011 UEEs respectively. Item 24 in the 2008 UEE was the last item in listening comprehension; this item required test-takers, after a long conversation, to fill in a table with a word limit. It not only tested their listening skill but also vocabulary and spelling, and grammatical competences. Similarly, Item 84 in the 2011 UEE was the last item in reading comprehension. This passage with short-answer question items required test-takers to complete the statement with the word limit. This item not only tested test-takers' reading ability, but also integrated writing ability.

6.3.2.2 Discrimination index

As well as knowing how difficult an item is, it is important to know how well it distinguishes between students at different levels of language ability. The average discrimination indexes of the item for the 2008 and 2011 UEEs are +0.30 and +0.36 respectively, which are reasonably good (See Table 44). Item designers are usually pleased with a discrimination index of +0.40 or above (Ebel, 1979). By looking at the average discrimination indexes of individual sections, it is noticed that only the average D.I. for the listening section are over 0.4 for the 2008 and 2011 UEEs (0.46 and 0.41 respectively) whilst the indexes of other sections are comparatively low, especially in the 2008 UEE. However, there are no particular rules as to what values are acceptable because it depends on the test type and format and the range of ability of the test-takers (Alderson et al., 1995).

Table 44: Classical Item Analysis – Discrimination Index for the 2008 and 2011 UEEs

	UEE 2008	UEE 2011
Average discrimination index of Total Score	0.30	0.36
Highest D.I.	0.80	0.70
Lowest D.I.	-0.05	-0.15
Average D.I. of each section:		
Listening	0.46	0.41
Grammar & Vocabulary	0.25	0.30
Cloze	0.21	0.36
Reading	0.23	0.37

The highest discrimination indexes for the 2008 and 2011 UEEs are +0.80 (for item 10) and +0.70 (for item 83) respectively, which means that most of the participants in the top group get an answer right and very few participants in the bottom group get an answer right. On the

other hand, two items in the 2008 UEE and one item in the 2011 UEE have negative D.I., which indicates that more participants in the bottom group were correct than in the top group. There is apparently something wrong with those items and they should be revised or discarded. For example, the D.I. of item 57 in the 2011 UEE is -0.15. This could possibly be explained by the performance of one or more of the multiple-choice distractors. Moreover, as discussed later in Chapter 8, this item involved a finance background rather than a language background. It is possible that weaker participants had better finance knowledge than stronger ones or they made better guesses in the mock exams.

6.3.3 Validity

6.3.3.1 Construct validity

One way of measuring the construct validity of the UEE is to correlate the different test components with each other: listening, grammar and vocabulary, cloze, and reading. Since these test components are each assessing something different, the measurements contribute to the overall picture of language ability. We should expect these correlations to be comparatively low – possibly in the order of +0.31 to +0.67; suggesting that there is no crucial overlap (Wall et al., 1994, p.340). If any two test components correlate highly with each other, for example, +0.9, we might say that two test components are assessing the same characteristics of the language ability. In this case, we may only need one of the two test components.

Table 45 and Table 46 show the subtest inter-correlations, the correlations between the subtests and the total score for both the 2008 and 2011 UEEs. Since the subtests are intended to test different aspects of language, they are not expected to correlate very highly with each other. The correlations between listening, grammar and vocabulary, cloze, and reading are therefore satisfactory; they are between +0.465 and +0.699. Some common variance would be expected since they were all testing English language ability; however, sufficient unshared variance showing the measurement of different skills is observed. The +0.617 and +0.699 correlation between listening and grammar and vocabulary is higher than the others for the 2008 and 2011 UEEs respectively.

Table 45: Correlation Matrix for the 2008 UEE

	Listening	Grammar & Vocabulary	Cloze	Reading	Total Score	Total minus self
Listening		.617***	.480***	.465***	.871***	.616***
Grammar & Vocab			.530***	.503***	.797***	.681***
Cloze				.525***	.700***	.606***
Reading					.785***	.522***

Note: ***p<0.001 (N=60)

Table 46: Correlation Matrix for the 2011 UEE

	Listening	Grammar & Vocabulary	Cloze	Reading	Total Score	Total minus self
Listening		.699***	.563***	.494***	.846***	.674***
Grammar & Vocab			.625**	.526***	.834***	.734***
Cloze				.505***	.759***	.757***
Reading					.824***	.579***

Note: ***p<0.001 (N=60)

As mentioned, if different subtests are testing different constructs, the correlations between them should not be very high; this is called discriminant validity. A low correlation or not very high correlation means that different subtests can be distinguished from each other. As can be seen from Table 45 and Table 46, correlations between listening and reading are relatively low ($r=+0.465$ and $r=+0.494$ for the 2008 and 2011 UEEs respectively, $p<0.001$). Relatively low correlations indicate they are discriminant from each other. In this case, the listening section must be different from the reading section, because the two sections are testing two different skills. However, the correlations are not very low because the two sections are testing other similar competences such as vocabulary, comprehension and strategic competence, and listening and reading are related skills sharing similar cognitive processes.

All the subtests have correlations with the total score above +0.7. The correlations between each subtest and the total score are expected to be higher than that amongst the subtests, since the overall score is taken to be a more general assessment of language ability than each

individual component score. It shows that all these subtests contribute to the whole test. The coefficients of listening with the total score are the highest ($r=+0.871$ and $r=+0.846$ for the 2008 and 2011 UEEs, $p<0.001$). This component shares 76 and 72 percent of variance with the total score respectively. It shows that this test component significantly contributes to the total score.

Not unexpectedly, reading has the lowest correlation with the total score minus self ($r=+0.522$ for the 2008 UEE and $r=+0.579$ for the 2011 UEE, $p<0.001$). This may well be due to the fact that this subtest has three different types of items: multiple-choice, heading-matching, and short-answer questions. Nevertheless, all correlation coefficients are still over $+0.5$ after the reduction. These strong correlations between the subtests and the total show how significant these subtests are in the UEE battery (Alderson et al., 1995).

6.3.3.2 Concurrent validity

Concurrent validity can be gauged by associating the UEE scores with scores on the IELTS, which is a standardised international test offered by Cambridge; it is considered to be a highly reliable and valid measure of the English proficiency test (Fulcher, 1997). Table 47 displays the correlation coefficients between the 2008 and 2011 UEEs and the IELTS sample test. The correlation coefficients are low ($r=+0.508$ for the 2008 UEE and $r=+0.435$ for the 2011 UEE, $p<0.001$). The relatively low correlations between the two tests and IELTS can be explained as follows. These two tests have some differences: different exam formats, different scoring systems, different test specifications, different vocabulary ranges in the materials, different test designers, and different test materials.

Table 47: Correlation Matrix for the Two UEEs and IELTS

	Listening UEE 2008	Listening UEE 2011	Reading UEE 2008	Reading UEE 2011	Total Score UEE 2008	Total Score UEE 2011
Listening_IELTS	.352**	.499***				
Reading_IELTS			.328*	0.249		
Total_IELTS					.508***	.435***

Notes: *** $p<0.001$, ** $p<0.01$, * $p<0.05$ (N=60)

When the Listening or Reading Comprehension sections of the UEEs are correlated separately with the IELTS test sample, the coefficients are quite low, around +0.35. The 2011 UEE Reading Section, especially, has correlation with the IELTS Reading Section of only 0.249, which is very low. A few reasons could account for the low coefficients:

- The IELTS test format was very different from that of the UEE. The students were not really familiar with the IELTS test format. However, they were very familiar with the UEE format, as they had been practising it since high school;
- The passages in IELTS were much longer than those in the UEE, and students could not finish within a time limit of an hour. The vocabulary range was also greater, which means that passages in IELTS were more difficult than the UEE;
- There was a large range of test tasks and question items in IELTS in both listening and reading – matching the given information or statements for identified paragraphs, short-answer questions, multiple-choice items, matching the information with answers or phrases, flow-chart or table completion, and identification of the information or statements. Most of the items in the UEE were multiple-choice items;
- In the listening section, the speed of the speakers in IELTS was much faster and more natural than in the UEE, which was slower and more easily understood. The recordings of the conversations and passages in IELTS also were only played once, whereas those in the longer conversations and passages were played twice;
- The scoring system of IELTS was different from the UEE. There were 40 items in each section, equally weighted, whereas the total score of Paper I in the UEE was 105; some items were two points and some were awarded by one point only;
- IELTS is an international English language proficiency test whereas the UEE is partly a proficiency test and partly an achievement test, as test-takers are provided with a vocabulary handbook listing out the vocabulary words, which are supposed to be learned before sitting in the UEE.

6.3.4 Reliability

6.3.4.1 Internal consistency

If all the items are considered to assess the same skill in the same way, then the test items will inter-correlate highly; and the reliability index for the test will be high (Alderson et al., 1995). From Table 48, the Spearman-Brown reliability coefficients are +0.873 and +0.925 for the 2008 and 2011 UEEs respectively. The result indicates that both exams have shown a relatively high reliability, with both figures over +0.870.

Table 48: Reliability Coefficients for the 2008 and 2011 UEEs

	Spearman-Brown reliability coefficient	Cronbach coefficient alpha	Kuder-Richardson 21 reliability coefficient
UEE 2008	.873***	0.878***	.963***
UEE 2011	.925***	0.909***	.969***

Note: ***p<0.001 (N=60)

Splitting the exam into two equal halves is problematic. One of the problems is that the UEE items are not completely independent of each other for both halves. For the UEE items, only items of the short conversations from the listening section and the grammar from the grammar and vocabulary section are discrete-point, 'stand-alone' multiple-choice items. However, some items, from passages and longer conversations in the listening section and passages in the reading section, comprise sets of question items based on the same passages. Two assumptions must be fulfilled in order to use the Spearman-Brown split-half method. First, since the two halves are treated as parallel tests, it is assumed that they have equal means and variances. However, in this case, the two sets of data were not exactly the same. Therefore, the reliability is underestimated. Second, the two halves were experimentally independent of each other. One of the examples is from the grammar and vocabulary section. In this section, there were 10 options provided in a box in a vocabulary passage; if one of the options was chosen by the test-takers, it would not be picked again for other items of the same passage. Another example is that of heading-matching from the reading section. There were six-heading options to be chosen by test-takers for a five-paragraph passage. Once a heading option was chosen, it would not be chosen again for the other paragraphs. In this case the assumption that the two halves are independent of each other is much more problematic. Thus, the split-

half method may create a situation that the items are not independent, and that the halves are consequently not independent. Hence, the reliability is overestimated.

Table 49: Cronbach Coefficient Alphas for the 2008 and 2011 UEEs

	UEE 2008	UEE 2011
Total Score	0.878***	0.909***
Listening	0.836***	0.810***
Total minus Listening	0.794***	0.875***
Grammar & Vocab	0.729***	0.792***
Total minus G & V	0.844***	0.882***
Cloze	0.516***	0.727***
Total minus Cloze	0.864***	0.893***
Reading	0.593***	0.779***
Total minus Reading	0.880***	0.900***

Note: ***p<0.001 (N=60)

From Table 49, Cronbach coefficient alphas of the total score are +0.878 and +0.909 for the 2008 and 2011 UEEs respectively. The result indicates that both UEEs have revealed a relatively high reliability. Also, most of the coefficients for different sections in both 2008 and 2011 UEEs are around +0.800, which are reasonably satisfactory, except values for cloze (+0.516) and reading (+0.593) sections in 2008.

Low coefficient means the internal consistency of the test items of a particular section is low. The lowest coefficient is obtained in the cloze section because of the following reasons: 1: there was a total of 15 items in the cloze section, every blank in this section measuring the same competences could be difficult; 2: the gaps of the cloze passage were rational; 3: the difficulty level of each gap would be different – some were more difficult than others as some blanks were provided with more information in the passage or context, and 4: different gaps were testing different competences: grammar, vocabulary, reading comprehension, and strategies for dealing with unknown words. These items were not consistently testing the same competences; therefore; the internal consistency of the test items is low.

Reliability of a test can be expressed and estimated in different ways. Other than the Spearman-Brown reliability coefficient and Cronbach coefficient alpha, as mentioned above, there is another approach called KR-21, which was developed by Kuder and Richardson (Bachman, 1990), involving the calculation of the means and variances of the items that establish the test. The KR-21 reliability coefficients for the 2008 and 2011 UEEs are 0.963 and

0.969 respectively, which are considered to be significantly high; hence, they are reasonably reliable. High reliability indexes were obtained due to the examinations items being well-constructed and objective (Alderson et al., 1995).

On the other hand, these two KR-21 reliability coefficients are a little higher than the other two coefficients as can be seen in Table 48. One of the possible reasons is because of the small number of participants; only 60. If there had been over 100 participants, these two sets of coefficients would probably have been similar (Alderson et al., 1995).

6.3.4.2 Stability and parallel-form reliability

From Table 50, the 2008 UEE has positive correlations with all parts of the 2011 UEE. The correlation coefficients between the 2008 and 2011 UEEs for listening ($r=+0.480$, $p<0.001$), grammar and vocabulary ($r=+0.706$, $p<0.001$), cloze ($r=+0.332$, $p<0.001$), reading ($r=+0.378$, $p<0.001$), and total score ($r=+0.607$, $p<0.001$) are comparatively low.

Table 50: Correlation Matrix for Stability and Parallel-form Reliability – 2008 versus 2011

2008\2011	Listening	Grammar & Vocab	Cloze	Reading	Total Score
Listening	.480***				
Grammar & Vocab		.706***			
Cloze			.332***		
Reading				.378***	
Total Score					.607***

Note: *** $p<0.001$ (N=60)

It is found that the figures for the grammar and vocabulary section and total score are comparatively higher, which are +0.706 and +0.607. In the grammar section, there are 16 multiple-choice discrete items. The UEE designers tested similar grammatical competences, for example, tense and voice, modal verbs, adverbial clauses, and noun clauses. In the vocabulary section, test-takers were tested on their vocabulary usage and forms in the chosen passage; the formats were very similar. The reliability coefficient for this section is comparatively high; hence, the grammar and vocabulary section is quite stable.

On the other hand, low correlations were obtained for the cloze and reading sections (+0.332 and +0.378, $p<0.001$). One of the reasons is that different reading passages were chosen and the question items were designed by two groups of UEE designers. Also reading section of the

2011 UEE format differs slightly from 2008's. For the 2011 UEE, the items for the last passage are short-answer questions, instead of multiple-choice questions, as in 2008.

In summary, the participants did not obtain similar scores for the two UEE papers because of the following reasons: different materials and content were covered in the two UEE papers; different vocabulary was tested in each UEE paper; formats were slightly different – short-answer question items; difficulty level and facility values of the items were all different, and participants were in different moods or statuses in the two mock exams.

6.3.5 Rasch analysis

6.3.5.1 Test-takers' ability versus test difficulty

Rasch measurement is a model that tries to demonstrate the relationship between different aspects of a test's conditions (McNamara, 1996). The most common aspects in the condition are test-takers' ability and test difficulty. The item difficulty and person ability maps of the Rasch model indicate a reasonably good match between the test-takers' ability and the test item. On the left are the units of measurement on the logit scale extending from -4 to +4. The average item difficulty has been set at 0 logit. Figure 10 and Figure 11 show the person-item maps for the mock exams (2008 and 2011).

The numbers (101-130 and 201-230) on the left region represent the 60 participants in the mock exams. The ability of individual participants is entered on the scale; the higher the number, the more capable the participant. Participants located higher on the scale have a higher probability of answering all the UEE items correctly. Ss0211 in the 2008 UEE and Ss0204 in the 2011 UEE were more likely to answer all the items correctly than the other participants. On the right are the UEE items, numbered 1 to 84; the higher the item number on the logit scale, the more difficult the question item; the most difficult items were 24 and 64 for the 2008 and 2011 UEEs respectively whereas the easiest items were 36 for the 2008 UEE, and 25, 35, and 36 for the 2011 UEE respectively.

In the Rasch model, the same gauge is used as an expression of the measurement and relationship of the test-taker's ability and the item's difficulty. This denotes that person ability and item difficulty can be measured and compared for a specific group of test-takers and test items, to see how fit they are matched (McNamara, 1996). It is found from the maps that there were 24 and 18 items (out of 84) below the least capable student on the logit scale in the 2008

and 2011 UEEs respectively. Meanwhile, there were 6 participants (out of 60 participants) above the most difficult item on the scale in both the 2008 and 2011 UEEs respectively. The person-item maps show that the UEEs are neither too difficult nor too easy for the sample, just well-matched.

The Rasch model routinely sets the probability of success for any test-taker on an item located at the same level on the item-person logit scale at 50 percent. Looking at the person-item map for the 2011 UEE, if a student's logit ability estimate is equal to an item's difficulty estimate, that student has a 50 percent chance of passing this item, for which he or she is equally matched. The logit ability estimate of the participant 115 (Ss0115) is equal to the difficulty estimate for items 2, 41, and 43, and thus it is concluded that Ss0115 has a 50 percent chance of passing these items. Moreover, the probability of their success increases to 75 percent for an item which is 1 logit easier, or decreases to 25 percent for an item which is 1 logit harder. For example, Ss0115 is approximately 1 logit lower than the difficulty estimate for items 14 and 55; it means that Ss0115 has only a 25 percent chance of passing these items. Or if Ss0115 is approximately 1 logit higher than item 37's difficulty estimate; then Ss0115 has a 75 percent chance of answering these two items correctly. If the item is 2 logits less, the chances are closer to 90 percent.

The measured person ability estimates of +1.25 and +0.97 on the logit scale for the 2008 and the 2011 UEE and the mean of measured item difficulty at -0.05 and 0.00 logit indicate that test-takers find the tests reasonably appropriate (See Table 51 and Table 52). When the mean person estimates are closer to the mean of a measured item it indicates that the tests are well-matched. On the other hand, measured items tables for both the 2008 and 2011 UEEs show that most of the items have relatively small standard errors (MODEL S.E. = 0.38 and 0.33) associated with their difficulty estimates, because ability estimates for the test-takers are close to the same level as those items (Bond & Fox, 2001).

variable-maps-2008
 ZOU851WS.TXT Aug 20 13:35 2014
 TABLE 1.0 2008_whole.xlsx
 INPUT: 60 PERSON 84 ITEM REPORTED: 60 PERSON 84 ITEM 2 CATS WINSTEPS 3.81.2

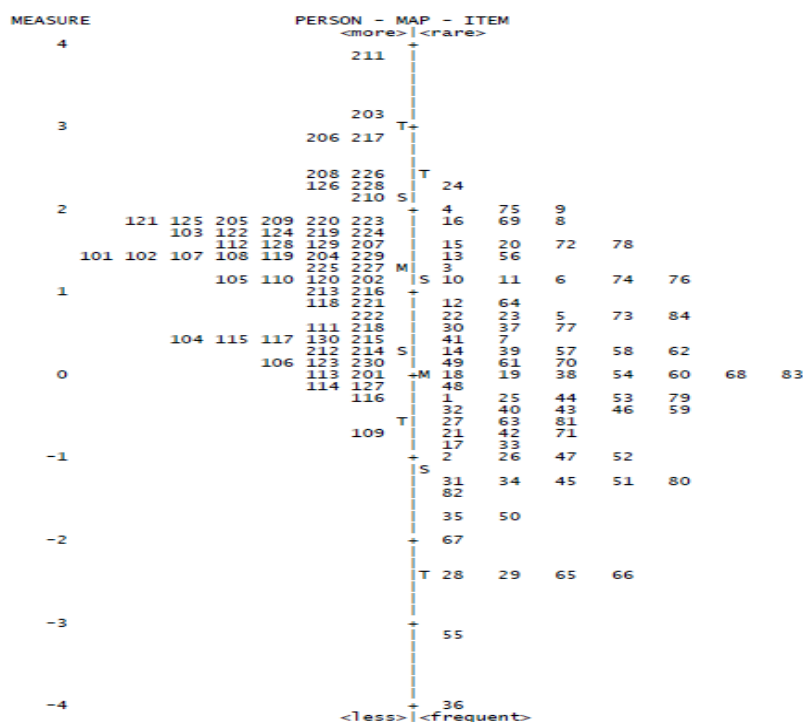


Figure 10: Person-item map_2008

variable-maps-2011
 ZOU445WS.TXT Aug 25 14:54 2014
 TABLE 1.0 2011_whole.xlsx
 INPUT: 60 PERSON 84 ITEM REPORTED: 60 PERSON 84 ITEM 5 CATS WINSTEPS 3.81.2

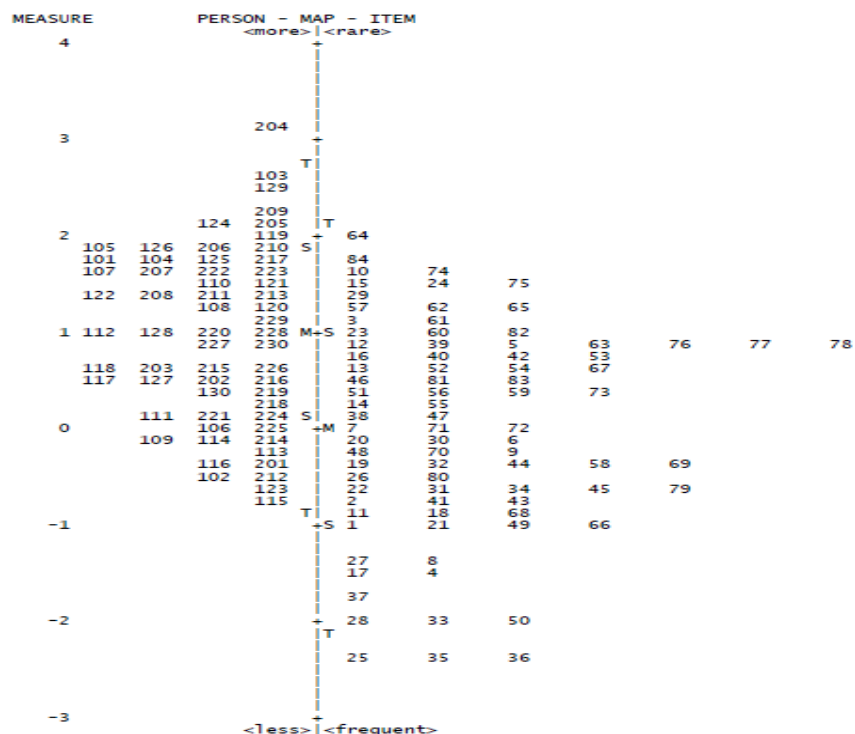


Figure 11: Person-item map_2011

Table 51: Summary of Measured Person and Item for the 2008 UEE

SUMMARY OF 60 MEASURED PERSON									
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		
					MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	59.6	84.0	1.25	.29	1.00	.0	1.04	.0	
S.D.	11.3	.0	.90	.06	.11	.9	.38	1.0	
MAX.	81.0	84.0	3.86	.60	1.28	1.8	2.36	2.7	
MIN.	32.0	84.0	-.64	.25	.79	-1.6	.58	-1.4	
REAL RMSE	.30	TRUE SD	.85	SEPARATION	2.78	PERSON RELIABILITY			.89
MODEL RMSE	.30	TRUE SD	.85	SEPARATION	2.86	PERSON RELIABILITY			.89
S.E. OF PERSON MEAN = .12									
PERSON RAW SCORE-TO-MEASURE CORRELATION = .98									
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .90									
SUMMARY OF 84 MEASURED ITEM									
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		
					MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	42.6	60.0	-.05	.38	1.00	.0	1.04	.1	
S.D.	11.0	.0	1.29	.20	.11	.9	.38	1.0	
MAX.	60.0	60.0	2.32	1.83	1.29	2.1	3.12	2.9	
MIN.	17.0	60.0	-4.41	.28	.75	-2.8	.23	-2.4	
REAL RMSE	.44	TRUE SD	1.21	SEPARATION	2.75	ITEM RELIABILITY			.88
MODEL RMSE	.43	TRUE SD	1.21	SEPARATION	2.80	ITEM RELIABILITY			.89
S.E. OF ITEM MEAN = .14									

Table 52: Summary of Measured Person and Item for the 2011 UEE

SUMMARY OF 60 MEASURED PERSON									
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		
					MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	58.2	84.0	.97	.26	1.01	.1	.97	.0	
S.D.	14.4	.0	.91	.04	.15	.9	.23	.9	
MAX.	83.0	84.0	3.15	.46	1.49	2.8	1.60	2.5	
MIN.	29.0	84.0	-.77	.23	.71	-2.0	.57	-1.5	
REAL RMSE	.28	TRUE SD	.87	SEPARATION	3.15	PERSON RELIABILITY			.91
MODEL RMSE	.27	TRUE SD	.87	SEPARATION	3.27	PERSON RELIABILITY			.91
S.E. OF PERSON MEAN	= .12								
PERSON RAW SCORE-TO-MEASURE CORRELATION = .99									
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .92									
SUMMARY OF 84 MEASURED ITEM									
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		
					MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	41.6	60.0	.00	.33	1.00	.1	.97	.0	
S.D.	11.6	.0	1.05	.08	.13	1.0	.26	1.1	
MAX.	76.0	60.0	1.97	.60	1.45	3.8	1.80	3.8	
MIN.	18.0	60.0	-2.32	.17	.75	-2.3	.44	-2.3	
REAL RMSE	.35	TRUE SD	.99	SEPARATION	2.87	ITEM RELIABILITY			.89
MODEL RMSE	.34	TRUE SD	1.00	SEPARATION	2.93	ITEM RELIABILITY			.90
S.E. OF ITEM MEAN	= .12								

Each participant's ability has error estimation as well. Measured person tables also show that most of the participants, except a few highly capable ones, have relatively small standard errors (MODEL S.E. = 0.29 and 0.26) because they have more items close to their ability level (Bond & Fox, 2001).

Based on an analysis of variance (t-test), there was no statistically significant difference between 2008 and 2011 test performance ($p > 0.05$). This suggests that the 2008 UEE might have been as difficult as the 2011 UEE. In other words, the scores in one test do not vary much more than the scores in the other one. This finding makes sense, in that both UEE papers were testing the same target domains, covered the same standards, and contained similar format and tasks.

6.3.5.2 Content aspect of construct validity

Various analyses were carried out to provide evidence for the content aspect of construct validity within the Rasch model. The content aspect of construct validity refers to “content relevance, representativeness, and technical quality” (Baghaei & Amrahi, 2011, p.1052).

Table 53 and Table 54 show the fit indexes for some of the items in the 2008 and 2011 UEEs. The items are arranged from misfit to overfit items. The first column, ‘ENTRY NUMBER’, indicates the number given to each item in the test (ranging from 1 to 84). The second column, ‘TOTAL SCORE’, represents the number of participants who answered that item correctly. The third column, ‘COUNT’, is the total number of participants. The difficulty estimates for the items are shown in the fourth column labelled ‘MEASURE’. The fifth column, ‘MODEL S.E.’, shows the standard error of the item difficulty measures. ‘MNSQ’ and ‘ZSTD’ are abbreviations for ‘mean-square’ and ‘z standardised distribution’ respectively, and are provided for ‘INFIT’ and ‘OUTFIT’ columns.

Fit indexes were used to check the relevance of the test content to the intended construct. For the 2008 UEE, infit figures (MNSQ) are between +0.75 and +1.29, and infit figures (ZSTD) are between -2.8 and +2.1. Similarly, for the 2011 UEE, infit figures are from +0.75 to +1.45 and infit figures (ZSTD) are from -2.7 to +3.8. According to McNamara’s (1996) rule of thumb, infit figures in the range of approximately +0.75 to +1.30 for ‘MNSQ’ and from -2.0 to +2.0 for ‘ZSTD’ are believed to be acceptable. Infit figures less than +0.75 for ‘MNSQ’ and -2.0 for ‘ZSTD’ indicate significant overfit and those above +1.30 for ‘MNSQ’ and +2.0 for ‘ZSTD’ reveal significant misfit. Misfit items mean that the items lack predictability and that the test-taker’s performance on one item could not be forecast from that performance on other items. Items 10, 11, and 62 for the 2008 UEE (See Table 53) and items 40, 54, 57, 67, 73, and 75 for the 2011 UEE (See Table 54) are outside of the acceptable range. These items outside of the acceptable range should be reviewed because of lack of fit to the model. These items are measuring something other than the intended content and construct. That is, they are construct-irrelevant. In sum, most of the test items are within the acceptable range; therefore, the 2008 and 2011 UEEs items are construct-relevant in general.

Two significant criteria – person-item maps and item strata – are used to verify the representativeness of the test items. Noticeable gaps in the item difficulty indicate that some area of the construct domain has not been included by the test (Baghaei, 2008, as cited in

Baghaei & Amrahi, 2011). From the person-item maps (See Figure 10 and Figure 11), the only noticeable gap between the 2008 and 2011 UEEs is the area above the most difficult item (items 24 and 64 respectively). There are no items whose difficulty levels are above the top six participants' ability levels in both UEEs (211, 203, 206, 217, 208, and 226 in 2008 and 204, 103, 129, 209, 124, and 205 in 2011); it means that some items are required to cover this area of the construct domain. However, only 10 percent of the participants were not covered. Overall, the items show an acceptable degree of representativeness.

Item strata, which refer to "the number of statistically distinct regions of item difficulty that the persons have distinguished" (Smith, 2001, as cited in Baghaei & Amrahi, 2011, p.1054), is another indication of whether or not the items are appropriate representatives of the intended content. Item strata are labelled as 'SEPARATION' for the 84 measured items of the 2008 and 2011 UEEs in the Table 51 and Table 52. The minimum value for item strata is 2. The separation values given are +2.75 and +2.87 respectively, which are acceptable. Therefore, one can rely on the representativeness of the UEE items.

Technical quality of the test items can be measured through item-measure correlations because they are an indicator of "the degree to which the scores on a particular item are consistent with the average score across the remaining items" (Wolfe & Smith, 2007, as cited in Baghaei & Amrahi, 2011, p.1054). Regarding the expected values of the item-measure correlations, Wolfe and Smith recommend that the values should be positive, showing that the scores on the item are positively correlated with the average score on the remaining items (Baghaei & Amrahi, 2011). From Table 53 and Table 54, the 'PT-MEASURE' column indicates the observed (CORR) as well as the expected (EXP) correlation between performance on each item and the ability estimates of the participants who have answered that item correctly. It can be seen that there is only one item in the 2008 and 2011 (item 29; $r=-0.14$ and item 57; $r=-0.08$ respectively) UEEs respectively showing the negative value of the item-measure correlations. The remaining items have positive values, indicating that both the 2008 and 2011 UEEs have very satisfactory technical quality.

In summary, from the resultant data of the Rasch analysis, both the 2008 and 2011 UEE items show a reasonable degree of relevance, an acceptable degree of representativeness, and a satisfactory degree of technical quality. Therefore, it can be concluded that both the 2008 and 2011 UEEs have a satisfactory degree of construct validity.

Table 53: Item Statistics – Misfit Order for the 2008 UEE

ITEM STATISTICS: MISFIT ORDER													
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	EXP.	EXACT OBS%	MATCH EXP%	ITEM
29	58	60	-2.47	.73	1.10	.4	3.12	1.8	A-.14	.14	96.7	96.7	29
50	56	60	-1.71	.53	1.01	.2	2.29	1.6	B-.08	.19	93.3	93.3	50
40	48	60	-.36	.34	.99	.0	1.83	2.2	C-.27	.31	78.3	80.2	40
52	53	60	-1.06	.42	1.02	.2	1.70	1.4	D-.17	.25	88.3	88.3	52
71	51	60	-.74	.38	1.09	.4	1.69	1.6	E-.11	.27	85.0	85.0	71
28	58	60	-2.47	.73	1.03	.3	1.44	.7	F-.07	.14	96.7	96.7	28
26	53	60	-1.06	.42	1.10	.4	1.43	1.0	G-.07	.25	88.3	88.3	26
62	41	60	.34	.30	1.29	2.1	1.43	2.0	H-.04	.35	63.3	71.6	62
74	32	60	1.08	.28	1.19	1.9	1.43	2.9	I-.14	.38	58.3	66.5	74
60	45	60	-.03	.32	1.24	1.5	1.38	1.4	J-.06	.33	73.3	75.9	60
42	51	60	-.74	.38	.97	-1.1	1.35	.9	K-.26	.27	85.0	85.0	42
16	23	60	1.80	.29	1.14	1.3	1.25	1.6	L-.21	.38	60.0	67.8	16
72	26	60	1.55	.28	1.16	1.6	1.25	1.8	M-.20	.38	60.0	66.3	72
37	39	60	.52	.29	1.23	1.8	1.21	1.2	N-.14	.36	58.3	70.0	37
75	20	60	2.05	.30	1.05	.4	1.22	1.2	O-.28	.37	71.7	71.1	75
58	41	60	.34	.30	1.07	.6	1.21	1.0	P-.26	.35	66.7	71.6	58
61	43	60	.16	.31	1.15	1.1	1.19	.9	Q-.18	.34	68.3	73.6	61
39	41	60	.34	.30	1.10	.8	1.18	.9	R-.24	.35	70.0	71.6	39
34	54	60	-1.24	.44	1.07	.3	1.17	.5	S-.14	.23	90.0	90.0	34
57	41	60	.34	.30	1.03	.2	1.17	.9	T-.30	.35	73.3	71.6	57
BETTER FITTING OMITTED													
47	53	60	-1.06	.42	.92	-.2	.70	-.5	t-.37	.25	88.3	88.3	47
55	59	60	-3.19	1.02	.92	-.2	.23	-.5	s-.27	.10	98.3	98.3	55
44	47	60	-.24	.33	.91	-.5	.77	-.7	r-.43	.31	78.3	78.7	44
32	49	60	-.48	.35	.90	-.5	.81	-.5	q-.41	.30	83.3	81.8	32
19	45	60	-.03	.32	.89	-.7	.80	-.7	p-.46	.33	76.7	75.9	19
20	26	60	1.55	.28	.89	-1.1	.88	-.8	o-.49	.38	73.3	66.3	20
45	54	60	-1.24	.44	.88	-.3	.50	-1.0	n-.42	.23	90.0	90.0	45
56	27	60	1.47	.28	.88	-1.2	.83	-1.3	m-.51	.38	68.3	66.0	56
4	20	60	2.05	.30	.87	-1.1	.82	-1.0	l-.51	.37	75.0	71.1	4
17	52	60	-.89	.40	.86	-.5	.60	-.9	k-.44	.26	86.7	86.7	17
3	29	60	1.32	.28	.85	-1.7	.80	-1.6	j-.55	.38	70.0	65.9	3
9	21	60	1.96	.29	.84	-1.4	.75	-1.6	i-.55	.37	73.3	70.0	9
14	41	60	.34	.30	.82	-1.5	.73	-1.4	h-.55	.35	80.0	71.6	14
82	55	60	-1.46	.48	.82	-.4	.41	-1.1	g-.46	.21	91.7	91.7	82
11	31	60	1.16	.28	.81	-2.1	.77	-1.9	f-.58	.38	75.0	66.2	11
43	48	60	-.36	.34	.81	-1.0	.63	-1.2	e-.53	.31	81.7	80.2	43
63	50	60	-.61	.36	.81	-.8	.62	-1.1	d-.50	.29	85.0	83.4	63
48	46	60	-.14	.38	.79	-1.3	.63	-1.5	c-.56	.29	84.7	77.2	48
21	51	60	-.74	.38	.78	-.9	.52	-1.3	b-.53	.27	85.0	85.0	21
10	32	60	1.08	.28	.75	-2.8	.71	-2.4	a-.64	.38	78.3	66.5	10
MEAN	42.6	60.0	-.05	.38	1.00	.0	1.04	.1			77.3	77.6	
S.D.	11.0	.0	1.29	.20	.11	.9	.38	1.0			10.5	9.7	

Table 54: Item Statistics – Misfit Order for the 2011 UEE

ITEM STATISTICS: MISFIT ORDER														
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	EXP.	EXACT OBS%	MATCH EXP%	ITEM	G
25	57	60	-2.32	.60	1.09	.3	1.80	1.1	A-.00	.18	95.0	95.0	25	1
82	61	60	-.95	.17	.88	-.7	1.64	2.2	B-.53	.60	63.3	51.3	82	2
57	26	60	1.29	.28	1.45	3.8	1.60	3.8	C-.08	.39	48.3	67.6	57	1
31	48	60	-.64	.34	.97	-.1	1.46	1.4	D-.31	.32	81.7	80.1	31	1
34	48	60	-.64	.34	1.07	.4	1.43	1.3	E-.20	.32	81.7	80.1	34	1
75	24	60	1.46	.29	1.26	2.2	1.41	2.5	F-.10	.39	58.3	68.0	75	1
67	35	60	.57	.29	1.32	2.7	1.39	2.5	G-.06	.39	55.0	68.5	67	1
73	38	60	.33	.29	1.31	2.4	1.34	1.9	H-.08	.39	60.0	69.9	73	1
74	22	60	1.62	.29	1.22	1.8	1.34	2.0	I-.14	.38	63.3	68.9	74	1
40	33	60	.74	.28	1.24	2.1	1.26	1.9	J-.16	.40	55.0	68.1	40	1
83	76	60	.50	.18	1.24	1.3	1.26	.9	K-.53	.59	51.7	55.9	83	2
32	46	60	-.42	.33	1.00	.1	1.25	1.0	L-.29	.34	80.0	77.2	32	1
52	34	60	.66	.28	1.18	1.6	1.22	1.5	M-.21	.40	63.3	68.2	52	1
12	31	60	.90	.28	1.11	1.0	1.21	1.6	N-.27	.40	65.0	67.7	12	1
29	25	60	1.38	.28	1.14	1.3	1.21	1.5	O-.24	.39	60.0	67.8	29	1
72	42	60	-.03	.30	1.12	.9	1.20	1.0	P-.25	.37	63.3	72.9	72	1
58	46	60	-.42	.33	1.08	.5	1.17	.7	Q-.25	.34	80.0	77.2	58	1
55	39	60	.24	.29	1.09	.8	1.15	.9	R-.28	.38	66.7	70.6	55	1
77	32	60	.82	.28	1.15	1.4	1.15	1.1	S-.26	.40	60.0	67.9	77	1
84	34	60	1.77	.18	1.12	.8	.99	.1	T-.50	.53	56.7	55.7	84	2
BETTER FITTING OMITTED														
59	37	60	.41	.29	.91	-.8	.89	-.7	t-.48	.39	71.7	69.4	59	1
64	18	60	1.97	.30	.91	-.7	.84	-.7	s-.47	.36	73.3	72.4	64	1
79	48	60	-.64	.34	.91	-.5	.82	-.5	r-.43	.32	81.7	80.1	79	1
47	40	60	.15	.30	.90	-.8	.86	-.7	q-.48	.38	75.0	71.2	47	1
21	51	60	-1.03	.38	.88	-.4	.63	-1.0	p-.45	.29	85.0	85.0	21	1
27	53	60	-1.35	.42	.88	-.3	.66	-.7	o-.41	.26	88.3	88.3	27	1
44	45	60	-.32	.32	.88	-.8	.77	-.9	n-.49	.35	78.3	76.0	44	1
37	55	60	-1.75	.48	.87	-.3	.50	-.9	m-.41	.22	91.7	91.7	37	1
65	26	60	1.29	.28	.87	-1.3	.82	-1.4	l-.53	.39	75.0	67.6	65	1
71	42	60	-.03	.30	.87	-1.0	.78	-1.1	k-.52	.37	76.7	72.9	71	1
63	31	60	.90	.28	.85	-1.5	.81	-1.6	j-.56	.40	75.0	67.7	63	1
1	51	60	-1.03	.38	.83	-.7	.55	-1.2	i-.50	.29	85.0	85.0	1	1
16	33	60	.74	.28	.83	-1.7	.81	-1.5	h-.57	.40	78.3	68.1	16	1
18	50	60	-.89	.36	.83	-.8	.66	-1.0	g-.50	.30	83.3	83.3	18	1
48	44	60	-.22	.31	.82	-1.3	.73	-1.2	f-.55	.36	78.3	74.9	48	1
19	46	60	-.42	.33	.81	-1.1	.61	-1.6	e-.57	.34	76.7	77.2	19	1
22	48	60	-.64	.34	.80	-1.1	.58	-1.1	d-.56	.32	78.3	80.1	22	1
46	36	60	.49	.29	.79	-2.0	.74	-1.8	c-.61	.39	76.7	68.9	46	1
54	35	60	.57	.29	.76	-2.3	.71	-2.3	b-.64	.39	75.0	68.5	54	1
69	46	60	-.42	.33	.75	-1.6	.57	-1.8	a-.62	.34	80.0	77.2	69	1
MEAN	41.6	60.0	.00	.33	1.00	.1	.97	.0			74.3	74.8		
S.D.	11.6	.0	1.05	.08	.13	1.0	.26	1.1			11.0	9.5		

6.3.6 Post mock exams interviews

6.3.6.1 Is the UEE able to test your language ability?

Two interviewees, Daisy and Danny, believed that the University Entrance Exam was significant and authoritative in China. They also thought that the UEE covered a wide range of aspects: listening comprehension, vocabulary, grammar, cloze, reading comprehension, translation and guided writing. When asked if the UEE was able to test their language ability, Daisy and Danny answered:

Yes, I think the UEE is able to test my language ability, especially, vocabulary, reading and grammar. In a word, I think the UEE makes perfect sense. (Daisy)

Well it is, more or less. However, the point is that it only focuses on written English without spoken English, which is less comprehensive than IELTS. (Danny)

6.3.6.2 Content – the UEE and IELTS compared

Both participants thought that they had to read a whole passage very carefully to answer all different kinds of question items in the IELTS Reading Section, while they only picked up a certain part of the passage to answer the UEE questions by skimming and scanning. Moreover, question item types for the IELTS Reading Section were more varied.

Daisy thought the question types in the IELTS Listening Section were very diverse. The IELTS test involved note-taking, table or form completion, locating items on a map, and multiple-choice items. In the UEE the questions were mainly multiple-choice items for short conversations and passages, and two simple tables that involved filling in the blanks for longer conversations.

Danny felt that listening in IELTS was a good test because it presented a real environment and situations concerning English listening skills, whereas the UEE was merely for the examination's sake, and had no simulation. They both thought that it was not authentic at all. On the contrary, the IELTS Listening Section was able to test their listening skills. It may also contribute to future communication skills with foreigners, for those students who planned to study abroad.

6.3.6.3 Difficulty – the UEE and IELTS compared

When talking about the difficulty between these two tests, students thought that more vocabulary was involved in IELTS. It was not necessary for test-takers to understand all the words in IELTS as it was a proficiency test. However, for the UEE, students were informed of a list of vocabulary in the Vocabulary Handbook and that this list should be taught in school.

Danny thought passages were longer and sentence structures were more complex and the question item types covered whole passages in the IELTS Reading Section, while passages in the UEE were shorter and simpler.

Daisy believed the IELTS Listening Section was more difficult, while questions in the UEE were more straight-forward and simple. Also, longer conversations and passages sections were played twice in the UEE, while the tape was only played once in IELTS. She stressed that the listening section in IELTS covered a variety of situations and the speech was much faster than in the UEE.

In summary, both interviewees, Daisy and Danny, responded favourably to the University Entrance Exam and they agreed that the UEE was able to test their language ability; that is, they confirmed face validity. By comparing the content and difficulty between the UEE and IELTS, both participants had similar responses that question-item types for the IELTS Reading Section were more diverse and the content for the IELTS Listening Section was more authentic and practical. This information from the interviews partly supports the result that the correlation coefficients between the UEEs and the IELTS sample test were low; the concurrent validity of the UEE is comparatively low.

6.4 Conclusion

The distribution of the UEE scores is generally satisfactory. Both the 2008 and 2011 UEEs are all of a suitable level of difficulty; the distribution of scores is negatively skewed, which means that the UEEs were comparatively easy for the participants. The figures for both skewness and kurtosis statistics are within the range from -2.0 to +2.0, indicating that scores in both UEEs have a normal distribution.

The average facility values of the test items for both UEEs are 71 percent and 67 percent respectively, which indicates that both UEEs are very consistent. The mean UEE scores performed by the participants are 69 percent and 65 percent for the 2008 and 2011 UEEs; those figures are very close to the average score of 66.7 percent the UEE designers are trying to maintain. Also the values of the discrimination indexes are +0.30 and +0.36, which are very satisfactory.

The correlations between the subtests in both the 2008 and 2011 UEEs, which range from 0.480 to 0.699, show that each subtest is contributing some elements to the total test but no individual subtest could replace another; each subtest is testing different skills, hence satisfactory *construct validity*. The high correlations, over +0.850, between the subtests and total UEE scores reveal that the test components significantly contribute to the final total UEE scores.

On the other hand, while a *concurrent validity* study for total score was carried out, the individual listening or reading comprehension sections were not subject to such an exercise. The total scores of the 2008 and 2011 UEEs have quite low correlations (+0.508 and +0.435) with the IELTS sample test, which show that the *concurrent validity* is low.

The *face validity* of the UEE was judged by interviewing two participants following all the mock exams. They were asked if the UEE was testing their language ability and to compare its content to the IELTS sample test, a standardised international English test. Both interviewees believed the UEE was able to test their English language ability, especially the reading comprehension section.

Three different reliability coefficients – Spearman-Brown reliability coefficients, Cronbach coefficient alphas, and KR-21 reliability coefficients – show that the *internal consistency* is quite high for both the 2008 and 2011 UEEs, with indexes of above +0.870. On the other hand, the correlations between the 2008 and 2011 UEEs for different subtests show the *stability and parallel-form reliability* of the test, indicating how reliable the UEE scores are over a period. The correlation coefficients are from +0.332 to +0.706, which are quite low. Figure 12 shows the major test qualities – validity and reliability – are significant in language testing; especially a high-stakes exam, like the University Entrance Exam.

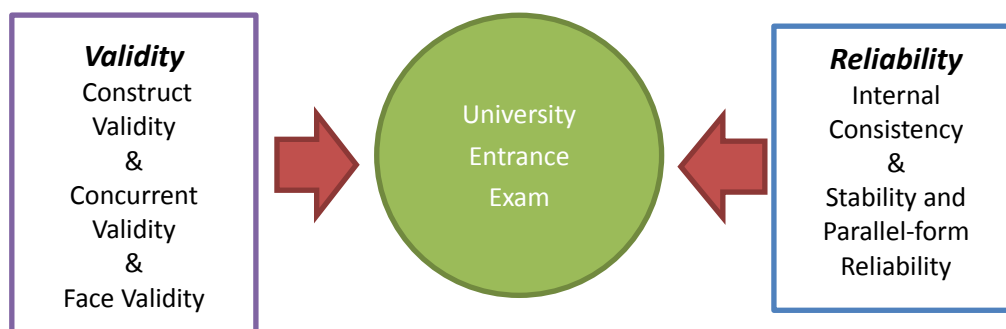


Figure 12: Test qualities – validity and reliability

The Rasch measurement is a sophisticated model to demonstrate the relationship between the test-takers and the difficulty of an item in a test. The person-item maps show that the UEE is neither too difficult nor too easy for the sample, but well-matched, which is very consistent with the figures obtained by the classical item analysis. The item-fit function by the model can also show how well each item in the test fits inside the underlying construct. Rasch analyses also provide evidence for the content aspect of construct validity, which refers to content relevance, representativeness, and technical quality. The resultant data show that most of the UEE items are within the acceptable range; the 2008 and 2011 UEEs reveal a reasonable degree of content relevance, a high degree of representativeness, and a very satisfactory degree of technical quality.

This chapter has summarised the present study's findings, and discussed them with reference to the evidence of two test qualities: validity and reliability. The findings have confirmed the satisfactory construct validity, low concurrent validity and satisfactory face validity of the English Language Exam for university entry in Shanghai. In addition, the reliability coefficients have revealed that for both the 2008 and 2011 UEEs, the internal consistency is reasonably high. However, the correlations between the 2008 and 2011 UEEs for different subtests have shown that the stability and parallel-form reliability is comparatively low.

Chapter 7: Consequential validity of the English Language Exam

7.1 Introduction

It has been claimed that language tests have an influence on teaching and learning independent of pedagogic purposes (Alderson & Hamp-Lyons, 1996; Hughes, 1989; Morrow, 1986; Peason, 1988). This influence, if it does occur, takes place in classrooms, and is commonly known in language testing as 'washback' (Alderson & Wall, 1993; Messick, 1996). Some consider 'washback' as a positive impact on teaching and learning, and indeed argue that testing should drive both curriculum and instruction (Fredrickson, 1984; Popham, 1987). Others argue that tests have an effect of narrowing the curriculum (Linn, 1983; Madaus, 1988). Wall and Alderson's (1993) study on examining washback revealed that washback had both positive and negative effects with an overall neutral result (pp.46-47; see Table 55).

*Table 55: Different Consequences between Positive and Negative Washback
(Wall and Alderson, 1993, p.46)*

	Positive Washback	Negative Washback
Content of teaching	Teachers would "teach the textbook".	Teachers would "teach the textbook selectively".
Method of teaching	Teachers would use the general approach and methods suggested by the Teacher's Guides.	Teachers would use whatever methodology that helped students for the examination.
Approach of assessing	Teachers would write tests which mirror the textbook's content.	Teachers would write tests which were similar to the content of past examination papers.

Washback is considered as a part of testing consequences relevant to assessing validity (Messick, 1996). In this study, the term 'impact' is the preferred usage indicating a powerful effect something has on a situation or person. The examination has an impact not only on teachers and students in the classroom, but also on the school, the educational system, and wider society (Hamp-Lyons, 1997). Such impacts are considered as the consequential validity of the examination (Messick, 1989, 1990, 1996).

Messick (1989) proposes "a unified validity framework" to reveal validity as a unitary but multi-faceted concept with "construct validity" but with an extra element in each facet (p.20; see Table 56). He explains that the consequential aspect of construct validity comprises "evidence

and rationales for evaluating the intended and unintended consequences of score interpretation and use in both the short-term and long-term” with positive and negative impact on teaching and learning (Messick, 1996, p.251); consequential validity then is a “complementary form of evidence to be integrated into an overall judgment of construct validity” (Messick, 1998, p.37). Cumming (1996) supports Messick’s idea and points out that the social consequences of test use must be evaluated to measure the long-term implications of the implementation of a test.

Table 56: Progressive Matrix for Defining the Facets of Validity (Messick, 1989, p.20)

Source of justification of testing	Function of outcome of testing	
	Test Interpretation	Test Use
Evidential Basis	Construct validity	Construct validity + Relevance / Utility
Consequential Basis	Construct validity + Value implications	Construct validity + Relevance / Utility + Value implications + Social consequences

The purpose of this chapter is to investigate the consequential validity of the University Entrance Exam on English language teaching and learning in a senior high school in Shanghai. The study included content, beliefs and behaviour of both teacher and students based on class observations, interviews and materials taken from the classroom. In addition, an interview was conducted with one of the UEE designers in which questions were asked about the UEE test development process and what he thought about the University Entrance Exam. A description of the study, analysis and findings from these classroom observations, interviews and artefacts, and a discussion of the nature of the impact of the UEE on society are then presented. The conclusion summarises the impact of the UEE on teaching, learning, textbook, the curriculum, and assessment itself.

7.2 Research method

7.2.1 Purpose

The major purpose of classroom observations and interviews was to ascertain the consequential validity of the University Entrance Exam; that is, to what extent the UEE influenced the ‘what’, ‘how’, and ‘how much’ of teaching and learning in the classrooms. A

teacher's class for one unit of teaching was observed and audio-recorded, and real-time notes were taken. Post-observation interviews were carried out – the teacher was asked why she was teaching particular materials and the students were asked what they did after class. Teaching materials including testing syllabi and textbooks were collected for analysis; they were examined to determine the impact on teaching and learning.

7.2.2 Participants

One senior high government school, located in Central Shanghai City, was selected. One of the senior II classes was selected randomly. All 40 students were native Chinese speakers and over 16 years old. English was considered as a foreign language in school. They were to take the UEE 15 months after the class observation took place. A Chinese teacher, Bonnie (pseudonym), taught two classes of senior II in the high school. She taught two lessons of 40 minutes every day. She had been teaching in this high school for eight years, since graduating from a university in Shanghai, majoring in Teaching English. Consent from all participants, including Bonnie, was obtained on a signed form.

7.2.3 Instruments

The teacher's lessons for one teaching unit were observed for about two weeks. Each lesson lasted 40 minutes. Before the beginning of the class, an audio recorder was placed on the teacher's desk to record the best effect of her voice. A worksheet template included today's topic, language skill and knowledge, materials used in class, time of each task, any specific notes related to the Exam, and any particular questions needing to be clarified in the interviews. Notes were jotted down during the class observations. The notes included any particular student activities and what materials the students were using. Particular attention was paid as to what special language (if any) the teacher used about the UEE. While the researcher was taking notes, the exact time was being marked so that, when checking the recordings, the researcher could trace the exact wording the teacher used in class.

7.2.4 Materials

The school had a standard teaching textbook for students, *The Oxford English for Senior II*. This textbook was written by native English speakers, published by Oxford Publishers, and approved by the Shanghai Municipal Education Commission. There were a total of six units in one textbook. The textbook came with an exercise book, the content of which followed the

textbook, with exercises including grammar, vocabulary, and translations. There is also a supplementary exercise book called *English Cloze Test for Senior II*, which only covers cloze exercises. Besides these, the students subscribe to a newspaper – *English Coaching Paper for Senior II*. Teaching materials were given to the researcher for analysis.

7.2.5 Procedures

The set up was ready prior to the 40-minute class observation. The researcher sat at the back of the class without interrupting the class. The researcher also told the teacher that he would not ask any questions or be involved in any class activities while the class was observed and audio-recorded. The researcher took notes and marked down class activities and specific comments. If there were any questions that needed to be clarified with the teacher or students, the questions were tagged and raised in the interviews. The researcher wrote detailed descriptions of each lesson. When checking whether the UEE might have an impact on the content of teaching in the lesson, he observed if the teacher was using the textbook or materials following the English Curriculum Standards or if she was using some other materials or sources related to the UEE.

7.2.6 Post-class observation interviews

After each of the five classes, the teacher and one student were invited for a post-class interview. The five students were Alan, Betty, Candy, David, and Eric (pseudonyms). Short interviews with the teacher after class were arranged in school. The questions were about whether the teacher followed every task provided in the textbook and how much she thought what she taught was related to the UEE. The students' interviews were arranged and conducted at lunch time, after class, in school. The students were interviewed separately. The questions covered the students' learning beliefs and behaviour. The interviews with both teacher and students were in Chinese, were semi-structured and structured respectively, and lasted up to 10 minutes. Interview questions are listed in Appendix E.

7.2.7 Interviews with UEE designer

The consequential basis of test use is "the appraisal of both potential and actual social consequences of applied testing" (Messick, 1989, p.20); Reckase (1998) believes that this appraisal contains problems for the test developer. After class observations and post interviews, the researcher had an opportunity to meet one of the UEE designers from Shanghai.

His name is Andrew (pseudonym), and he had previously constructed the English Language Exam paper for university entry in Shanghai. The interview was in Chinese and semi-structured, and lasted about two hours. The questions in the interview (See Appendix E) were about the test development process, the characteristics of the UEE, the impact on society, the traditional English teaching in school and his comments on the quality of the examination.

7.2.8 Data analysis

Classroom observations were analysed both quantitatively by looking at the time allocation in the teacher's teaching content in class and qualitatively by focusing on the teacher's teaching beliefs and behaviour when teaching. Interview data, on the other hand, were analysed qualitatively in order to reveal beliefs and behaviour of both teacher and students in view of the University Entrance Exam, and the UEE designer's experience. The audio-recordings of classroom observations and participants' interviews were transcribed and translated into English. The transcripts were then coded, analysed, summarised qualitatively, and categorised into different topics, and, hence, used to analyse classroom observations and interview data.

7.3 Findings and discussion

7.3.1 Content of teaching

Table 57: Time Allocation for Different Teaching Areas in Class

(Time in minutes)	Day 1	Day 2	Day 3	Day 4	Day 5	Total	By %
Listening Skills	0	0	0	0	0	0	0
Grammar & Vocabulary	5	15	5	29	10	64	32
Cloze	10	0	10	0	10	30	15
Reading Skills	21	5	25	5	10	66	33
Translation	4	20	0	6	10	40	20
Writing Skills	0	0	0	0	0	0	0
Speaking Skills	0	0	0	0	0	0	0
Total	40	40	40	40	40	200	100

The researcher observed a unit of teaching for five days, audio-recorded the lessons and real-time notes were taken. Each lesson lasted for 40 minutes. The times allotted for different teaching areas are listed in Table 57. Summing up the five observation days, the teacher (Bonnie) spent 33 percent of class time on reading, 32 percent on grammar and vocabulary teaching, 20 percent on translation practice, and 15 percent on cloze exercise practice. She did not spend any time on writing, listening or speaking; this could have been an impact of the UEE, in that about half of the examination is focused on the grammar and vocabulary, cloze and reading comprehension sections.

The following sections show how the UEE affected the content of teaching in terms of grammar and vocabulary, cloze, reading, translation and writing, listening and speaking, and monthly tests in the classroom. Supplementary information was provided by interviewing the teacher, and examining teaching materials she gave to the researcher during the classroom observations.

7.3.1.1 Grammar and Vocabulary

The teacher spent about one-third of the class time on teaching grammar and vocabulary. She explained to the researcher that learning grammar and vocabulary not only helped with that section, but helped students understand vocabulary in reading comprehension and learn how to use correct vocabulary and grammar in both the translation and guided writing sections. When asked if she spent much time on teaching vocabulary in class, Bonnie replied:

Those words I taught in class may be seen in the translation or reading comprehension sections. So I have to teach them those words. Those words are key words and very likely appear in the UEE. (Bonnie\Day2\02'18")

At the beginning of the class on Day 4, Bonnie started her lesson with a dictation. She said that the purpose was to see to what degree her students had mastered the key words and the text in class. She said that the dictation task was not directly related to the UEE; however it was a very important exercise to build up students' vocabulary. The dictation was mainly the words from Unit Two, a unit that was more related to the textbook than to the UEE. However, it was found that the key words were not only from the textbook, but also part of the Examination Specifications, which meant students were required to understand the UEE.

It was noticed that the teacher was checking the answer keys in the Unit Two grammar exercise of noun clauses with the students. Bonnie told the researcher that every unit had a special subject with a certain grammar point. The grammar points were included in the textbook. She thought that the grammar points were designed in the course according to the English Curriculum Standards. Bonnie said that she had been teaching grammar in high school from senior I until the first term of senior III. She had to reinforce her students' grammar. By senior III, their grammar foundations would be strong enough.

Bonnie mentioned to the researcher that one part of the UEE had changed in 2014. There would be no multiple-choice questions for the grammar section; they were replaced by two cloze passages requiring the students to fill in the blanks according to the given key words or context. She commented:

I think this change brings more difficulty to students because, in the past, students could guess the answers if they could not choose the right one ... They used to spend only 10 minutes to finish this part but now it takes them a longer time. They have to understand the context and content, and then you have to think what kind of words should be filled into the blanks. Conjunctions? Adverbials of time? Attribute clause? You have also to change the key words according to grammar. It is more difficult now. (Bonnie\Day4\02'21")

7.3.1.2 Cloze

Bonnie admitted that the cloze section was the most difficult part of the UEE paper and a very crucial part of the UEE; it accounted for about 25 marks and 15 marks respectively in the Grammar and Vocabulary and Reading Comprehension sections. Since her students did not perform well in this part, Bonnie asked them to work on the *English Cloze Test*, which had been chosen by the English teachers in her school, three to four times a week. That exercise book had the same format of exercises as the UEE. She thought it was a good opportunity for her students to practise the cloze exercises in order to get higher scores. Bonnie usually spent the first 10 minutes nearly every day on checking and reviewing the answers of the cloze exercises with her students in class. When she gave the correct answers, she asked which answers her students chose and why they chose such answers:

After telling them the correct answers I will ask them which one they chose and why they chose the wrong ones. Then I can find a way to rectify their mistakes and change their thinking process. [Is it also a strategy?] Yes, it is.

(Bonnie\Day3\01'20")

Bonnie also realised the cloze section was the weakest point for her students. Therefore, she always spent more time on this part and let her students practise more on the new UEE format. It was noticed that she always asked her students to pay attention to that and spent more time on reviewing the exercises. She had to ensure that her students remembered and knew how to answer the UEE questions. That was why they had extra supplementary exercises from the *Coaching Paper* in addition to the *English Cloze Test* exercise book. Since the newspaper and the textbook were from the same publisher, the teachers in school thought the newspaper could help supplement the content from the textbook and reinforce students' knowledge. She added:

English teachers from the same grade discussed and ordered this newspaper [the Coaching Paper] because we want our students to consolidate the language knowledge they learned in the textbook. That's why we ordered the newspaper.

(Bonnie\Day4\01'04")

7.3.1.3 Reading

The teacher spent 33 percent of her teaching time on reading skills. This matched the reading comprehension section, which accounted for one-third (50 points out of 150) of the UEE. She told the researcher that more time should be spent on this section so that her students would be able to get higher marks in the UEE. The teacher mentioned that students could be introduced to reading strategies and taught how to choose the most appropriate answers for the questions. She said:

I took the University Entrance Examination myself. According to the distribution of the marks in the examination paper, the part of reading comprehension accounts for the most points in the UEE. Therefore I hope that my students can enhance their reading ability. We only practise a little about reading comprehension in class so I hope they could practise more in their spare time.

(Bonnie\Day5\00'20")

On Day 3 of the classroom observations, after Bonnie checked answers with her students on the cloze exercise for the first 10 minutes, she started to focus on the reading passage of Unit Two in the textbook. She went through the passage with students in the beginning by playing the tape. Then she analysed the structure and the content of each paragraph. The researcher noticed that she taught “skimming and scanning” when going through the third and fourth paragraphs. Finally, Bonnie summarised the main idea of each paragraph with her students.

There were three exercises following the passage: vocabulary gap-fillings, true and false statements, and question discussions. Bonnie finished only the first two exercises but did not go through the last speaking activity. She explained that vocabulary and true/false exercises were more related to reading comprehension and also time was limited in class (private conversation during lunch time, 10 March, 2014). Nevertheless, the teacher gave the researcher the impression that her teaching was very teacher-centred; it was hard to see any student discussions in class.

7.3.1.4 Translation and Guided Writing

From the classroom observations, it was noticed that the teacher spent 20 percent of her class time on translation. She told the researcher that translation practice was a good way to learn how to write sentences. Through translating Chinese into English sentences, students could learn how to make use of new words and correct grammar to write long sentences. The more they practised, the easier the students found it to write sentences.

Bonnie explained to the researcher that her students did not have enough vocabulary and grammar knowledge to write a good essay. She thought it was a waste of time to practise writing in senior I and II. She would prefer to spend more time on teaching vocabulary and grammar, and practising translation exercises. Then the students would have a better foundation to write sentences, and, of course, better essays. Not until senior III were students taught how to write in class; the topics given would be very similar to those in the UEE.

Students were given translation tasks to practise their knowledge of vocabulary to ensure that they knew how to use the words. Bonnie said that students had to know how to use the words first. Then they could know how to translate the sentence correctly so as to get higher scores in the UEE. In her opinion, grammar and vocabulary were always involved in translations. From

her students' performance, she knew more about her students' language standard. Therefore, working on a translation exercise was a good way to train her students.

Bonnie emphasised that translation tasks could train students' integrated writing ability. She thought if students translated well then it would be easier for them to practise other forms of exercises well. Bonnie explained why she had not yet started to teach writing in class:

This is actually a problem because I asked them to write before but they could not write well. There were many grammatical mistakes and the organisation and structure of the essays were not good. There was no content for what they wrote. Therefore, I mainly focus on the translation part. In fact, the translation part is helpful to their writing. If they can translate well, they can write articles more easily. So we usually practise translation first and then if there are some texts suitable for writing, e.g. there are texts for writing letters and story-telling, I will ask them to write but not very frequently. Writing will be left for teachers in senior III. (Bonnie\Day5\03'09")

7.3.1.5 Listening and Speaking

The materials used in the school included *English Oxford Shanghai Edition for Senior II* (Textbook) and its exercise book, *English Cloze Test for Senior II* and *English Coaching Paper for Senior II*. For the school textbook, there were a total of 16 pages in Unit Two and this unit consisted of Reading, Speaking, Writing, Listening, Grammar, Vocabulary, and Cloze. From the class observations, the researcher noticed that six pages in the textbook Unit Two covered listening, speaking and writing skills, but Bonnie did not spend any time on these skills in class. When asked why she did not follow every task in the textbook, Bonnie replied:

We only choose some parts of the textbook. Because our students do not have a strong foundation for English learning, if they focus on every part of the textbook, the time will not be enough. The listening and speaking sections, sometimes even some of the writing parts in the textbook, are not related to the UEE. And it is difficult to see students make great progress after this kind of practice. Therefore we ignore those parts. (Bonnie\Day5\01'46")

In school, students had monthly tests; that was the only time they had opportunities to practise listening, the format of the monthly tests being exactly the same as the format of the UEE. The teacher evaluated the answers of the monthly test. She explained and taught the

students strategies for choosing and finding the answers for the listening conversations and passages. In the second term of senior III, students would spend more time practising for the exam; therefore, they would spend more time on listening. When asked if she would teach her students some strategies to cope with the UEE by then, she firmly answered:

Yes, of course. According to different types of question-items and test-tasks we will have different lectures to help students. For example, in the part of listening comprehension, we will teach them how to guess or how to understand the implied meanings of the conversations. (Bonnie\Day5\04'42")

Since speaking was not compulsory and only the marks of students applying for the English Major in university would be considered, teachers usually did not spend time in practising speaking tasks in class. The teacher usually started to 'train' her students on speaking exercises according to the Oral English Test format in class three months before the Exam, which was held in December when students were in senior III. When asked how she would help students cope with the Oral English Test, Bonnie answered:

In senior III, before the Oral English Test, we will let the students focus on this part. We will ask them some questions in class, also print many daily talks for them and ask them to recite them and go to the computer rooms to become familiar with the software for the Oral English Test. (Bonnie\Day5\02'38")

7.3.1.6 Monthly tests

In the class on Day 5, the researcher noticed that the teacher was checking out the answers with her students. She told the researcher that it was a test related to Unit Two; it was also the monthly test. The monthly tests, designed by university professors and high school teachers, were provided by the publishers of the *English Coaching Paper*. The content of the test was the same format as the UEE. Only the translation section in the test was related to the textbook and contained key words taken from the textbook. Listening comprehension, grammar and vocabulary and reading comprehension sections were not related to the textbook, but to the UEE. The students were practising for the UEE, becoming familiar with its format. When asked if the test this month was related to Unit 2 of the textbook and what the question-items were like, Bonnie answered:

Only the translation part is closely related to Unit 2 which contains the key words from the textbook. Listening comprehension and reading comprehension are irrelevant to the textbook. The question-items are similar to those in the University Entrance Exam. (Bonnie\Day5\01'09")

In summary, about 80 percent of class time focused on the grammar and vocabulary, cloze and reading comprehension sections. This is similar to Lam's (1994) study which found that teaching emphasised those parts accounting for most marks in the examination. When the teacher found any tasks in the textbook irrelevant, for example, listening and speaking tasks, she would ignore those tasks in class. The only opportunity to practise listening tasks was using the listening section in monthly tests. Since the Oral English Test score was not counted in the University Entrance Exam, the teacher usually trained her students in class three months before the Oral English Test. The teacher also thought her students did not have enough vocabulary for writing. Therefore, the best way in their senior II was to practise translation so as to learn more words and how to write sentences. For the materials used in her class, other than the textbook, she used *English Cloze Test* three to four times a week. That exercise book had the same format of exercises as the UEE which was considered "examination oriented" (Spratt, 2005, p.10).

7.3.2 Teacher's beliefs and behaviour

The UEE not only had a huge impact on teaching content, but it also affected the teacher's beliefs and behaviour in the classroom. It was easily observed that what the teacher said in class was always related to the UEE directly or indirectly. Some examples are listed below:

1. The teacher taught some important phrases, or collocations from the textbook or exercise books. She thought those phrases might appear in the UEE; therefore, she would remind her students and ask them to remember them or even memorise the phrases:

*We have this phrase – stick to something (黏住, 卡住). This is a very important phrase, you **have to** memorise it – stick to something. (CO\Day2\13'38");*

2. The teacher told the researcher that translation was an essential part in teaching. If students managed the translation well, she would know how much they understood the vocabulary they learned in class. When the teacher was checking the answers in the

textbook, she would sometimes ask the students to translate some English sentences from the textbook into Chinese in order to practise translation:

The government successfully reversed the country's economic decline. *How do you translate this sentence into Chinese? The government successfully reversed the country's economic decline* (政府怎么样成功的逆转国家的衰退). Reverse; *reverse* (逆转). (CO\Day2\24'00");

3. When the teacher was checking the answers in the cloze exercise, she explained the content of the passage in Chinese and told students why the particular answers were chosen, and not others. For example, in the *English Cloze Test* pp.15-16, the passage was about the exchange programme between China and the U.S: "Our (9) goal is not to create volume, but to provide opportunities for people who are going to be future leaders. That's why we (10) select the students very carefully." The teacher explained to the students why the answer on (10) was not C (elect), but B (select):

Item No.10 is B select; select ... Therefore, no.10 is select. What answer did you choose? (Some students said C). C elect; elect (选举). Understand? Election; election. Presidential election; presidential election (总统选举). B select; select (挑选). Select leaders; select leaders (挑选领袖). (CO\Day3\07'56");

4. The teacher usually had a 'Dictation' after finishing a unit of teaching. In the beginning of the class on Day 4, the students had a dictation of Unit 2 Vocabulary. She said to the researcher that the purpose of having a dictation was to learn to what degree her students had mastered the key words and textbook. At the end of the dictation, the teacher said to her students:

We will have the dictation again at lunch time tomorrow for the same content. The new words from the textbook were very important in the UEE. (CO\Day4\13'12");

When asked why she would have the dictation again the next day, Bonnie sighed and explained:

I hope to reinforce my students' vocabulary, which is very useful, I believe, on reading comprehension and translation and writing in the UEE. (Bonnie\Day4\00'28");

5. When the teacher reviewed the answers to the translation in the *English Coaching Paper*, she always asked her students to “remember that (记住啊!)”. She urged them to remember the collocations because she told her students that they might probably see the same collocations again in the mid-term examination. It was interesting to note that the teacher asked her students to remember particular words because of the examination:

The doctor should not put emphasis on the importance of losing weight! (医生不应该过多地强调减肥的重要性) Emphasis! Please tell me what is the collocation of emphasis? Do you still remember? [A student said, ‘emphasis on.’] What emphasis on? What is the verb? Three verbs – we may use ... put / place / lay (emphasis on). The word ‘emphasis’ – I am 80 percent sure that it will appear in the mid-term examination. (CO\Day4\35’03”);

6. Students had the monthly test the day before the Day 5 class observation. Before the teacher evaluated the test paper with the students, she mentioned which part of the UEE was more important than the others. The teacher emphasised that the total score of the UEE was 150 – reading comprehension and cloze passages together accounted for one-third of the total score. She asked her students to put emphasis on these two sections and remember what she had just said:

The total score of UEE is 150. Passages in the reading comprehension items account for two marks per item. Also longer conversations and passages in the listening comprehension section also account for two marks each item. Then total score for the translation section is 20 marks and for the writing section 25 marks. Are you clear? (明白了吗?) Therefore, we have to remember! Which sections do you think are easier to lose marks? ... Reading comprehension and cloze – the total score for these two sections is 50 marks. Be careful! (CO\Day5\00’23”);

7. After the teacher analysed the score allocation for each section of the Exam, she announced the scores of the monthly test to the students, telling them their class (Class 3) performance was not satisfactory. She told her students that only five students had passed this test, compared to the 14 people who passed in another class, much more students than their class. Before she reviewed the answers with her students, she

reported to the class which five students passed the test and their scores:

The scores you got in your class are not really satisfactory! Am I right? Compared to Class 4, there is a big difference. Guess how many students passed the monthly test in Class 4? There are 14 in Class 4 and only five in your class (Class 3). There is a big difference as I said! Am I right? When I speak in class, you should listen carefully and jot down the notes. If you have any questions, feel free to ask me after class. Only five classmates passed the test in your class. The best one in your class is classmate xx – 78 marks. She always gets very high scores in dictations and monthly tests. She is very industrious, isn't she? (CO\Day5\01'23");

8. After she finished the class on Day 3, the researcher had opportunities to talk to the teacher and asked her what she thought about the Curriculum Standards and Examination Specifications. Bonnie studied the Curriculum Standards (CS) when she was an undergraduate, and when she was doing her masters degree in education, including a course about the National Curriculum Standards. When asked how much she understood about the CS, Bonnie replied to the researcher:

Uh...I looked though the whole content and know what should be taken into account in my teaching. [For example?] Reading, listening, writing and speaking. We should know to what degree the student can manage language, to what extent we [teachers] should extend one certain knowledge point and which aspect should be paid more attention to. (Bonnie\Day3\01'50");

Apart from the language skills and language knowledge, teachers should also teach about culture, affect and attitudes as embodied in CS. However, she added:

I think they are very abstract and they only build a general frame for teaching but different schools choose different textbooks. They cover little about the real teaching in schools. They are just a frame to guide teachers as to what goals should be achieved and in terms of the way to teach and the strategy to adopt; they don't help us a lot. They involve little about the more practical examples. (Bonnie\Day3\02'43");

The Examination Specifications were about guiding teachers and students in respect of the content, the format and the purpose of the UEE. They included listening, speaking, reading, writing, grammar and vocabulary. The Specifications gave some examples of the scope of the Examination. A list of vocabulary words and phrases that would be

tested in the UEE was also attached, meaning that students should understand and memorise those words before the UEE. When asked whether her main focus in preparing for the lesson would be the Curriculum Standards or the Examination Specifications, she answered with confidence:

The Examination Specifications, of course, would be my main focus because the Curriculum Standards are too abstract and not very detailed. The Curriculum Standards do not mention which words are important and to what extent they should be mastered. The standard is not very clear. (Bonnie\Day3\02'43");

However, Bonnie thought the examples given in the Examination Specifications were too limited; they only demonstrated some of the phrases. She did not think it was enough to just teach students vocabulary in class, feeling that the information given should be more detailed. She was disappointed that the guidelines were not enough. When she prepared her teaching lessons, she had to use different materials and exercise books, ones that were close to the content of the UEE. Bonnie explained:

The Examination Specifications do demonstrate some simple grammars but the relevant exercises and the extension of the grammar is lacking. They only show some examples. The examples are limited and they only demonstrate some of the phrases. I don't think it is enough just to teach students vocabulary. (Bonnie\Day3\04'20");

When the researcher asked what she would do when preparing the lessons, she added:

We have relevant and good exercises, very close to the content of the University Entrance Examination. Those exercises are our emphasis for students. (Bonnie\Day3\05'02").

In summary, from the above examples observed in the classroom and interviews with the teacher, it was very apparent that what the teacher did or mentioned in the class or interviews were either explicitly or implicitly related to the University Entrance Exam. When she was teaching reading from the textbook, she had to ask her students to memorise the vocabulary or practise translation for some of the useful sentences in class. When she finished a unit in the textbook, Bonnie had a dictation in order to build up her students' vocabulary. When checking the cloze exercises, Bonnie taught her students strategies on how to choose the correct answer and analysed how important the cloze section was. The most surprising

phenomenon was that Bonnie announced every student's monthly test score in class. She also explained to the researcher that she had to use extra exercises, which were close to the content of the UEE, since the Curriculum Standards and the Examination Specifications provided limited information for teaching.

7.3.3 Content of learning

7.3.3.1 Materials learned in class

The students thought materials learned in school were enough for them to cope with the UEE. For instance, the phrases learned in class were very useful in the translation section. The teacher also trained them in grammar with exercises. Students believed that it was adequate to tackle the UEE. Most of them completely agreed that the materials they learned in most of their lessons, such as vocabulary, phrases, grammar and cloze, and translations, were related to the UEE. One of the students thought that only vocabulary and grammar were important for the UEE, because the UEE mainly emphasised comprehension and vocabulary. When asked if what the students learned in class was enough for or related to the University Entrance Exam, David and Eric answered:

What we usually learn in class are mostly grammar and collocation, which are exactly what will be tested in the examinations together with translation. So I think they are definitely helpful. (David)

The methods or ways of answering the questions and the grammar and vocabulary for the translation part are very useful for the UEE. (Eric)

7.3.3.2 Strategies learned in class

Students thought the methods or techniques they learned in class and the new words they learned would be very helpful for future UEE. When asked how the students would learn any techniques to deal with the cloze passages in the UEE, Alan and Candy responded:

We should read the entire passage first before we fill in the blanks. And we may look at the context to see if it has any relation to the options in the questions. When making decisions, we will judge from both grammar, such as collocations of prepositions like 'with' and 'to', and meaning, which may be enough to make correct choices. (Alan)

Practising cloze exercises in class is of great value; we are able to gain knowledge of new words. Also we learn some techniques, like how to judge the parts of speech of the missing words or understanding the content of the passage. (Candy)

One of the students realised that reading comprehension learned in class was a part of the UEE; she knew that although the passages would not be the same, the techniques and methods she acquired in class were similar and worth accumulating and learning. Except in the reading comprehension exercises, her teacher often told them how to define parts of speech with missing words or how to relate the questions to the context in passages.

In summary, in the students' views, most of the materials they learned in class were related to the University Entrance Exam and they thought the content they learned was sufficient for coping with the UEE. They were also able to learn some techniques and strategies on how to answer the question-items for the UEE in class, which they believed would be very helpful.

7.3.4 Learner's beliefs and behaviour

7.3.4.1 UEE impact on student lives

Students usually reviewed new words and several sentence translations learned from the class at home. They tried to remember how to use the new words and focused on how to use them in sentences, for example, in the application of phrases. They also reviewed the key answers in the monthly test in class, the format exactly the same as that in the UEE. They would go over the entire paper and pay specific attention to the mistakes in the test. If they had any problems, they would go to ask their teacher the next day. Most of the students believed that what they learned in class was very important. When asked what they would usually do after school, Candy and David answered:

I usually take notes in class. So normally I can review the word usages and the main ideas of the texts after school. I think cloze and reading comprehension are more significant, especially the latter mentioning how to construct and organise English compositions. She [her teacher] helped us analyse the structure of English essays, which will be definitely useful for our writing. (Candy)

For dictation, I will correct the answers and examine my mistakes, and for correction, I will review the notes about some key words and collocations that I have taken during the class. (David)

7.3.4.2 The teacher's change of focus as the UEE drew closer

Students believed their teacher's focus in class would change as the UEE approached. In senior II they were learning new texts from school textbooks, but the next year (senior III), they would be reviewing the lessons they have learned in the entire three years and practising more test papers in order for the UEE. Students also expected more intensive classes were to come as the training included grammar, cloze and listening comprehension exercises. Candy and Betty thought their teacher's focus on teaching in class would change in senior III:

I think she will pay more attention to revision and consolidation next year, compared to the texts we have learned in the first two years in order to make students triumph in the UEE. [What exactly would be revised and consolidated?] It includes the vocabulary, applications, grammar, and so on, along with some more new words and grammar knowledge. (Candy)

There may be more intensive training. [In what aspects?] In several aspects such as grammar, listening, and so forth, all for the University Entrance Exam. (Betty)

7.3.4.3 Private English tutorials

It is a common practice that students in China attend private tutorial classes outside school in order to get higher scores in the UEE, and those tutorial classes cost extra and are usually paid by their parents. During the interviews, the researcher wondered if the participants attended any English private tutorials. The five participants had different responses with respect to this topic.

Alan told the researcher that he did not attend any private English tutorials outside class because he thought his school work and revision after school were adequate to deal with the exam. He guessed the tutorial materials would be similar to what he learned from school.

Betty said that she attended private English tutorials after class. The tutorial was run by an experienced English teacher and held at her home. There were about eight students in a class. She said that the materials, for instance the vocabulary and reading passages, learned from the tutorials, were more difficult. She explained that perhaps the tutor taught in a key school in Shanghai, since the private tutorials covered what she could not learn from school. Betty felt that there were some similarities, such as the skills for reading comprehension.

Candy did not take any private English tutorials outside the class; she thought her English language ability could cope with the lessons at school. In other words, she was able to follow what the teacher said in class for the moment. She guessed that English private tutorials mainly focused on exercises, while classes in school paid more attention to texts from the textbook. She would prefer to learn English in school because she thought the exercise-training mode in private tutorials were of very little help in other fields except the University Entrance Examination, as the English language needs comprehension above all.

David was attending a tutorial established by a teacher from another high school. There were about six people in his class. Actually the tutorial was aimed at reinforcing what he had learned at school. The tutor was an experienced teacher from a prestigious high school; therefore, he knew what students were learning in school and he then strengthened their knowledge and taught them how to answer well in the examinations. The purpose of having private English tutorials was mainly to prepare students for the University Entrance Exam. David felt that there was not much difference between school and tutorial learning. They both placed most emphasis on grammar.

Eric did not attend any private English tutorials outside the class because he thought successful English learning required one's own efforts to keep the knowledge in mind. He believed the help from the private tutorials would be minimal. According to his experience, the tutors usually introduced to them some strategies and gave them a great deal of practice. He thought his teacher was doing the same thing so he did not go to any private tutorials.

7.3.4.4 English learning beliefs and behaviour if there was no UEE

Examinations play a significant role in the Chinese educational system. Children are obsessed with examinations once they start their school life (Gu, 2014a). They never think examinations could disappear in their lifetime. The researcher was interested to know what their English learning beliefs and behaviour would be if the UEE were cancelled.

Alan thought his English learning beliefs and behaviour would be the same if there was no UEE. He realised that English was definitely significant and it was impossible to find a decent job in foreign companies without good English skills, especially in contemporary Shanghai. He agreed English was a must for stepping out into the larger world. He added that what he was learning in school prepared him for his future profession as well as the UEE.

Betty, in contrast, would feel happier without UEE; she could learn in a freer way, as she could have more options and follow her heart (to learn what she liked to learn). She told the researcher that she loved art; she would pay more attention to the vocabulary of art in English instead of learning grammar. Nevertheless, she thought that what she was learning in class was very helpful and useful in her future study and career.

Candy would focus more on communication with foreigners and less on memorising vocabulary and grammar if there was no UEE. Candy would spend more time learning English according to her own interests and hobbies; she loved watching American TV series, as a way of learning how to carry out daily conversations. Candy thought the materials she learned in school were not enough for future study and career, only for the UEE.

David also thought his learning beliefs and behaviour would be a little bit different and, like Candy, he would have more free time to pay attention to English conversations rather than grammar in order to communicate better with foreigners in future. He definitely agreed that what he was learning from school would help his future study and career, as future jobs might require all kinds of knowledge and skills.

Eric was very honest to say that he was not interested in learning English. He would feel very relaxed and less stressed without the University Entrance Exam. He would focus less on English language and spend more time learning Chinese as he was fond of Chinese. He did not think the English language he was learning in school would help with his future university study or career because the future jobs would require more speaking skills, while what they had learned now was more about grammar, which was not helpful in English.

In summary, the University Entrance Exam not only had an enormous impact on the content of learning, but also affected students' beliefs and behaviour in their lives. Students reviewed materials the teacher covered every day after school. They believed their teacher's focus would change as the UEE approached. Some students attended English private tutorials outside school because they believed the tutorial would help and reinforce what they had learned at school. In most cases, it was interesting to find that their English learning beliefs and behaviour would be very different if the UEE were cancelled.

7.3.5 Interviews with UEE designer

Andrew (pseudonym) was an experienced English language teacher in Shanghai City. He had past experience in designing examination papers and had previously constructed the English Language Exam paper for university entry in Shanghai. The researcher had an opportunity to interview him and asked him about the process of setting up the examination paper and what he thought about the University Entrance Exam in Shanghai.

7.3.5.1 Test development process

The UEE designers were invited by the Examination Centre to design and construct the UEE paper every year. The UEE designers included some university professors and experienced high school English teachers. Unlike IELTS or TOEFL, the Examination Centre did not have a test bank; they set up their examination paper just one month before the UEE. Once the UEE designers accepted the invitation, they entered the camp, where they set up the examination and could not be reached. They were allowed to make phone calls when they were in the camp, but all calls were recorded. They were not able to access the internet in the camp. Since the University Entrance Exam was a 'National Exam', if the questions were leaked, all related staff would be under suspicion.

The designers prepared two sets of UEE papers: paper A and paper B. Paper B was a back-up paper. In case an unexpected situation occurred, paper B would be used. The UEE designers had some basic guidelines when they set up the UEE papers. The first guideline was the Curriculum Standards. The Examination Specifications, including the nature and objectives of the UEE measurement, the contents and requirements of the UEE, and question examples, were then formed according to the Curriculum Standards. Other than the Specifications, there was a vocabulary handbook – the University Entrance Exam Vocabulary Handbook for English Language. This handbook contained a list of recommended vocabulary that senior high school students were to learn and understand before attending the UEE.

Andrew explained to the researcher the general procedure of the examination operation. There were about six UEE designers. They prepared some materials for the camp. In the beginning, they selected suitable materials, then they set up the first draft. They would decide which materials they were going to adopt after a detailed discussion. They checked whether the materials were suitable in the UEE. If the materials were from one of the designers, he or

she would not be involved in setting up the questions. The aim was to keep a higher degree of confidentiality and fairness.

After the first draft was designed, all UEE designers had a group discussion. They went through the question paper again and again, and then looked at each item and discussed it in detail. They then had to check if there were any problems in the language itself, and check if there were any errors in the questions. For multiple-choice questions, they had to check whether the choices were reasonable. When they found some problems or errors, they would discuss and amend accordingly.

After the UEE designers finished setting up the questions, some other teachers were invited to join the camp as proof-readers. First of all, the proof-readers worked on the paper by themselves to see if there were any problems or mistakes in the UEE paper. They looked and checked the UEE paper from different angles. Then the proof-readers had a meeting with the UEE designers. The proof-readers discussed with the UEE designers when they picked up any errors from the UEE paper. When there were any errors or mistakes in the paper, the UEE designers would have to adjust and amend the question items.

The Examination Centre had its own piloting process, but the piloting size was comparatively small. There were some students from different districts in Shanghai who were current UEE candidates. After the students finished the piloting, they were asked to make comments on the UEE paper. The main concern was to check if they had ever done or read the same materials before. Then the students were asked if they had guessed any items in the piloting test, which items had given them a very strong impression, and how they felt generally about the whole exam paper. If there were any problems in the questions, the UEE designers would amend the question items in the UEE papers.

Andrew thought the interviews with the piloting students were very important. The students could not pass the piloting test if they had read the materials somewhere else before. Guessing answers was also not really acceptable. If they found quite a few students answered the same question by guessing, then the UEE designers had to check the question item again. Andrew thought that think-aloud protocols, which they did not use, would be a good tool to check the response validity of the UEE paper – they could see what students thought when they answered the questions and check if they were thinking what they were required to think. However, this was impossible to apply with these students, because they were current

candidates. Instead of taking the UEE in early June, the students took the UEE two weeks earlier in the camp. After the piloting, these students had to stay in the camp until the University Entrance Exam was finished. They did not need to take the UEE again in June. The results from the piloting would be their results.

7.3.5.2 Proficiency test versus achievement test

It has been debated whether the University Entrance Exam was a proficiency test or achievement test. Andrew believed the UEE was a selection exam. The major purpose of the English language exam for university entry was to select the best students who had learned their subjects well according to the English curriculum, the most talented students who would study academic subjects in universities, and potential candidates who were going to study subjects in English. Therefore, he thought it was not an achievement test, but proficiency test. Although there were Curriculum Standards, he said that the Standards were just used for reference. Not all the UEE questions were restricted to the Standards.

In Shanghai, there was another public examination called 'Shanghai High School Graduation Examination'. Senior III students would sit this exam before they graduated from high schools. Basically, this graduation exam was considered as an achievement test. Compared to the UEE, Andrew said that this graduation exam was comparatively easier. Students were required just to achieve the minimum requirements set by English Language Teaching Basic Requirements (SMEC, 2005). Since English language was a compulsory subject in Shanghai, all high school students had to pass this exam before they were qualified to graduate.

7.3.5.3 Impact on school and society

Andrew thought the UEE designers would consider the concept of 'wash-back' when they constructed the UEE paper. That was the reason why they changed the format from time to time. In the past there were many multiple-choice questions. For multiple-choice questions, students could easily get high scores, especially in the grammar section. Their teachers also taught them techniques to answer those questions. He thought most of the students in China were preparing for the UEE when they were in school.

When asked if there was more positive or negative impact on school, Andrew felt that there was generally more negative impact on teaching in school; teachers and students were preparing in class just for the UEE. In Shanghai, there were three grades in senior high schools.

In senior I and II, they were learning English in class under the high school curriculum. Then in senior III, they were practising the examination format papers for the whole year in order to prepare for the UEE.

The UEE not only had impact on students but also society. Andrew shared an interesting incident:

During the UEE period, some parents usually block the main road in order to minimise the traffic noise. If the date of the UEE is held on the “eighth” of June, you know, the “eighth” is a very popular date for people to start their business, business people will have fireworks to celebrate the opening, and the fireworks might affect test-takers’ performance. (Andrew)

7.3.5.4 Traditional teaching

Andrew thought that the purpose of learning language was to communicate with people. People might not have to pay attention to grammar when they communicated. There are several reasons why teachers in China focused on teaching vocabulary, translation and grammar in class. First, there was not much change in learning the English language in China; it was a traditional method of learning English. Second, to Chinese students, English was not their native language, also not an official language in China, so they did not have any language learning environment around them. They had very few opportunities to interact with English speakers, only learning English in class.

It was found that the instructions of the guided-writing section in the UEE paper were written in Chinese. Andrew thought that they were testing students’ writing ability but not reading comprehension, therefore students should be given written instruction in their native language and be told what to write. The UEE designers were afraid that some students might misunderstand the English instruction, which might affect their writing content. If the instruction was written in English, some students might not be able to write what they wanted to write because they did not comprehend the test-task.

7.3.5.5 A fair and good examination

When asked if the UEE was a fair examination, Andrew proudly answered:

Relatively speaking, the University Entrance Exam is comparatively fair. Firstly, when the UEE paper was constructed, the UEE designers would neither choose

any popular topics nor unfamiliar topics. We were trying to avoid picking any hot topics. We were afraid that it would not be fair if some students read those related materials and some did not. For reading comprehension passages, the UEE designers would choose some topics which were not experienced by students. Secondly, the UEE follows the Examination Specifications and Vocabulary Handbook and it is related to the Curriculum Standards. Therefore, if the students study in high school for three years, they should not have any difficulties in finishing the UEE. (Andrew)

Andrew added that the English Language Exam paper was a good paper:

Indeed, nothing was perfect; there must be some parts in the UEE which are not satisfactory. For example, the writing section – the UEE designers would like not only to test students' writing ability, but also critical thinking ability; however, the UEE designers are not happy with the guided writing topics they set. In reading comprehension, the UEE designers hope to have different kinds of question types instead of multiple-choice and short-answer questions in order to test students' reading ability. These still need to be improved. Nevertheless, I think the UEE paper is not a bad one. (Andrew)

Although the fairness of university entry is still a big concern in China (Davey et al., 2007), Andrew thought so far the UEE was the fairest examination in China for university entry in terms of the test development process and the content. However, there were many aspects that needed to be improved and changed in order to make this UEE paper better.

7.4 Conclusion

The class observations, post-class interviews with the teacher and students, and the interviews with the UEE designers and the materials collected showed that the UEE had a strong impact on different aspects of pedagogy (See Figure 13): teaching, learning, textbook, curriculum and assessment.

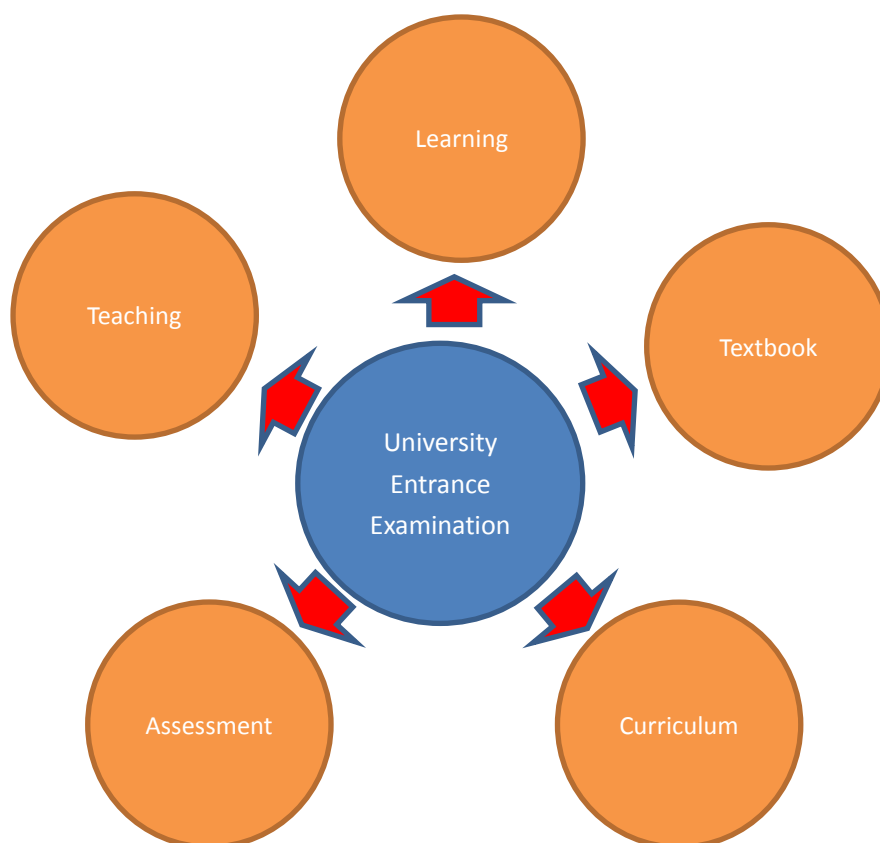


Figure 13: Impact of the UEE on different aspects of senior high school education

7.4.1 Teaching

In investigating whether there might be any impact of the UEE on the teaching content in the classrooms, it was noticed that the teacher did not teach the whole unit from the textbook because some skills, such as listening and writing, were not assessed to any significant degree (certainly not when compared with reading), whilst speaking was not compulsory in the UEE. The researcher asked the teacher why she omitted the listening and writing sections from the textbook, as both sections were included in the UEE. She said the students would be trained for the listening section by practising listening exercises and for the writing section by learning how to write short essays in senior III class, the last year of senior high school, with training

beginning about a term before the UEE. At the moment, for listening, the students only had opportunities to practise listening tasks in their monthly tests. The teacher thought her students lacked a good foundation in writing essays; it was better for them to learn more vocabulary and practise more translation exercises in order to write better sentences because in this way they would equip themselves better to write a short essay. At the time the class was observed, the teacher thought that it was more useful and worthwhile spending more limited class time practising grammar, vocabulary, cloze and reading in class.

In effect, the teacher just used whatever method she felt comfortable with and which was the most expedient means of preparing the students for the UEE. When the teacher was checking the cloze exercise answers with her students, she asked if they had any questions in that exercise. The students shouted out question numbers and the teacher explained the answers one by one. The researcher was surprised that the teacher would not let the students find out the answers by themselves using group discussion or peer feedback. This is understandable because the teacher had only 40 minutes in each lesson and, as she mentioned to the researcher, had a great deal of information to cover.

During the period the class was under observation, the teacher taught one of the units from the textbook. It was a reading comprehension passage. The researcher observed that the class was very teacher-centred; the teacher read whole passages in class, translated the meaning of the passage into Chinese, and worked in the textbook exercises by asking students questions individually in class. There was neither student-student interaction nor group discussion in class. On the other hand the teacher spent much of the class time teaching vocabulary and asking students to translate sentences into Chinese. She told the researcher that vocabulary building was very important in the UEE and that she was testing knowledge of words that might appear in the translation and reading comprehension sections. The teacher seldom followed the teacher's guidebook because the information given by the guidebook was too general for the UEE.

7.4.2 Learning

Generally speaking, having to sit a test changes motivation and thus behaviour (Alderson & Wall, 1993). Different beliefs and behaviour were generated by the UEE for students, or test-takers. If a test was imminent, students would usually review new words and exercises at home every day after class. They also spent time memorising vocabulary just for the UEE and had

monthly tests in school. After the tests, the teacher would spend class time studying the correct answers and examining why the students made mistakes. Also, when working on cloze exercises or reading comprehension exercises, students realised they were also learning techniques and strategies to deal with different kinds of questions in the UEE papers. Students also noticed that the teacher's focus in class would change as the UEE approached; more time would be spent on doing practice exams and more intensive training would be given in senior III class.

Parents as well as students cared about their study and the UEE results. Some students told the researcher that they attended private English tutorials because their parents thought they would thereby acquire more knowledge and strategies for the UEE. Most students thought their English learning beliefs and behaviour would be very different if there were no UEE. They would learn whatever they wanted to learn in English and also develop more communication skills in order to communicate well with foreigners. Parents, on the other hand, would not be persuaded that learning that did not directly assist the students to pass an examination would be of any use. However, the data gathered has not allowed us to judge what extent the beliefs and behaviour of teachers and students would change if the UEE were cancelled.

7.4.3 Textbook

The teacher neglected some text types or activities from the textbook, feeling that these never appeared in the examination papers and were irrelevant to the UEE and therefore not worth spending time on. The teacher thought the textbook alone was not enough for the UEE and she claimed that the school and teachers discussed and decided to use other materials such as cloze exercise books and newspaper supplementary exercises, as the textbook had no cloze exercises. Those materials or exercises mirrored the UEE format. The purpose of such exercises was to familiarise students with the format and style of the UEE. Practising past-paper-like tests designed by publishers might lead to a narrowing of the curriculum to conform with the UEE, but might still help students preparing for it (Smith, 1991).

The teacher informed the researcher that many publishers noticed that school teachers did not follow every task in the textbook; therefore they published supplementary exercises to fulfil teachers' needs. As mentioned, students were required to subscribe to the *English Coaching Paper* in the school bi-weekly because the paper provided much information and extra exercises related to the UEE. In school, the students had monthly tests provided by the

publishers of the *English Coaching Paper*; the test's content mirrored that of the UEE. In sum, textbook writers or publishers adjusted their publishing strategies in order to fulfil users' needs because of the impact of the UEE.

7.4.4 Curriculum

Qi (2004, p.181) has found that "teaching to the test" is a widespread practice that has hindered the implementation of the curriculum. The format of the UEE has become the format for any English test in schools (Liu, 2010). At a macro level, Gu (2012) also states that there is no communication between the Curriculum developers and UEE designers, and he wonders if the UEE designers had read the English Curriculum Standards carefully. A communication channel should be opened between these two parties in order to minimise the narrowing of the curriculum to match the examination in the classroom.

The teacher understood the content of the English Curriculum Standards, however, she thought the English Curriculum Standards were quite abstract and vague and bore little relationship to real teaching in schools. She hoped more details would be provided that could be applied in her teaching. Another document, the Examination Specifications, provides information about the content of the UEE. The teacher thought the examples given in the Specifications were limited; she had to use additional materials and exercise books, which were close to the content of the UEE.

7.4.5 Assessment

The teacher set up the tests format, which was exactly the same as the content of the University Entrance Exam, rather than the content of the textbook. She told the researcher that the contents would be drawn from different test reference books provided by the publishers and past examination papers. Besides the end of term exam, the students were given tests every month; these tests were similar to the format of the UEE. Every time the students finished their monthly tests, the teacher reviewed the answers and gave students detailed feedback because she thought it was very important that students not repeat the same mistakes. The teacher adopted the marking criteria and rubrics used by the UEE when she marked the monthly tests.

Other than monthly tests, students were assessed daily. They were asked to work on the cloze exercises at home in the *English Cloze Test*, which was selected by the English teachers in her

school. The teacher usually spent the first 10 minutes of the class checking the answers. Those exercises took the same form as in the UEE. Since students habitually did not perform well in this part, the teacher emphasised it and made them do relevant exercises. She thought it a good opportunity for her students to practise the cloze exercise in order to get higher scores. On the other hand, when the teacher finished every unit in class, she gave students a dictation task in class in order to ensure they understand the textbook's keywords. When the teacher was asked why they had different assessments in class, she replied that the choice of assessment was driven by the UEE. This demonstrates that the UEE has a very strong impact on assessment in school.

A similar impression came from the UEE designer interviewed. The UEE designer believed the University Entrance Exam had a negative impact on teaching and learning in school. From time to time, he noted, they tried to change the test-task format in order to make the UEE material more practical and authentic. The Examination Centre sent their staff to schools to exchange their ideas on teaching and learning so as to reduce any negative washback in school from the UEE.

In conclusion, the University Entrance Exam appears to have a very strong impact on teaching, learning, textbook, curriculum, and assessment. It seems that those impacts were more unintended consequences, such as narrowing of instruction and curriculum, pursuing the UEE outcomes, publishers issuing extra materials suited to the UEE, focusing language skills and knowledge but ignoring other areas in the English Curriculum Standards, and using examination preparation materials in class that were closely linked to the UEE.

This chapter has summarised, analysed and discussed data from the class observations and post-class interviews with the teacher, students, and a UEE designer. The results indicate that the University Entrance Exam has an enormous impact on students, teachers, and schools, and also on parents, textbook publishers, Curriculum developers, and UEE designers.

Chapter 8: Response validity of the English Language Exam

8.1 Introduction

Henning (1987) describes response validity as “the extent to which examinees responded in the manner expected by the test developers” (p.96). He also explains that test-takers’ responses may not reveal their true ability if they respond in a random manner or are not familiar with the test format and instructions. In this case, the test may be said to lack response validity. Thinking-aloud protocols are used to capture participants’ internal thinking processes during test tasks of the examination paper. They reveal if the test item or task engaged the test-taker’s language ability the way in which it is designed (Nunan, 1992). These verbal reporting methods are designed to help researchers reveal a comprehensive picture of the thinking process (McDonough & McDonough, 1997).

Verbal protocols are spoken records of thoughts, given by participants “when thinking aloud during or immediately after completing a task” (Kasper, 1998, p.358). There are two types of verbalisation: concurrent and consecutive. Concurrent verbalisation takes place when participants are working on the task; consecutive think-aloud reports take place immediately after finishing (Kuusela & Paul, 2000). Thinking aloud is different from introspection in that it does not involve participant’s interpretation (van Someren et al., 1994). Participants are requested to talk aloud while solving a problem; this request is required if the participants stop talking during the problem-solving process, and they are encouraged to keep saying what they are thinking in their minds.

Think-aloud protocols gathered from participants contain knowledgeable, strategic and reasoning information that explains why they chose the answers they did (Cohen, 1984; Gu, 2014b; Norris, 1991). The function of verbal protocols is not only to capture the real-time thinking behind the process, but they also enables more reliable inferences about reasoning rather than just the answer itself. When a participant is working on a test item or task, his or her thinking aloud will reveal if a test item or task engages the test-taker’s language ability. In other words, is there any interaction between the test-taker and the test task? If there is, how is the language ability of the test-taker involved, and what kind of language knowledge and strategic competence does the test-taker use and execute?

Bachman and Palmer (1996) define interactiveness as “the extent and type of involvement of the test-taker’s individual characteristics in accomplishing a test task” (p.25). The individual characteristics are test-takers’ language ability, topical knowledge, and affective schemata. The focus, in this study, has been on language ability, including language knowledge and strategic competence. The interactiveness of a given language test task can thus be symbolised by engaging the test-taker’s areas of language ability in the test task. The interactiveness exists in the two-way interaction between the test-taker and test task. This interaction requires the use of language knowledge and strategic competence, otherwise, inferences about language ability based on test-taker’s performance will not be able to be made (Bachman & Palmer, 1996).

Response validity is a measure of the ways test-takers respond to the test tasks and the psychological processes involved in completing the test tasks. If think-aloud protocols reveal that the completion of the task requires background knowledge and strategic handling of input, and these kinds of actions are indeed expected by the test designers, then response validity is high because test takers have responded as expected. Response validity focuses on one type of evidence during the validation process, i.e., the interactiveness of the test.

Verbal reports of thinking provide strong evidence for theories regarding reasoning processes (Norris, 1991) and analysis of them supports extra important validation (Ericsson & Simon, 1993). Therefore, Norris (1991) believes that participant thinking aloud is pertinent to construct validation. Buck (1991) recommends using verbal report methodology to investigate how listening tests process and claims that this methodology delivered valuable insights into language thinking process and test performance. Cohen’s (1984) report on obtaining think-aloud data shows how test-takers execute language strategies when taking language tests; in particular cloze and reading comprehension. Similarly, Meyers et al. (1990) describe how think-aloud techniques demonstrate that children carry out problem-solving strategies by asking them to describe their thoughts while working on the task. The study of Mician et al. (2000) on response validity of the IELTS Writing Subtest proves that using verbal protocols could identify participants’ writing behaviour and such valuable responses provide insights into response validity for language testing. However, up to the present very little research on response validity or verbal report on language testing has been carried out; this is especially the case with respect to the Shanghai University Entrance Exam.

This chapter reports on the study into the response validity of the UEE. Did the responses reflect what UEE designers planned to test in the UEE? In this study, think-aloud protocols were used among a group of senior high school students in Shanghai in order to capture the ‘online’ thinking behind the processing of these test tasks. For example, in reading a passage or during listening comprehension, a student was trying to choose among four multiple-choice options. Concurrent or consecutive thinking aloud uncovered whether or not a question item or test task indeed engaged the test-taker’s language ability the way in which it was designed. The chapter concludes by assessing whether test-takers did respond to what test designers wanted to test, whether individual characteristics (language knowledge, strategic competence, topical knowledge, and affective schemata) were involved while participants were taking the tests, and whether any other abilities have been tested.

8.2 Research method

8.2.1 Participants

Eight students – Aaron, Ben, Cathy, Dennis, Elaine, Fanny, Grace, and Helen (all pseudonyms), were randomly selected from a senior high school in Central Shanghai City. These eight students were in their senior II class. They were all native Chinese speakers and over 16 years old. All had over 10 years of English learning experience. They were to take the University Entrance Exam 13 months after the think-aloud protocols. It was explained that the data was to be handled in strict confidence. Consent from the participants was obtained on a signed form.

8.2.2 Instruments

The eight students were asked to take one of two past UEE (2008 and 2011) papers. They were divided into four groups – two groups worked on the 2008 UEE, and two on the 2011 UEE. Within each group, one member worked on Grammar and Vocabulary, and Cloze and Reading Comprehension and the other member worked on Listening Comprehension and Writing (included Translation and Guided Writing). The tests were held on four Saturdays (in April and May 2014) in school.

8.2.3 Procedure

Two students from each group took part in this think-aloud session on Saturday morning in school. Data collection proceeded as follows:

Introduction: Before the session started, a classroom was booked and all equipment and tools, such as examination papers, pens, an mp3 recorder, and bottles of water for the participants, were ready. The researcher welcomed and expressed his gratitude to the participants for their time to take part in this activity.

Objectives: The participants were given an explanation about the aim of the study, about what was going to happen and about the protection of the collected data. It was also important to tell them that the work they did would not influence their grades in school. They were reminded that the researcher was interested to see how they solved problems and worked on the test tasks but was not interested in their opinions on the UEE.

Setting: The researcher had to check if all materials and equipment were ready. He also had to make sure that a comfortable environment was set up so that the participants felt at ease. The room should be quiet and the seats should be comfortable. The situation should be focused on the task and the researcher should interfere as little as possible in order to avoid influencing the process.

Instructions: The participants were given a past UEE paper. They were reminded to keep talking out loud while working out the tasks or problems, and voice how and why they chose particular answers. They were instructed to try to say everything that went through or came into their mind when they were reading the passages or questions. They were also reminded to read out every item number when they started each new item; if they thought they had finished an item, then they should go to the next one and read out the new item number. During the process, the researcher was not going to ask participants any question. He was not asking what they thought about the questions, what their opinions of the questions were; he was just recording what came to their minds when they were working on the tasks.

Warm up: A brief training session and practice was provided to the participants before the think-aloud task was started. They were also told clearly what thinking aloud was. For the day that they were working on the tasks, a brief reminder or practice was given before it started. Practising not only gave the participants an opportunity to familiarise themselves with thinking aloud, but also gave the researcher a chance to train the participants to confine themselves to verbalising their thoughts and not interpreting them. The real session started once the participants felt confident and comfortable with the tasks.

Recording: The activity was recorded on audio mp3. The participants were informed that the researcher would start to record when the participant started the tasks. The researcher intervened only when the participants stopped talking; then the researcher prompted the participant by displaying 'keep on talking' sign. During recording, the researcher jotted down some notes about different facets of the exercise, including the participant's facial expressions. When answers were particularly interesting he recorded the exact time for future reference.

Post think-aloud tasks interviews: Once the participants had finished their think-aloud tasks, short interviews were conducted. Only one student was interviewed at a time. The participants were asked what they thought about the UEE, if they thought if the UEE tested their real language ability, and which part of the UEE they thought should be improved. The interviews were audio-recorded, five to ten minutes long, in Chinese and semi-structured.

Transcription: After the think-aloud protocols and post think-aloud tasks interviews had been recorded, participants' verbal responses and interviews were transcribed and translated into English. In addition, notes taken during the protocols were retained as supplementary information.

8.2.4 Data Collection – Recording Details

8.2.4.1 Listening Comprehension

Section A – Short Conversations: The mp3 tape was paused after each participant had listened to each short conversation so that they had enough time to voice aloud what they were thinking when they chose the answer. The pause was about five to ten seconds for the participants to speak aloud because in reality there were about five seconds between two conversations.

Section B – Passages: There were two passages in Section B. Participants listened to the first passage, then the second passage; participants listened twice and then when the question came, there was a five-to-ten-second pause before the next question so that they had enough time to voice aloud; they spoke out what they were thinking. During the listening, participants' behaviour was observed.

Section C – Longer Conversations: There were two longer conversations. Participants were not able to voice their thinking while they were listening and filling out the form at the same time.

Therefore, participants finished listening and filling out the table, then told the researcher how they got the answers. For his part, the researcher asked them to explain their actions.

8.2.4.2 Grammar and Vocabulary

Section A – Grammar: In this section, there were 16 multiple-choice question items. For each item, participants read out the item number first, then read out the question content, and voiced everything that went through or came into their minds, and finished by telling the researcher which option they chose for the answer. Then they moved on to the next item. The researcher observed that participants voiced much grammatical terminology in Chinese when they problem-solved because in school grammar rules were explained in Chinese.

Section B – Vocabulary: In this section, there was a cloze passage with nine blanks; ten words were given in the box and each word could only be used once. Participants usually started to read out the words given in the box, and then read through the whole passage. They were reminded to voice all their thoughts. The researcher did not intervene. Participants skipped the item if they did not know the options; they later returned to the skipped items for another try.

8.2.4.3 Cloze and Reading Comprehension

Section A – Cloze: Cloze included 15 missing items in the passage with four multiple-choice options provided. Similar to the vocabulary section, participants read out the passage, they stopped at the blank and then read the multiple-choice options. They voiced whatever was in their minds when they answered the items. They tried to explain how to obtain the answer from gaining an understanding of the whole passage or its context.

Section B – Passages: Participants usually started to read the first question and multiple-choice options provided; then they identified which paragraph related to the question and read the paragraph very carefully. Similarly, they read out loud and told the researcher how they had worked out the answers from the passage.

Section C – Heading Matching: In this section, the passage contained five paragraphs with the missing headings for each paragraph. There were six headings provided for test-takers to choose to match with the paragraphs. Participants usually started to read out the options provided in the box and then read out the paragraphs one by one. They tried to figure out

which option matched with the particular paragraph. In the process, they voiced what was in their mind and explained how they chose the answer from the box.

Section D – Short-answer Questions: Similar to Section B Passages, there were four short-answer questions without multiple-choice options. Test-takers were required to provide the answers with no more than eight words. Participants voiced how they obtained their answers from the passage and read out their final answers at the end of each item.

8.2.4.4 Writing

Section A – Translation: Participants were firstly asked to read out the item number and the Chinese sentence to be translated into English. Then they spoke out the English word provided (in bracket after the Chinese sentence). Participants then started to work on the translation and spoke aloud whatever they were doing and what came through their minds. At the end of each item, they read out the complete English sentence they had just translated.

Section B – Guided Writing: Unlike the Translation Section, the topic was given to the participants first before recording. They read the topic and the instructions, then they started to write a 120-150 word essay. After the participants finished their writing, they explained to the researcher how they had comprehended the topics and the procedures required as they wrote their essays. At the end, they were asked to read their essay to the researcher.

8.2.5 Data analysis

The test items were checked to see if they were indeed engaging the ability that was meant to be elicited through the items. When analysing the data, the Examination Specifications were compared with the participants' responses in the think-aloud protocols. The findings showed how students responded to the question items and any particular strategies they used when solving the problems. Table 58 lists what standards the UEE designers would like to test in each single section.

Table 58: Objectives of English Language for Examination Specifications

Section	Description of Objectives
<i>Listening Comprehension</i>	1. Acquiring factual information from conversations; 2. Making simple judgement(s) from factual information; 3. Understanding hidden or implied meaning of discourse; 4. Summarising the main idea of a discourse.

Grammar and Vocabulary	<ol style="list-style-type: none"> 1. Identifying, comprehending and using language knowledge correctly in a specific context; 2. Comprehending and using vocabulary correctly in specific context.
Cloze	<ol style="list-style-type: none"> 1. Speculating on meaning by understanding the context and word repetition phenomena; 2. Finding answers by speculating on the passage; 3. Finding answers by understanding details within the sentence; 4. Analysing the passage by using language knowledge; 5. Obtaining answers by referring to similar words or meanings from the passage.
Reading Comprehension	<ol style="list-style-type: none"> 1. Comprehending details in the passage; 2. Inferring the implied meaning from the context; 3. Summarising the main idea of the passage; 4. Speculating on meaning by understanding words and/or sentences correctly in context.
Writing	<ol style="list-style-type: none"> 1. Translating fluent sentences by using language knowledge; 2. Writing correctly, coherently, and aptly according to the meaning of question item or topics.

8.3 Findings and discussion

8.3.1 Response validity: Do test tasks match the Examination Specifications?

The two UEE papers contain the following sections: listening comprehension, grammar and vocabulary, cloze, reading comprehension, translation and guided writing sections. In most cases, the participants responded as the designers expected when they performed in the think-aloud protocols. Examples in each section are listed as below in detail.

8.3.1.1 Listening Comprehension

1. Acquiring factual information directly from conversations:

2008:1 M: Will Mrs. Johnson be here tomorrow for the meeting?

W: Probably not. He called this afternoon to put off the meeting to next week.

Q: When will the meeting be held?

He asked if there is a meeting tomorrow, the meeting is delayed to next week.

Therefore I would choose C. (Ben)

First choose C, the man mentioned tomorrow, the meeting would be held tomorrow, he called this afternoon and cancelled the meeting, and it was delayed to the next week. Therefore, I choose C. (Helen)

2011:5 M: Hi, Jane. It's been ages. You haven't changed a bit!

W: Except for the hair!

Q: What does the woman mean?

The man said that she had not changed much. The woman said except for the hair, it means her hair has changed. (Dennis)

The woman said except for the hair, the man said that she had changed a bit; so the woman felt she was not satisfied with her hair, Choose B. (Fanny)

2. Making simple judgement(s) from factual information:

2008:9 M: Would you tell me what time the dormitory doors will be locked.

W: No problem. In fact I can give you a copy of the dormitory rules.

Q: What does the man want to know?

It is about time. The answer should be A dormitory hours. (Ben)

The dormitory rules – choose B. The doors were locked. And then the woman gave him the rules of the dormitory. (Helen)

2011:3 M: Here is your room key, and the checkout time is 12 noon.

W: Thanks for reminding me.

Q: What is the most probable relationship between the two speakers?

He said that it was about the room card and what you had to do before 12 noon, they were about the hotel guest and hotel receptionist; therefore choose A. (Dennis)

He mentioned about the key of the room and the registration, therefore choose A hotel receptionist and guest. (Fanny)

3. Understanding hidden or implied meaning(s) of discourse:

2008:2 W: It seems Nancy never wants to do anything but play tennis.

M: That's what she likes to do most.

Q: What can be inferred about Nancy?

I didn't hear the first part very clearly. From the second part, the man said that she liked to play tennis. The man said that tennis was her favourite sport. The answer should be C. (Ben)

The man said that Nancy always liked playing tennis, so would not be A, also not B. D she is a professional athlete that is not mentioned in the conversation, therefore C. (Helen)

2011:8 M: Everybody is here except John. Shall we start the meeting?

W: If we wait for John, we might be here all night.

Q: What does the woman mean?

The man asked if we should wait for John. But the woman said, if we waited for John, we could not start our meeting at night. It can be inferred that the answer is C. (Dennis)

The woman said, if we waited for John, we would start our meeting at night. Therefore, they felt that they should not wait for John and should start the meeting then; therefore, the answer is C. (Fanny)

4. Summarising the main idea of a discourse:

2008:7 M: It's impossible to find a place to park before the class at 10:00.

W: Yeah, we seem to make more parking areas here.

Q: What can we learn from the conversation?

They were looking for a parking area. But I didn't hear clearly. I guess personally. But I am sure A, B, C are incorrect. Therefore I choose D. (Ben)

Choose B. The man said that he could not find the car park before the class. It can be concluded that parking areas are full before 10am. (Helen)

2011:16 Q: What is the speech mainly about from the passage?

The main idea of the passage, firstly, I can eliminate the answer A and C, the answer is D, oh no, no, the answer should be B. (Dennis)

I choose B Roles of knowledge in students' growth. (Fanny)

8.3.1.2 Grammar and Vocabulary

Section A: Grammar

1. Modal verbs:

2008:29 According to the air traffic rules, you 29 switch off your mobile phone before boarding.

It is about the traffic rule, so before boarding, about the obligation; therefore choose should. (Aaron)

You need to turn off your phone before boarding. The answer is D should. (Helen)

2011:28 I 28 worry about my weekend – I always have my plans ready before it comes.

I always have my plans; that is I don't need to worry about my weekends; therefore the answer is D needn't. (Cathy)

I have already got my good plan for the coming weekend; therefore I don't have to worry, the answer is D needn't. (Elaine)

2. Tense and voice:

2008:32 In recent years many football clubs 32 as businesses to make a profit.

This football clubs are under operated, so passive voice. In recent years, therefore choose present perfect tense; hence the answer is have been run. (Aaron)

In recent years, it should be present perfect tense. Is it passive voice? It should not be? The football clubs are run? Therefore choose B have been run – present perfect tense and passive voice. (Grace)

2011:31 After getting lost in a storm, a member of the navy team 31 four days later.

Since it is a passive voice, I eliminate A and C. It is because he had been already rescued; therefore I choose D had been rescued. (Cathy, who recognised that this item tested tense and voice; but picked up the wrong tense)

After getting lost in a storm, because four days later, it should be past tense; therefore choose A rescued. (Grace, who chose the correct tense, but did not realise that it was a passive voice)

3. Adverbial clause of time:

2008:31 X: Are you ready for Spain?

Y: Yes, I want the girls to experience that 31 they are young.

I want the girls to experience that when they are still young, so choose while.

(Aaron)

Yes, I want the girls to experience when? They are young. I think the answer is A

while; while they are still young. (Grace)

2011:36 If a lot of people say a film is not good, I won't bother to see it, or I'll wait 36 it comes out on DVD.

I will wait. I wall wait. It is about time; I choose D until. (Cathy)

I will wait xxx it comes out on DVD. I will wait 直到 [Chinese words mean until];

I choose D until. (Grace)

4. Noun clause:

2008:34 As his best friend, I can make accurate guesses about 34 he will do or think.

This is a noun clause; he thinks and does, and what he thinks and does; there the answer is A what. (Aaron)

I guess either 'which he will do or think' or 'what he will do or think'; I feel that the answer is A what. (Helen)

2011:35 There is clear evidence 35 the most difficult feeling of all to interpret is bodily pain.

This is an appositive clause that the second clause is to explain the first clause; therefore, D that should be used. (Cathy)

It is because the second clause is to modify the evidence and evidence is a noun; therefore I should choose D that. (Grace)

5. Attributive clause:

2008:38 We went through a period 38 communications were very difficult in the rural areas.

The word form before the blank is a noun and after the blank the word is also a noun; therefore, it should be a preposition. The answer is C in which. (Aaron, very sure about the answer)

Communications were very difficult ... I think it is a guess, I guess the answer is C in which! (Helen)

2011:39 You'll find taxis waiting at the bus station 39 you can hire to reach your host family.

The second part of the sentence is showing the purpose, and it is talking about the taxis, which help to reach your family; therefore I choose which A. (Cathy)

I think the second part is describing the bus station; therefore, I choose where

B. (Grace was confused that the question talked about the location so she obtained a wrong answer)

Section B: Vocabulary

1. Adjective:

2008:41 He became a 41 voice in the year before the Civil War.

What voice? I would choose a leading voice, just guess! (Aaron)

This became what voice? It should be an adjective. Adjective! (Grace)

2011:41 Owing to the work that has gone into developing intelligent materials, this may not be as 41 as it sounds.

As ... as, we should use an adjective; in this sentence, it means that if you wear the same underwear for several weeks and the underwear is made of very special and good materials, it would not upset people with the smell. Therefore I choose B unpleasant. (Cathy)

As ... as, we should use an adjective, as unpleasant as; not satisfactory, as unpleasant as it sounds; therefore 41 I choose B unpleasant. (Elaine)

2. Noun:

2008:45 Douglass grew up to become the first black 45 to hold a government office.

Douglass grew up to become the first black 45; it is a noun, the first black, should be the first black citizen, citizen. (Aaron)

To become the first, become the first black, black citizen? Citizen? Black citizen to hold a government office? 45 should be a noun. (Grace)

2011:46 A special 46 on these windows not only prevents dirt from sticking to their surfaces, but also allows dust to be easily washed off by the rain.

A special xxx, it should be a noun; ... A special what on these windows not only prevents dirt from sticking to their surfaces. It should be this window has a special coating; F coating should be the answer. (Cathy)

... It should be a noun after special; special, I should choose advances J. (Elaine obtained the wrong answer although she knew it was a noun for the blank)

3. Verb:

2008:47 Douglass was involved with the Underground Railroad, the system 47 up by antislavery groups to bring runaway slaves to the North and Canada.

In the 1850s, Douglass was involved with the Underground Railroad, the system 47 here it should be set up, so 47 should be B set. (Aaron)

The system set up by, set up by, 47 choose B set. (Grace)

2011:47 In fact, these new windows have already been 47 to some cars."

Passive voice should be used here for 47; only option A launched is left, have been launched to some cars, have been launched. (Cathy noticed that the answer should be a verb and passive voice)

New windows have already been 47; have already been fixed to some cars.

Therefore it is G fixed. 47 choose fixed. (Elaine)

8.3.1.3 Cloze

UEE 2008

1. Speculating on meaning by understanding the context and word repetition phenomena:

2008:57 Many parents go to children's sporting events and shout insults at other players or cheer when their child behaves 57.

Aggressively – the passage mentions this word several times. Although I don't understand the meaning of 'aggressively', I guess it has negative meaning; therefore I chose C aggressively. (Aaron)

This word appears several times in the passage – a hint for 57; I should choose C aggressively. (Grace)

2. Finding answers by speculating on the passage:

2008:50 People think children should play sports. Sports are fun, and playing with others.

However, playing sports can have 50 effects on children.

There is a word 'however', and before 'however' it talks about sports are fun, it must be opposite of fun; therefore 50 is a negative word, negative effects. (Aaron)

However, playing sports can have, not good effects; it should be 'not good', B negative. (Grace)

3. Finding answers by understanding details within the sentence:

2008:53 Many researchers believe adults, especially parents and coaches, are the main 53 of too much aggression in children's sports.

53 is about the main reason; therefore I choose A resource. (Aaron mistook resource as cause; he chose a wrong answer although he understood the meaning in the context)

53 is about the main reason which induces the bad phenomena; therefore I choose B cause. (Grace)

4. Obtaining answers by referring to similar words or meanings from the passage:

2008:51 Of these, 18,000,000 say they have been 51 at or called names while playing sports.

They have been what at or called names while playing sports; therefore they yell. I choose D shouted. (Aaron)

Or called names? Or they are shouting their names? 51 should be shouted D. (Grace)

5. Analysing the passage by using language knowledge:

2008:52 This leaves many children with a bad 52 of sports. They think sports are just too aggressive.

They think sports are just too aggressive. Here the blank 52 should be a negative phrase – sports are a bad impression; therefore the answer is A impression. (Aaron)

It mentions about negative phenomena; therefore I choose a bad impression for 52 A impression. (Grace)

UEE 2011

1. Speculating on meaning by understanding the context and word repetition phenomena:

2011:54 In constantly changing 54, this is not surprising is the fact that few companies have any idea how many customers they have lost.

In constantly changing 54, changing prices? Expenses? Tastes? Markets? *In constantly changing what?* What is surprising is the fact that few companies have any idea how many customers they have lost. In constantly changing, *in constantly changing market; choose A markets*. (Cathy)

Markets, *markets*, prices, *prices*, in constantly, *it talks about* customers, *here it should be talking about markets; therefore I choose A markets*. (Elaine)

2. Finding answers by speculating on the passage:

2011:55 Only now are organisations beginning to wake up to those lost opportunities and calculate the 55 implications.

Only now are organisations beginning to wake up to those lost opportunities and calculate what implications – financial? *Economical* economical? C financial *financial*. (Cathy)

Calculation! It talks about calculation, therefore here culture A culture should not be correct, it should be financial or economical; implications, implications, implications, financial – it should be C financial. (Elaine)

3. Finding answers by understanding details within the sentence:

2011:58 A customer who receives a poor quality product or service on their first visit and 58 never returns, is losing the company thousands of dollars in potential profits.

A customer receives a poor quality product or service on their first visit, they will not come back again, it shows cause and effect situation here; the answer is A as a result. (Cathy)

As a result, *as a consequence*; A customer receives a poor quality product or service on their first visit, and as a result, *as a consequence*, answer is A as a result. (Elaine)

4. Obtaining answers by referring to similar words or meanings from the passage:

2011:62 62 customers tend to buy more, are predictable and usually cost less to service than new customers.

What customers tend to buy more, are predictable and usually cost less to service than new customers? *It talks about the customers who have been established; therefore, C established.* (Cathy)

Established? Respected? Customers tend to buy more, are predictable and usually cost less to service than new customers. Established *estimates*? Unexpected *disappointed*? They tend to be less price. *Estimated customers, it should be C established.* (Elaine mistook established as estimated; nevertheless, she obtained the answer correctly)

5. Analysing the passage by using language knowledge:

2011:56 Cutting down the number of customers a company loses can make a big 56 in its performance.

Cutting down the number of customers a company loses can make a big 56 in its performance. Make a big difference. (Cathy)

Cutting down the number of customers a company loses can make a big make a big difference. *Make a big difference; therefore I choose D difference.* (Elaine)

8.3.1.4 Reading Comprehension

UEE 2008

1. Comprehending details in the passage:

2008:65 Why are the children not allowed to play football in the playground?

My children go to a primary school where they are not allowed to play football in the playground for fear that a child might be hurt. *From the first sentence of the first paragraph, we can find the answer for 65, the answer is A the school is afraid that children might be injured.* (Aaron)

They are not allowed to play football in the playground because they are afraid children might be hurt. School is afraid that children might be hurt. Yes, I choose A for 65. (Grace spotted the answer immediately on the passage)

2. Inferring the implied meaning from the context:

2008:68 When working as a PR consultant in London, Zoe thought she lived a _____ life.

From the passage, she was a successful PR consultant and life was going well. Therefore, it means that when she was in London, her life was very good, very wonderful. She was a successful PR. Her life was fantastic. She had a good job, a good flat, her life was very satisfying, the answer is A satisfying. (Aaron)

Zoe was a successful PR consultant and life was going well – life is well! So D Boring – *not correct!* B Tough; C Meaningless; A Satisfying; *therefore, the answer is A satisfying. (Grace)*

Participant Aaron was not so sure if ‘satisfying’ was the answer. He reconfirmed by using eliminating strategy and discovered that the other three options (tough, meaningless, and boring) were negative and only ‘satisfying’ is positive.

3. Summarising the main idea of the passage:

2008:71 Which of the following is closest to the main idea of the passage?

Answer A is talking about the friendship, but here the passage is talking about her life, not friendship. So A is not the answer. Answer B means if you have a dream or wish, you will have a target to focus. This sentence is to encourage people to pursuit their dreams. I don't really understand what answer C is talking about. Answer D One stone kills two birds. I am sure this is not the answer. This passage never mentions about the hope or dream, therefore it is not the answer. Although I don't know the answer C, I still choose C as the answer. (Aaron)

Test-takers were asked to use an idiom to conclude the main idea of the passage in item 71. Aaron only knew three of the options and he was sure that none of those three were the answer. He used the eliminating strategy again to choose the option he did not understand. He chose the correct answer.

UEE 2011

1. Comprehending details in the passage:

2011:66 What does Archy and Mehitabel in Paragraph 3 probably refer to?

A book of great fun; a writer of high fame; a serious masterpiece; a heart-breaking play. *Because she mentioned that when I was a child I had a strong liking for humour; therefore, I will never forget. It must relate to humour; therefore I choose A a book of great fun.* (Cathy read the third paragraph and then read the answer options)

From paragraph 3, it mentions from the beginning, my aunt also took down books from her extensive library; therefore it should be A a book of great fun. (Elaine read from the middle of the third paragraph)

Both participants obtained the answer correctly but from different explanation. It seemed that Elaine matched the content of the third paragraph with the meanings of the options rather than comprehending the paragraph, whereas Cathy related 'humour' and 'joy' from the passage with the answer of 'great fun'. Interestingly, different participants used different strategies when they worked on the test tasks.

2. Inferring the implied meaning from the context:

2011:67 Aunt Myrtle recommended the author to a newspaper editor mainly to ____.

Mainly to, *mainly to* (A) *her writing ability, upgrade her writing ability*; (B) *give her a chance to collect material. It is mainly to, should be*, B *give her a chance to collect material.* (Cathy read the fourth paragraph, then read the fourth paragraph again when she was not sure)

My column, begun when I was fourteen, was supposed to cover teenage social activities ... after graduation from ... was my passport to a series writing jobs. *It is mainly to, mainly to*, (C) *involve her*, (D) *offer her; mainly to enrich her social life. It should be* C *involve her in teenage social activities.* (Elaine looked at paragraph 4)

However, both Cathy and Elaine were not able to infer the implied meaning from the context. As a result, they did not obtain the correct answer; which is C.

3. Summarising the main idea of the passage:

2011:68 We can conclude from the passage that Aunt Myrtle was a teacher who ____.

Like a diamond, she has reflected a bright, multifaceted image of possibilities to every pupil who has crossed her path. *She emphasised, possible, what was*

practical or possible for pupils. *Therefore, I choose C.* (Cathy looked at the last paragraph)

Through the whole passage, we can conclude that: A. trained pupils to be diligent and well-disciplined; B. gave pupils confidence in exploiting their potential; C. emphasized what was practical or possible for pupils; D. helped pupils overcome difficulties in learning. *The answer should be B gave pupils confidence in exploiting their potential.* (Elaine)

Elaine was very sure that the answer B was the concluding statement, whereas Cathy thought the last paragraph summarised the passage; accordingly she read the last paragraph directly after reading the question. The reason Cathy chose option C was that it contained the words ‘possibilities for pupils’ which matched directly from the words in the passage. Test-takers sometimes thought they did not have to look in the text for surface matches but just paired the words with those in the options without understanding the test item.

8.3.1.5 Translation

In this part, the test-takers’ ability to translate Chinese into English sentences by using language knowledge was tested. Translation involved an extra process on top of understanding the given text (in this case, Chinese): looking for the targeted vocabulary that can be used in translation. Test-takers were then required to reconstruct the English sentences precisely. In the translation section, three areas were assessed: use of vocabulary, sentence structures, and grammar knowledge. Examples are shown as below:

2008:2 他设法把游客及时送到了机场。 (manage)

He *managed* to send the tourists to the airport in time.

- Use of vocabulary: managed to, send the tourists, in time;
- Sentence structure: He managed to do something;
- Grammar knowledge: using simple past tense – ‘managed’.

I remember the phrase ‘manage to do’; therefore, I first write it down, ‘manage to send’. 设法把游客及时送到了机场。 He managed to, then ‘送’ should be ‘take’.

I write it down, ‘send the airport on time’. *Because it happened in the past, therefore, add ‘d’ for ‘manage’.* He managed to send the airport on time. (Ben)

Then, manage, manage to. He managed to send. ‘游客’ mean the tourists, to the airport. 及时是; on time 是准时, and in time 是及时. So item 2 – He managed to send the tourists to the airport in time. (Helen)

Helen could distinguish between on time and in time, but Ben could not. Ben noticed that it happened in the past; he added ‘d’ after manage, where Helen also could but she did not speak out about using the past tense.

2008:3 你今晚能来参加我的生日聚会? (possible)

Is it *possible* for you to attend my birthday party this evening?

- Use of vocabulary: to attend, my birthday party, this evening;
- Sentence structure: is it possible for somebody to do something;
- Grammar knowledge: it is a question – ‘is it ...?’

Then possible, there is a sentence structure, is it possible ..., is it possible for you to come to my birthday party tonight? (Ben)

Then possible, 你今晚能来参加我的生日聚会? Is it possible for you to attend my birthday party? So item 3 – Is it possible for you to attend my birthday party tonight? (Helen)

Both Ben and Helen knew that it was a question by using ‘is it possible to’. On the other hand, Ben used ‘come’, whereas Helen used ‘attend’.

2008:4 应该鼓励年轻人按照自己的特长选择职业。(encourage)

Young people should be *encouraged* to choose their careers according to their own strengths / strong points.

- Use of vocabulary: be encouraged, careers, according to, their own strong points;
- Sentence structure: should be encouraged to do something according to ... ;
- Grammar knowledge: using passive voice – ‘should be encouraged’.

There is no subject; it should be passive voice. Therefore ‘年轻人’ is passive.

Then I write the word ‘The young people’ for ‘年轻人’; ‘应该被鼓励’ should be encouraged to; ‘按照自己的特长选择职业’ encouraged to choose their jobs; ‘根据自己的特长’ according to their; ‘特长’ is ‘characteristics’. The young people

should be encouraged to choose their jobs according to their characteristics.

(Ben)

Then encourage; there is no subject, so we should use passive voice. Therefore, the youth should be encouraged to choose the job, ‘按照自己的特长’ according to their; ‘特长’ is, ‘特长’ is, what is ‘特长’? Advantage? Therefore item 4 is –
The youth should be encouraged to choose the job according to their advantages. (Helen)

Both of them noticed that there was ‘no subject’ in this sentence and therefore, ‘passive voice’ should be used. Both Ben and Helen understood and used the phrase ‘according to’. However, they had difficulties in using some of the unfamiliar vocabulary; such as ‘career’ and ‘strengths’; instead, they used ‘job’ and ‘advantages’.

2011:2 我常把王海误认为他的双胞胎弟弟，因为他们长得太像了。(mistake)

I often *mistake* Wang Hai for his twin brother because they look so similar / very much alike.

- Use of vocabulary: twin brother, look so similar / very much alike;
- Sentence structure: mistake somebody for somebody;
- Grammar knowledge: The given word ‘mistake’ is a verb, not a noun.

‘误认为’ mistake, mistake A from B; I always/often, *simple present tense should be used*. ‘双胞胎弟弟’ is twin, twin brother. ‘长得太像了’ because they look like; ‘因为他们长得太像了’ because they look like too familiar. I often mistake Wang Hai from his twin brother because they look like too familiar. (Dennis)

I always mistake Wang Hai as his twin brother, because they look so familiar.
(Fanny)

Both participants did not know the phrase ‘mistake A for B’; instead, they used ‘mistake A from B’ or ‘mistake A as B’. For the phrase – ‘look much alike’; instead, Dennis and Fanny used ‘look like too familiar’ and ‘look so familiar’ respectively.

2011:3 对父母而言，没有什么能与孩子的身心健康相比。(compare)

As for parents, nothing can be *compared* with their children’s physical and mental health.

- Use of vocabulary: compared with, physical and mental health;
- Sentence structure: As for parents, + a main clause;
- Grammar knowledge: using passive voice – ‘can be compared’.

As, ‘对父母而言’ As for parents; ‘没有什么能与孩子的身心健康相比’. ‘与什么相比是’ compare A with B, ‘身心健康是身心健康’ physical and mental health. As for parent, nothing, ‘没有’ is nothing, ‘与什么相比’ nothing can compare, ‘与孩子的身心健康相比’, ‘孩子’ children’s, ‘与什么相比是’ compare A with B, compare children’s, ‘身心健康’ physical and mental health. As for parents, nothing can compare children’s physical and mental health. (Dennis)

To the parents, nothing else can compare with their children’s health. (Fanny)

Dennis was not sure whether to use ‘compare’ or ‘compare with’. Both of them did not know that ‘passive voice’ should be used in this sentence; that is, can be compared.

In summary, participants responded as the UEE designers would have anticipated, judging from these think-aloud protocols. Usually, two participants working on the same tasks had the same reaction. However, participants found that the longer the Chinese sentences were, the more complex the English sentences would be, and the more difficult the vocabulary it would require. Some participants also struggled to find appropriate vocabulary and accurate grammar when they translated the longer sentences.

8.3.1.6 Guided Writing

Testing of writing skills in a foreign language might be assessed to determine how much the test measures both writing skill and language knowledge (Green, 1998). Using writing tasks in a test can show the language skill and proficiency of a test-taker. The guided writing section required test-takers to follow the given instructions and guidelines and write a 120-150-word essay. The writing task in the 2008 UEE required test-takers to give opinions on participating in a group dancing competition whereas that in the 2011 UEE required test-takers to write an application letter to apply for funding according to the information given in the memo. The task tested students’ ability to express themselves in writing correctly, coherently, and suitably according to the task’s meaning. The criteria included content and coherence, use of vocabulary and grammar, and structure and organisation in the essay. It tested test-takers’ ability to provide points to support their ideas and to organise their ideas coherently.

From the recording and transcription, the researcher examined how participants handled the UEE topics: how they followed the given instructions and responded to the information in the topic. They were asked to tell the researcher how they interpreted the topics and how they

wrote their essays. The data examined in this part was collected in order to identify how participants planned and wrote their essays and demonstrated their writing ability. This data was collected and analysed to obtain online information on the participants' actions and responses when they performed the tasks. It was discovered that participants had no difficulties in understanding or interpreting the instructions and meanings of the test tasks because the instructions were written in Chinese. The following extracts illustrate their responses to task interpretation and planning.

1. Ben seemed to be very clear about the instructions given in the 2008 UEE. Ben knew very clearly what he was going to write about because the structure and organisation of the essay were guided by the instructions. He summarised his essay at the end. It shows that his essay has a very good structure and organisation:

This essay includes two requirements. First of all I introduced the background of the essay. Then I explained the reasons why we decided to participate in the group dancing competition in threefold. Finally I summarised the above three reasons and stated our stance to join in this competition. (Ben)

2. Similarly, Helen was very clear about the instructions for writing the essay in the 2008 UEE. Her essay structure followed the instructions exactly. She had her ideas or concepts on why people should join the dancing competition. She also mentioned the advantages of participating in the competition:

My essay should mention: 1: reasons to participate in this group dancing competition, and 2: the purposes and advantages of participating in the competition to the contestants. Therefore, in my essay, first of all I made a very simple introduction, then I expressed my opinions whether we would join the dancing competition and explained why we would like to join. (Helen)

3. Dennis thought that the instructions restricted her ideas for writing in the 2011 UEE. She felt that guidelines in the memo were very confusing, especially, item 2 (your basic idea of the proposed project) and 3 (your plan on how to use the project funds). She thought both items were talking about similar things. Therefore, she combined both items into one. Besides, Dennis was clear that it was a letter to apply for funding; therefore, she wrote it as a letter:

First of all, the applicant – Li Ming, would like to apply for funding for a poverty alleviation project to help children in poor regions. The guidelines were given in

the instructions: 1: your basic situation; 2: your basic idea of the proposed project, and 3: your plan on how to use the funding. It seemed to me that the second and third points looked similar and a bit confusing. I had combined these two points when I wrote my essay. Anyway, there were three paragraphs in my essay: the introduction of my situation, how I would use this funding, and what I would do to help the children in the poor region. (Dennis)

4. Fanny was very clear about the instructions as shown in the 2011 UEE she needed to follow when she was writing the essay. It seemed that, similar to Dennis, Fanny combined item 2 and item 3 from the guidelines and described how she would use the money. Actually, she did not write anything about item 2 (Your basic idea of the proposed project) in her essay:

When I read the instructions, I found the format was very new to me. In the beginning of the letter, I just briefly introduced who I was and which school I was from. I did not describe myself too much. Then the second and third points were about the basic idea of the proposed project and how I would use the funding. I spent a lot of detail on how I would use the funds reasonably. Regarding the basic idea, I thought it was all about helping poor children. (Fanny)

In summary, all participants' responses seemed to be very clear about what they wrote for the task, as the instructions were given very clearly. From their essays, their content of writing was very similar and the essays were quite coherent because of good use of cohesive devices. However, the vocabulary the participants used was quite simple and the range of vocabulary was quite limited. Also, it was not difficult to pick up a few grammatical errors from their essays. For the structure and organisation, their writing was similar because they were restricted to the guidelines and instructions as shown in the items. The writing tasks required processes that involve planning, brain-storming ideas, accessing relevant information for the essay, organising the ideas, coordinating the essay, editing and evaluation. From the think-aloud protocols, the participants' performance reflected what the UEE designers wanted to measure: writing ability in the use of vocabulary and grammar, content and coherence, and structure and organisation of the essay.

8.3.1.7 Summary

The main aim of the use of the think-aloud methodology has been to examine if the UEE items or test tasks engaged with test-takers and if the test-takers responded to what UEE designers

wanted to test; that is, response validity. In this study, six different sections – listening comprehension, grammar and vocabulary, cloze, reading comprehension, translation and guided writing – were investigated. Results supported the finding that think-aloud methodology could provide valuable understanding of different aspects of language processing and into the ways in which student performance related to the Examination Specifications that the UEE designers would like to test.

Test-takers use the same skills in listening and reading test tasks except that obviously the stimulus materials are conveyed through different channels (Green, 1998). For example, in listening comprehension, test-takers were required to infer implied meaning from conversations to obtain the answer, whereas in reading comprehension they were required to infer implied meaning from the passages. From the think-aloud protocols analysis, when participants were asked, “What does the speaker mean?” in the conversations, they noticed easily that they were asked to infer or work out the meaning of the conversations.

In the grammar and vocabulary section, participants felt they were familiar with the grammatical terminology of the items on which they were working. In the grammar section, they could easily tell what grammatical competences were being tested in the sentences, which were exactly the kind of grammatical competences the UEE designers wanted to test. When they worked on the vocabulary cloze passage, different participants had the same strategies. They classified the word options into different forms – verb, noun, adjective, adverb – before they started to read the passage. They then settled on the word form of the missing word, so that fewer options needed to be considered. In other words, verbal reports of thinking contain information on the knowledge and strategies used (Norris, 1991).

Think-aloud protocols can provide direct information on cognitive processes and thereby uncover test-taker behaviour. If both tests measure the same construct, similar processes could easily be recognised from the think-aloud protocols (Green, 1998). Similarly, in the cloze and reading comprehension section, different cloze or reading passages tested the same construct. It was readily noticed that the participants had very similar responses and similar thinking processes in the think-aloud protocol analysis, which provided very strong support for the argument that they tested similar constructs. For example, in passage A of the 2008 UEE, both Aaron and Grace had very similar strategies for reading comprehension; before reading

the passage, they looked at the first question, and they could spot the answer very quickly from the passage for the question item.

In the translation section, when participants carried out the translation tasks, the planning processes of constructing sentences were revealed through the analysis of think-aloud protocols. Translation involves the additional skill of constructing another language on top of the understanding of the question items. Translation tasks from L1 to L2 are considered to be more difficult than that from L2 to L1, because the former requires the productive skill of vocabulary and sentence structure in the target language (Green, 1998). When the participants worked on the tasks, they relied on cognitive processes and underlying abilities in spheres such as language use and sentence structure, which were easily revealed from the think-aloud protocols. From the analysis, the findings showed that different participants had very similar strategies and methods for completing sentences.

The use of think-aloud protocols can also reveal the validation process for writing tasks. They show how the participants understood and interpreted the test tasks. In the guided writing section, different participants had different responses to the same writing tasks, including use of vocabulary, writing content and essay structure. Although they had different responses, the basic writing skills gained through cognitive processes were still able to be assessed by using think-aloud protocols. Student performance in the think-aloud protocols also allowed the UEE designers to measure what they wanted to measure: writing skills on expressing in writing correctly, coherently, and suitably according to the meaning of the task. The criteria included content and coherence, use of vocabulary and grammar, structure and organisation.

In sum, participants in the think-aloud tasks achieved what the UEE designers planned to measure; such performance provided very strong evidence to support the construct they wanted to measure. As a consequence, response validity is considered to be reasonably high.

8.3.2 Interactiveness

Bachman and Palmer (1996) define interactiveness as “the extent and type of involvement of the test-taker’s individual characteristics in accomplishing a test task” (p.25). The individual characteristics are test-takers’ language knowledge, strategic competence, topical knowledge and affective schemata (See Figure 14). In this section, other than those examples mentioned

in Section 8.3.1, examples show how participants engaged with these individual characteristics in the test items or tasks; that is, what level of interactiveness was involved in the UEE.

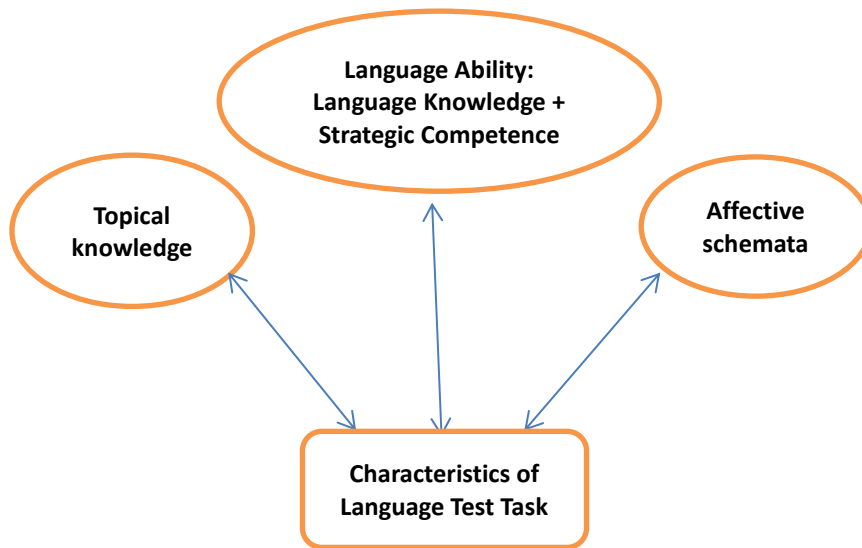


Figure 14: Interactiveness

8.3.2.1 Language knowledge

1. Testing vocabulary:

For 2011:4, test-takers were required to choose the appropriate word from the options to describe a man's feelings. If test-takers did not understand the meaning of the vocabulary, they would not be able to find the right answer.

2011:4 W: Sorry, sir. We are working on your order right now and we'll be delivering it soon.

M: Soon? How soon is soon?

Q: How does the man feel?

Answer: A. Excited. B. Dissatisfied. C. Bored. D. Exhausted.

The woman said that they would deliver the food immediately. But the man was very angry; therefore I chose D Exhausted. (Dennis)

The man complained that the people could not finish their task on time, he was very angry and not happy; therefore I think it is D Exhausted. (Fanny)

Both participants comprehended the conversation and realised that the man was very angry and they thought 'exhausted' meant angry. But they did not understand the word 'dissatisfied'; therefore, they could not get the correct answer.

For 2008:3, test-takers were asked to find out where the conversation takes place when the speakers are talking about the oil painting.

2008:3 M: So how do you like this oil painting?

W: If I had enough room in my apartment, I would buy it.

Q: Where does the conversation probably take place?

Answer: A. At a paint store. B. At an oil market. C. At a science museum. D. At a gallery.

I choose A, the conversation mentioned oil painting. The woman said if she had money she would buy the painting; therefore it was in a paint store. (Helen)

In this situation, participants were tested on vocabulary such as oil painting, paint, museum and gallery. Some words have more than one meaning; it depends on the context. For example, oil for cooking and oil in oil painting; and paintings and paints are different things. Helen mistook paintings as paints; therefore she chose A instead of the correct answer D.

2. Testing spelling:

In longer conversations of the listening section, test-takers were required to fill in the blanks with the information they heard. It was possible that test-takers understood and comprehended the conversations but they might forget the spelling of the words. For 2011:18, the conversation was about sending a sweater at the post office; however, participant Helen did not remember how to spell the word 'sweater'.

18 is 'sweater'? I don't know how to spell. (Helen)

3. Testing grammar:

Longer conversations in the listening section were very straight forward. Participants could easily fill in the form without difficulty. However, item 24 required test-takers to understand the conversation and they then had to modify the word as the blank was limited to a maximum of three words. It tested not only listening comprehension but also grammar; words had to be re-arranged in order to form a sentence in the table.

For 2011:24, participants were required to understand that the latter memory referred to long-term memory and they were also required to complete the sentence in the table – 'Long-term

memory is permanent'. Similarly for 2008:24, participants were required to comprehend the last section of the recording about Ms Ogata so that they were able to put the words – 'She is successful both in career and family.' – in the table.

8.3.2.2 Strategic competence

Think-aloud protocols from the participants contain strategic information which explains why participants choose the answers they do (Norris, 1991). Such an analysis is commonly used to trace which process participants use when they make a decision, and the data is then used to make inferences about the strategies they employed to make their choice (Kuusela & Paul, 2000). From the think-aloud tasks, it was found that participants executed different strategic competence to cope with difficult test items or tasks when they made their decisions.

1. Classifying words into different word forms:

Before participants started to read the passage, they classified the listed options into four categories: noun, verb, adjective, and adverb. Then when they looked at the blank and the sentence, they had to find out which form of the word was needed; then the probability of getting the correct answer rose.

2011:43 The name comes, of course, from the lotus leaves, which are famous for growing in muddy lakes and rivers while remaining almost 43 clean.

While remaining almost 43 xxx clean. Almost? *An adverb is needed to describe 'clean'. The answer is either D technically or H miraculously. I choose H.* (Cathy)

Before Cathy started the vocabulary section, she marked down the form of the words in the box. She wrote down 'adverb' for option D 'technically' and option H 'miraculously'. By classifying the words into different word forms, she had a 50 percent chance to get the answer correct since there were only two adverbs in the box.

2. Word collocation:

It was noticed that participants were familiar with some of the common collocations; for example, a bad impression (2008:52), face up to this problem (2008:60), make a big difference (2011:56), or price sensitive (2011:63). In these situations, participants could find out the answers by understanding the details within the clause or sentence. The better knowledge of collocations they had, the better they were able to get correct answers.

2008:60 As a society, we really need to 60 this problem and do something about it.

Answer: A. look up to; B. face up to; C. make up for; D. come up with.

We should face up to this problem; face up to B this problem and do something about it. (Aaron)

As a society, we really need to 60 this problem and do something about it. It should be: we need to face up to this problem; face up to, Choose B. (Grace)

3. Eliminating strategy:

Eliminating strategy is one of the most common strategies used in the examinations by test-takers. Participants used eliminating strategy to choose answers even though they were not clear about the conversations, and they still could get the correct answers.

Helen was very successful because she used this strategy. For 2008:2, although she could hear only part of the conversation, she was not sure of the answer in the beginning. By using this strategy, she decided the answer after she ticked off the other three options. Similarly, Helen used elimination strategy although she was not clear about the conversation for 2008:4.

2008:2 M: It seems Nancy never wants to do anything but play tennis.

W: That's what she likes to do most.

Q: What can be inferred about Nancy?

Answer: A. She doesn't play tennis well. B. She likes other sports as well.
C. She is an enthusiastic tennis player. D. She is a professional athlete.

He said, he always likes playing tennis, so would not be A, also not B. D she is a professional athlete that is not mentioned in the conversation, therefore C. (Helen)

2008:4 W: Peter, want to take a walk?

M: No, I have to go down to the yard now, I'm building some bookshelves.

Q: What will Peter probably do next?

Answer: A. Work in the yard. B. Buy some wood.
C. Go to the bookstore. D. Take a walk.

He mentioned to go down to the yard, so not D. Also he mentioned the bookshelf, neither B nor C, therefore A. (Helen)

8.3.2.3 Topical knowledge

Topical knowledge is the knowledge someone gets to help them "use the language with reference to the world in which they live, and hence is involved in all language use" (Bachman

& Palmer, 2010, p.41). Topical knowledge is one of the individual characteristics. From the think-aloud protocols, topical knowledge might sometimes appear in the examination papers and the researcher observed if the test-takers were able to cope with it or not.

For example, in the first passage in Section B of the listening section (2011), the topic was about the 'Commonwealth Organisation'. If test-takers read or heard any information about the Commonwealth Organisation, they would have some ideas about the passage. As a result, they could obtain the answers easily even if they did not listen to the passage. From this topic, some students might have this kind of topical knowledge; from 2011:11, the answer could be chosen without listening to the content. From the think-aloud protocols, neither participant possessed the relevant topical knowledge so they did not have any advantages in answering these questions.

Another example is 2011:57; this item involved topical knowledge – finance. "Research in the US found that a 5% decrease in the number of defecting customers led to 57 increase of between 25% and 85%." The answer was 'profit'. This item was a financial calculation: decrease in loss would increase in profit in business. Neither participant was able to obtain the correct answer for this item because they did not have any finance knowledge to support their language knowledge. However, if any test-takers had a finance background they would have an advantage in working on the task.

Led to 57 increases? What increase? Of between 25 and 85 percent. What increase? Cost? Budget, budget? Cost cost? led to cost. (Cathy had no idea about cost and budget)

8.3.2.4 Affective schemata

Affective schemata of test-takers in language testing may affect their thinking when they try to finish the test tasks (Bachman & Palmer, 2010). Their emotional responses to the tasks may either facilitate or limit their ability to make use of their knowledge and meta-cognitive strategies to achieve the tasks; as a result, their performance might be affected. In this study, no affective schemata on the part of any participants were noticed from their performance.

8.3.2.5 Summary

Different test items or tasks involve different extents or degrees of interactiveness. Some tasks engaged more language ability and strategic competence than others; for example, a cloze

passage without multiple-choice options engages test-takers' language ability and strategic competence more than a reading comprehension or listening comprehension task with multiple-choice items. After test-takers finish reading a passage and questions, they will be able to decide the answer from a set of four options very quickly. The test-takers are just eliminating the options. In terms of interactiveness, these kinds of questions are much shallower in the process of thinking and engaging, compared to the question-items in the cloze passage.

For the vocabulary section, test-takers were required to fill in the blanks (with a list of words given in the box) to make the passages coherent and grammatically correct. From the think-aloud tasks, it was found that this kind of integrated task involves not only test-takers' grammar and vocabulary knowledge, but also reading ability and strategic competence. Compared with the discrete items in the grammar section, this kind of test task is more engaging and interactive when test-takers have to make a decision for the answer, because the thinking process in a cloze passage involves eliciting a great number of language competences and skills. The test-takers have to integrate all related knowledge and skills together when solving the problems.

Similarly, translation tasks are more engaging and interactive than reading comprehension multiple-choice question-items. From the think-aloud tasks, participants were required to read and understand the Chinese sentences and use some given English keywords, then translate into these English sentences. By using targeted vocabulary and correct grammar, they had to construct the English sentences precisely. The translation tasks involve not only language use and grammar knowledge, but also awareness of sentence structures. Therefore, the tasks are quite engaging and interactive.

In order to be able to make inferences about language ability, the test items and test tasks in the UEE should involve the test-takers' areas of language knowledge and their strategic competence. Therefore, the more engaging and interactive the test tasks are, the greater the amount of language knowledge and strategic competence the test-takers manage, and the better the inferences about language ability that can be made. Interactiveness is thereby a crucial factor of language tests because it connects a significant link with construct validity (Bachman & Palmer, 1996).

8.3.3 Testing other abilities

8.3.3.1 Testing mathematics calculation

For 2011:2, simple calculation was involved. Participants had to understand and listen to the figures – there was a 10 percent discount and the cost of the meal was 50 dollars, and they were required to do a simple calculation. Participants did not expect calculation to be involved and they were not prepared for that. Participant Dennis thought 10 percent was 10 dollars, she did not expect that it was a mathematics calculation; as a result, she missed the answer, whereas Fanny was doing the mathematics calculation while she was listening.

2011:2 W: I wonder if there is a service charge for our meal.

M: I think so. The menu said the service charge is 10 percent.

Q: How much is the service charge if the food costs 50 dollars?

He said the discount was 10 percent; therefore they had a discount of 10 dollars. (Dennis misheard 10 percent as 10 dollars)

The meal cost them 50 dollars, and the service charge was 10 percent, therefore choose A 5 dollars. (Fanny was calculating the mathematics)

8.3.3.2 Guessing the answer

From the think-aloud protocols, some items were found to be guessed at by the participants. There are several reasons for guessing the answer in the test: 1: participants did not understand the question items because of their language ability or test difficulty; 2: they might find the multiple-choice options were confusing so they were not sure and, as a result, they guessed the answer, and 3: there might be some problems in the test items or those items were poorly designed.

Green (1998) states that negative comments are clearly an indication of a problem but it is important to establish whether this problem is due to task difficulty, lack of motivation or a poorly constructed test item. Think-aloud protocol can be recommended for test piloting; test designers can see whether the test items or tasks engage test-tasks or if there are any poorly designed test items. If they find most of the participants guess the same items, they should also investigate what wrong with the particular item; they might have to modify or remove that item if necessary.

In summary, testing abilities other than language ability might undermine the construct validity of the English language examination; however, very few such situations were encountered in this study. Most of the items tested what the UEE designers wanted to measure.

8.3.4 *Post think-aloud interviews*

Short interviews were conducted in school on the same day after the participants finished their think-aloud tasks. Only one participant was interviewed at a time. The interviews were in Chinese and structured. The questions the researcher asked included: 1: What do you think about the UEE? 2: Does the UEE test your English proficiency? 3: Is the UEE good or bad? 4: Is it a fair examination? 5: In what ways does the UEE need improvement? 6: What is your view on communicative English language? The eight participants – Aaron, Ben, Cathy, Dennis, Elaine, Fanny, Grace, and Helen (all pseudonyms) – had different views on the interview questions.

8.3.4.1 What do you think about the UEE?

Aaron thought the skills he needed for the UEE and the format of the UEE were similar to the tests in his school. The UEE covered much vocabulary that he did not understand – some of the words he felt were difficult to him.

Dennis thought the listening comprehension section was relatively easy except longer conversations in Section C. The translation was also easy, but she had limited vocabulary so she sometimes failed to translate certain words correctly. Speaking about the writing part, her reflection on writing was quite shallow.

Fanny thought her vocabulary was limited. For the translation part, she was totally unable to translate some of the sentences, not to speak of the fact that she would always make mistakes in grammar such as tense and word order. For the writing part, she believed she wrote many grammatically wrong sentences since she had not practised writing for the past two years.

Helen thought the grammar and vocabulary, and writing sections were easier while the listening and reading comprehension sections were more difficult. This was especially the case as there were so many new words in reading that she found it hard to infer what the articles were talking about.

8.3.4.2 Does the UEE test your English proficiency?

Ben felt the UEE was not comprehensive enough. It mainly focused on the grammar and vocabulary section, and ignored oral English. Many Chinese students could write in English but they might have difficulty in communicating with others orally; therefore, the UEE could not fully reflect students' actual language ability.

Fanny thought the UEE was able to show her English language proficiency. For example, the writing part could test her vocabulary, grammar and thinking as well as the ability to organise ideas and frame essays. Any lack of proficiency in either of the two necessary aspects of vocabulary and grammar would be easily and clearly shown in her writing.

Helen did not think so. She thought the UEE just tested students' written English including grammar, but the real case was that she tended to unconsciously and naturally make many grammatical mistakes when she spoke English with foreigners, whereas these might be deliberately avoided in such tests.

8.3.4.3 Is the UEE good or bad?

Ben believed that the good part was the reading comprehension section, which could evaluate his grammar, vocabulary and other abilities in a comprehensive way. The writing section was also good because it could test a student's level of vocabulary and their skill of organisation and structure of the essay.

Dennis thought the translation part was good. The translation was not fixed and it ranged from easy to difficult questions. It was progressive. However, the guided writing section was not so good because the topic of the test task was not very clear.

Helen thought the listening comprehension and writing sections were both good parts while the grammar and vocabulary section, which seemed of little help for communication, was a bad part. She thought the writing section was good because it required test-takers' language organising abilities as well as requiring clear minds to explicitly elaborate an event or express an idea in a second language such as English, the way of thinking of which might also be useful for future communication with foreigners.

8.3.4.4 Is it a fair examination?

Aaron thought the UEE was fair. All students were evaluated in the same ways. Everyone could have access to this examination. It was a fair form of examination but not the best option to evaluate students; the score was determined by their own efforts.

Fanny thought students were able to gain some certain marks as long as they studied hard in senior III; however, they should really be proficient enough in English so as to achieve a truly high score.

Helen thought it was fair because those students who lacked understanding were still able to catch the meanings of the passages by reviewing and utilising vocabulary and grammar, which still accounted for the larger part in the test.

8.3.4.5 In what ways does the UEE need improvement?

Dennis thought if the UEE really wanted to test students' abilities then multiple-choice questions should be eliminated from the examination papers. He said that teachers sometimes trained their students in the skill of choosing the right answers even when they did not understand the questions. Therefore this kind of format should be eliminated.

Grace did not think the UEE needed any improvement. If it did, it should be to make the UEE more comprehensive and in particular to merge grammar with other sections such as listening comprehension, reading comprehension, and writing and speaking sections, rather than testing grammar alone.

Helen would like the UEE to fit real life more, as spoken English was never tested while what was tested was actually not that useful in communication or daily life. She also thought that its vocabulary requirement in the UEE should not be too demanding; she spent so much time in memorising vocabulary.

8.3.4.6 Communicative English language

Ben did not think the materials in the UEE had anything to do with communication with foreigners. Although he had been learning English for 12 years, he was not confident enough to communicate with foreigners in English. He thought students should not be assessed on their performance in the listening section because listening could not help them talk to people fluently, so students should focus more on learning and practising speaking in class.

Cathy thought listening and speaking were both important because listening was basic for understanding what other people said, and speaking was how you expressed yourself to others. English language learning should not be limited to training in written English. She thought the students should learn to communicate in English lessons and that oral English was helpful and necessary for her future studies and career.

Fanny was not sure if the listening material in the UEE was relevant or helpful to the application of English in her daily life. She thought however that the questions in the listening section asking test-takers to infer the relationship between the two speakers from the context might be of some relevance or help. General communication was surely more important than grammar and vocabulary in daily life.

8.4 Conclusion

This chapter began with introducing some background about the *response validity* and the definition of *interactiveness*. The research method of the study was then introduced in detail. This chapter has also summarised the findings regarding one of the test qualities: interactiveness. The findings above show that the test items and tasks fulfilled the requirements of the Examination Specifications. Participants in the think-aloud tasks responded to what the UEE designers planned to measure, indicating high response validity. Test-taker performance showed the ways in which they engaged with individual characteristics (language knowledge, strategic competence, topical knowledge and affective schemata) in the test items and tasks; that is, a high level of interactiveness was involved in the UEE. In comparison to the multiple-choice comprehension tasks, integrated language tasks are more engaging and interactive because the thinking process involves eliciting various language competences and skills. The more engaging and interactive the test tasks are, the greater the amount of language knowledge and strategic competence the test-takers manage, the higher the degree of construct validity is, and the better the inferences about language ability which can be made. It is also interesting that abilities other than language ability were tested in the UEE – this might undermine the construct validity of the English language examination (however, very few such situations were encountered in this study). Finally, participants were asked what they thought about the UEE: if they thought if the UEE tested their real language ability, and which part of the UEE they thought should be improved.

Chapter 9: Summary of findings and discussion

9.1 Introduction

This chapter starts with a summary and evaluation of findings by using Bachman and Palmer's (1996) model of six test qualities. It then discusses the major issues arising from the findings.

9.2 Qualities of the UEE: Summary and evaluation of findings

Bachman and Palmer (1996) proposed a model of test usefulness that includes six test facets – reliability, construct validity, authenticity, interactiveness, impact, and practicality. This model is used here as a framework for summarising and evaluating the test qualities of the UEE.

9.2.1 Reliability

Reliability is always expressed as a measurement of consistency. A reliable test result is always consistent across different testing situations with different characteristics. In this research, 60 students were asked to take two past years' UEE papers (2008 and 2011) and one IELTS test sample, a total of three tests, over three months. In this study, three areas of reliability were considered: 1: internal consistency, which is concerned with sources of error from within the test; 2: stability, which indicates how consistent test scores are over time, and 3: parallel-form, which estimates the reliability of a test to examine two scores obtained from alternate forms of a test.

Spearman-Brown reliability coefficients, Cronbach coefficient alphas, and KR-21 reliability coefficients show that the internal consistency was reasonably high for both the 2008 and 2011 UEEs, with indexes of above +0.870. A high internal consistency meant that different test items or test tasks were measuring the same construct in both the 2008 and 2011 UEEs. Most of the sections measured the same competence in the same section, for example, the test items focused on measuring the grammatical competences in the grammar section. However, a low coefficient was found in the cloze section because different gap-fillings tested different competences and some gaps were more difficult than others. However, this was the exception; generally, the internal consistency for both UEEs was quite high.

The researcher did not examine rater consistency due to limited time and resources. This research only focused on Listening Comprehension, Grammar and Vocabulary, Cloze, and

Reading Comprehension (Paper I), in which most items were multiple-choice options or short-answer items and in which rater consistency was therefore not a factor. In Paper II –Writing, including Translation and Guided Writing (which accounted for one-third of the total score), the question items were subjective and rater consistency, including intra-rater and inter-rater, was therefore crucial.

The UEE papers from both 2008 and 2011 were seen as parallel forms of the same test. They could also be regarded as the same test taken twice over a month. Correlating scores from these two exams could thus be regarded as providing evidence for both stability and parallel-form reliability. In this study, the average value of the correlation coefficients between these two UEEs for different sections was low ($r=+0.500$), meaning low stability. The major reason was that the two sets of papers were designed by two different groups of designers at two different periods. One way to overcome this would be to establish a test bank, which UEE designers could draw on to construct test items and test tasks throughout the year. Up to now, only a small group of UEE designers have been assembled just one month before the UEE, to design the exam paper. Moreover the exam format would change every two to three years. Both stability and parallel-form reliability would be affected.

9.2.2 Construct validity

The term *construct validity* refers to the extent to which a given test score indicates the ability that is measured (Bachman & Palmer, 1996). In this study, content relevance by qualitative content analysis, and correlational analyses by using mock exams provided the evidence for the validation of the UEE.

Stage one in the qualitative content analysis examined the test items or tasks for the past 10 years' UEE papers against the objectives of the Examination Specifications and ECS. All objectives in the Examination Specifications for each of the different sections – listening comprehension, grammar and vocabulary, reading comprehension, writing and speaking – were fulfilled. However, measuring the test items or tasks against the ECS, while the UEE covered most of the objectives in the area of language skills and language knowledge, the UEE covered only a small part of learning strategies, and nothing at all in the areas of affect and attitude, and cultural understanding.

Regarding learning strategies in the Curriculum Standards, only cognitive strategy objectives were accomplished in the UEE. Metacognitive strategy, communicative strategy and resourcing strategy are not easily assessed by any tests or examinations. Those strategies would be more easily evaluated through classroom observation.

Affect and attitude and cultural understanding were not tested in the UEE. Affect and attitude are not easily tested because they are relevant to the daily lives of learners and relate to their interest and motivation in learning English. Intercultural communicative competence should be measured, but this kind of test is still under exploration. National and international awareness should be cultivated during the learning process rather than being tested because both have a strong subjective element. Various cultural ideas found in English-speaking countries should be introduced to students in both curricular and extra-curricular activities. Because these two areas are not easily tested in an exam, teachers need other ways of measuring the realisation of these objectives in the classroom.

Stage two in the qualitative content analysis involved comparing past UEE papers with Bachman and Palmer's model of language ability. The findings can be summarised as follows. Grammatical knowledge, textual knowledge and strategic competence were always the main components tested in the English language examination.

In respect of functional knowledge, ideational functions were tested in all the different sections in the UEE. Manipulative functions were the major components in the listening and writing sections. However, knowledge of the heuristic and imaginative functions was hardly tested.

In respect of sociolinguistic knowledge, the writing section tested knowledge of registers. When asked to write a letter, test-takers were required to understand the characteristics of different levels of formality in language-use. Moreover, knowledge of genres was tested not only in listening and reading, but also in speaking and writing. Other than that, there were very few items testing sociolinguistic knowledge.

Correlational analyses using mock exams demonstrated construct validity. Correlations between the subtests in the 2008 and 2011 Exams ranged from +0.480 to +0.699. This means that each subtest contributed elements to the total test, but no individual subtest could replace another; each subtest was testing different skills. The high correlations, over +0.850,

between the subtests and total UEE scores revealed that the test components had a very strong effect on the final total UEE scores. Overall, the construct validity was satisfactory. The UEE designers knew what they wanted to measure in each section.

On the other hand, the mock exams also showed that concurrent validity of the UEE with IELTS, an internationally recognised test, was low. Correlations for the 2008 and 2011 UEE total scores with the IELTS sample test scores were +0.508 and +0.435. That was a reasonable result but when Listening or Reading Comprehension was correlated separately with the IELTS test sample, the coefficients were lower ($r=+0.350$). At the post mock exam interviews, participants expressed the view that the IELTS test sample was more difficult than the UEE papers. One reason would be that IELTS includes many more language use tasks in the exam papers when compared with the UEE, which included mainly practice tasks focusing on grammar and vocabulary, reading comprehension and translation that test-takers normally took in school.

After the mock exams, the researcher had short interviews with two volunteer participants. Both interviewees thought that the University Entrance Exam was a comprehensive examination because it covered a wide range of aspects in the Examination Specifications. They also believed that the UEE was able to measure test-takers' English language ability; therefore, the face validity of the UEE was satisfactory.

In sum, the construct validity of this UEE is moderately satisfactory. However, in terms of content, the UEE items focused more on language knowledge and language skills when compared with the ECS. The UEE items also focused more on grammatical knowledge and textual knowledge, and strategic competence when compared with Bachman and Palmer's model. This suggests that the construct of the UEE items should be modified by taking in a wider range of language learning areas. This will be further discussed later in this chapter.

9.2.3 Authenticity

Authenticity is the extent to which the characteristics of the test task correspond to those of target language use (TLU) tasks. The question-items and test-tasks were considered to be not quite authentic in the UEE; a few examples from the qualitative content analysis could account for this:

- Gap-filling items of the passages were used in the grammar and vocabulary section;

- Most of the items in the listening and reading comprehension sections were multiple-choice items although the passages used were for the most part related to TLU tasks;
- In the translation section, Chinese sentences were required to be translated into English with key words given, but in real life people do not have to translate from Chinese to English to foreigners;
- In the guided writing section, test-takers were required to write only a 120-150 word essay in a given test-task. A 150-word essay is too short for either a university assignment or an editorial contribution to a newspaper;
- The Oral English Test was a 'one-way' communication; test-takers were required to talk to a computer only and read the information off the screen.

Moreover, from the empirical studies, the researcher had opportunities to interview several students after classroom observations and think-aloud tasks. They all commented that what they had learned in school was more about grammar and vocabulary which might be helpful in reading articles but not for carrying out conversations with foreigners. The test items in the UEE were focused mainly on grammar, vocabulary, translation, and reading comprehension. The students interviewed believed that future studies and careers would require more of their communicative skills; yet the current UEE ignored Oral English. They thought many Chinese students might have difficulty in communicating with foreigners in English. They felt that the UEE was not particularly authentic and could not fully test real world use of English.

Authenticity in a test is considered to be important because authenticity induces a “potential effect on test-takers’ perceptions of the test”; as a result, test-takers will perform at their best in the test (Bachman & Palmer, 1996, p.24). If test-takers think that the test tasks look very relevant and familiar, then it helps test-takers stimulate a positive emotional response to the test task, hence on their performance. Therefore, more authentic test tasks should be recommended in the UEE.

Most language test designers appear to care about authenticity implicitly when designing language tests (Bachman & Palmer, 1996). The UEE designer the researcher interviewed commented that in the past the question items were mainly multiple-choice types and test-takers did not have any opportunities to write and answer questions. He hoped the UEE would become more authentic; that was almost always the reason why the designers wanted to

modify the format. However, it has been shown in the study that actual changes were minor and slow to be implemented.

The researcher thinks that the authenticity of UEE is quite low. As mentioned before, the question items and test-tasks were considered to be not quite authentic in the UEE; for examples, the materials in the reading and listening sections, the test tasks are not authentic for the university study or future career, nor will test-takers encounter these tasks after taking the UEE. Therefore, the questions items and test tasks need to be improved to achieve a higher degree of authenticity. However, when test designers try to design an authentic test task, they have to consider the critical features of the task, which determine the degree of authenticity. The UEE designers should reconsider the main purpose of setting up the English Language Exam – is it to evaluate how much English the senior high school students have learned for the preceding 12 years, or to select best students for university admission? In other words, is the UEE an achievement test or a proficiency test? The debate about whether UEE test-takers belong to a language instructional domain or a real-life domain will be discussed later in this chapter.

9.2.4 Interactiveness

Interactiveness exists in the interaction between the individual and the test-tasks. Findings from the qualitative content analysis of this study, when compared with Bachman and Palmer's model of language ability, showed that most areas of language knowledge and strategic competence were tested in the UEE. But there was a limitation; most items focused on grammatical and textual knowledge but neglected functional and sociolinguistic knowledge. The question-items and test tasks in the past 10 years' UEE papers also involved testing some topical knowledge, but affective schemata were not engaged in the UEE. Some test tasks required test-takers to manage their topical knowledge and relate the test content to their knowledge; those tasks accordingly became relatively more interactive.

One of the empirical studies in this research, think-aloud protocols, provided direct information on the cognitive processes of the test-takers. The researcher was able to deduce from the think-aloud tasks that the UEE test items and tasks not only measured test-takers' listening and reading comprehension but also measured their command of vocabulary, grammar, pronunciation and spelling (language knowledge) in the various sections of the Exam. The data from the think-aloud tasks can also be used to infer the strategies adopted by test-

takers, particularly in difficult test items or tasks. These included classifying words into different forms, word collocations, and elimination strategy. Some test items required participants to cope with topical knowledge. It is also noteworthy that test-takers deployed abilities other than language ability, such as testing mathematics calculation and guessing answers. In sum, the researcher observed during the think-aloud protocols the variety of ways in which the participants engaged with individual characteristics in the test items or tasks.

Different test items or tasks involve differing degrees of interactiveness. Some tasks engaged language ability and strategic competence more than others, for example, a reading or listening comprehension task with multiple-choice items, or an integrated task with gap-filling items. In the think-aloud tasks, after participants finished reading or listening to a passage and questions, they chose an answer from a set of four options. In this instance test-takers had simply to eliminate options. These kinds of questions had a very limited amount of interactiveness compared to the gap-filling items in the cloze section. The integrated task tested not only grammar and vocabulary but also reading ability and strategic competence. This kind of test task was very engaging and interactive because of the range of language competences and skills involved.

9.2.5 Impact

Findings from one of the empirical studies in this research, classroom observations and interviews, showed that the UEE had a strong *impact* on different aspects of pedagogy: teaching, learning, textbook, curriculum and assessment.

- Teaching: Teachers did not teach the whole unit from the textbook because some skills, such as listening and writing, were not assessed to any significant degree, whereas speaking was not compulsory in the UEE. The teacher just focused on whatever method she felt comfortable with and which was the most expedient means of preparing her students for the UEE; she spent more limited class time practising grammar, vocabulary, cloze and reading in class.
- Learning: Different beliefs and behaviour were generated by the UEE for students. Students would usually review new words and exercises at home every day after class and spent time memorising vocabulary just for the UEE. They also believed that techniques and strategies could be learned to cope with different types of question

items in the UEE papers. Most students thought their English learning beliefs and behaviour would be very different if the UEE was cancelled.

- **Textbook:** The teacher thought the textbook alone was not enough for the UEE so other materials such as cloze exercise books and newspaper supplementary exercises were used in class. Those materials or exercises mirrored the UEE format. Many publishers noticed that school teachers did not follow every task in the textbook; therefore they published supplementary exercises to fulfil teachers' needs.
- **Curriculum:** The teacher thought that the ECS were quite abstract and vague and bore little relationship to real teaching in school. She hoped more information and details would be provided that could be applied in her teaching. Andrew, the UEE designer, said that the designers used the ECS as a reference when designing the exam papers. However, they had to read the Examination Specifications, which were written by the UEE designers, before they constructed the UEE papers.
- **Assessment:** In school, the students had monthly tests provided by the publishers of the school newspaper, and the content of the tests mirrored that of the UEE. The teacher reviewed the tests' answers and gave students detailed feedback every time they finished their monthly tests. Other than monthly tests, students were given cloze exercises every day.

The *impact* of test use operates at two levels: a micro level – the individuals who are affected by the particular test use – and a macro level – the educational system or the wider society.

At the micro level, test-takers and teachers are the individuals most directly influenced by the UEE. Regarding test-takers, the UEE not only has a major impact on the content of learning, but also affects test-takers' beliefs and behaviour. When they studied, they would consider whether the materials were helpful in the UEE. 'One test determines a life' is a Chinese saying, and most test-takers consider that their UEE result would set the direction of their life. Some senior high school students thought their attitudes to learning English would be very different if there was no UEE.

Regarding teachers, class teaching appeared to be directly affected by the UEE, and focused mostly on the grammar and vocabulary, cloze and reading comprehension sections, which account for most marks in the exam. Teachers believed that 'teaching to the test' would help students get higher scores in the UEE. One way to minimise this essentially negative impact on

teaching is to change the format of the UEE so that the test items and tasks relate more closely to the teaching programme.

At the macro level, the educational system or the wider society includes universities, senior high schools, and the publishing industry. The use of the test score interpretations has different consequences for different stakeholders in the educational system and in the wider society. University faculty members use the UEE to identify the best students, whereas senior high school principals use outstanding UEE scores to promote their school's reputation. The UEE can influence teaching practice, language programmes, and students' activity schedules in schools. School principals often maximise their use of resources in school in favour of preparing students for the UEE. When choosing school textbooks and supplementary exercises, schools also have to consider whether the exercises are useful to their UEE candidates. Textbook publishers publish supplementary exercises mirroring the UEE format, so that students can practise more tasks to prepare for the UEE.

Finally, when assessing the impact of the examination, facets of particular testing situations should be considered in terms of general individual and societal values and goals. These facets include purpose, target language-use domain, the profile of the test-taker, and construct definition of the test. This is especially important for a high-stakes exam such as UEE, as well as in terms of impact on the specific groups (both micro and macro) discussed immediately above.

9.2.6 Practicality

Practicality can be considered in two ways: internal to the test (test content and format) and external to the test (test use and administration).

In terms of test content and format, most questions are multiple-choice with only small parts devoted to gap filling and short answers. The time and the human resources required to mark the UEE is comparatively limited; thus the UEE is relatively practical. However, the translation and guided-writing test tasks require judgemental marking, which takes more time and resources. Indeed, in order to standardise the scoring method for the writing tasks, scorers need to be trained. Essay writing tasks or integrated tasks are more authentic than multiple-choice items and the real language ability of the test-takers can more accurately be measured; however, the scoring of these kinds of tasks also requires a great number of resources. Similarly,

the speaking tasks in the Oral English Test are conducted in a laboratory with the aid of computers. Test-takers had to answer the questions communicated via the screen and earphones through the microphone. It is also argued that the test is not authentic enough because test-takers never talk to the computer in their daily lives. Again, a speaking test like the IELTS Speaking Section involves many resources in terms of scoring, time and space resources for the exam.

In terms of test use and administration, three types of resources are relevant for assessing practicality: human resources, material resources, and time. Human resources include the people involved in test development, for example, UEE designers, scorers, administrators, technicians, and invigilators. Material resources include space for designing the UEE papers, space for printing equipment for the UEE papers, studios for recording listening tests, and computer systems for marking. Time includes development time for specific tasks such as constructing the papers, administering, scoring, analysing, and final reports.

The UEE is a very high-stakes exam. Every year over nine million students take this exam and consequently it demands a huge number of resources. Andrew, the UEE designer interviewed as part of this research, told the researcher that the government would expend as much as they needed to operate the examination smoothly. The UEE is held on the first weekend in June every year. All schools are closed to make the maximum number of classrooms available as exam centres, and to allow teachers to administer and invigilate the exams. Every year, the UEE designers assemble a month before the UEE to construct the question papers; raters scoring the UEE papers usually finish before the results are released in early July. After that, the exam designers issue an analytic report to the public in September.

Considerations of practicality are likely to affect policymakers' decisions at different stages in the process of test development and use. The considerations include the test content and format of the UEE, which are related to the Examination Specifications. The policymakers might have to revise the Examination Specifications at an earlier stage. However, changes in UEE content and format will affect the other five test qualities.

9.2.7 Conclusion

The most important considerations in designing and developing a language test are the six test qualities: reliability, construct validity, authenticity, interactiveness, impact, and practicality.

These qualities do not exist in isolation. In this section, the most important five inter-relationships among these six qualities will be analysed.

The inter-relationship between the first two qualities – *reliability* and *construct validity* – is crucial to the usefulness of any language test. Reliability is a necessary condition for construct validity, but not sufficient unto itself. A multiple-choice test of reading comprehension or another multiple-choice test of listening comprehension might yield very reliable scores, but not be sufficient to justify the integrated skills of reading/writing or listening/speaking respectively for the construct validity. Those integrated skills tasks require more subjective marking than multiple-choice test items. Although integrated skills tasks can achieve a higher degree of construct validity, they do undermine the reliability.

The second inter-relationship is that between *construct validity* and *authenticity*. Both construct validity and authenticity depend very much on how the construct of language ability for a language test is defined. Authenticity is a measure of the relevance between the test task and the target language use (TLU) domain. This relevance specifies what construct the test tasks include in a specific domain so as to achieve a higher degree of construct validity. Should the Examination Specifications for the UEE follow the English Curriculum Standards, or should the UEE follow the language ability theory model? That is, what construct should the UEE measure and what test content should the UEE contain? Policymakers should define the construct of the UEE very clearly.

The third inter-relationship is between *construct validity* and *interactiveness*. Similar to both authenticity and construct validity, interactiveness also depends upon how language ability is defined for a given language test. Interactiveness is a function of the relative involvement of aspects of language knowledge, strategic competence, and topical knowledge. It was found that the grammar section in the 2014 UEE was a gap-filling passage compared to the past multiple-choice discrete items. This integrated task tested not only grammar and vocabulary, but also reading ability and strategic competence. This kind of task was very interactive because a higher range of language competences and skills is involved. Since it requires very high involvement of areas of language knowledge and skills, it will provide a higher degree of construct validity.

The fourth inter-relationship is that between *authenticity* and *interactiveness* on the one hand, and *reliability* and *practicality* on the other. Some individuals regularly complain that the UEE

test tasks are not authentic and/or interactive enough because most of the questions are multiple-choice items that university students seldom use in the course of daily life (authenticity), and that do not adequately engage test-takers in their language ability (interactiveness). More integrated and real language use tasks could be included in the UEE, by using combined listening/reading and writing tasks or group discussion tasks. However, those kinds of tasks will undermine the *practicality* of the UEE because they are very costly and substantially more resources (time and raters) will be required in marking the subjective tasks. Since the assessment is subjective, those tasks will also weaken the *reliability*, which includes intra-rater reliability and inter-rater reliability.

The last important inter-relationship is between *authenticity* and *impact*. In the UEE situation, the teaching and learning tasks carried out in schools have no correspondence with the use of language in the real world. The question of whether the UEE is an achievement test or a proficiency test has a bearing on this. To the extent that the test is regarded as an achievement test, both authenticity and impact will be compromised. If the UEE is a proficiency test, the test task characteristics and therefore the instructional tasks become more relevant to a real-university-life domain, and the test becomes more authentic. If the test tasks are made more authentic, then the instructional tasks in school will also be more authentic; as a consequence, a positive impact on instruction will be induced.

It may seem that limited resources are available given the scale of the examination, with over nine million candidates in China and 50,000 candidates in Shanghai in 2015 (China Education Online, 2015). Surely the incorporation of additional language use test tasks to the UEE would burden the reliability and practicality of the exam. But the UEE is a crucially important examination and the candidates selected for the universities from the UEE will be the country's future leaders. This researcher believes that more resources should be devoted to raising the level of construct validity and reliability, and to devising a more interactive and authentic examination.

This need not be done nationally all at once. China is a big country and cities in the interior have not developed as fast as coastal cities. In particular, first-tier cities such as Beijing and Shanghai have more resources and qualified professionals for designing and developing the UEE. Examination reform should be introduced and implemented first in Shanghai and other well-resourced cities and provinces.

In summary, the six test qualities cannot be assessed separately. They must be evaluated in terms of their combined effect on the overall usefulness of the test and every effort should be made to achieve an optimal balance among them (See Figure 15). In a high-stakes exam like the University Entrance Exam in China, test designers should aim for the exam and the test tasks to achieve the highest possible levels of *construct validity* and *reliability*. Construct validity enhances higher degrees of *interactiveness* and *authenticity* in a test task, that in turn has a positive *impact* on teaching and learning in schools. Regarding reliability, in order to maintain high levels of internal consistency, stability, and parallel-form reliability, more resources should be devoted to training raters in subjective test-task marking and training professional exam designers to set up high quality test items in a test bank. In short, the researcher is certain that it is worth having a higher degree of construct validity. A higher level of construct validity would lead to interactiveness, authenticity, and positive impact, and reliability and practicality could be underpinned by devoting more resources to training more raters and exam designers in the UEE.

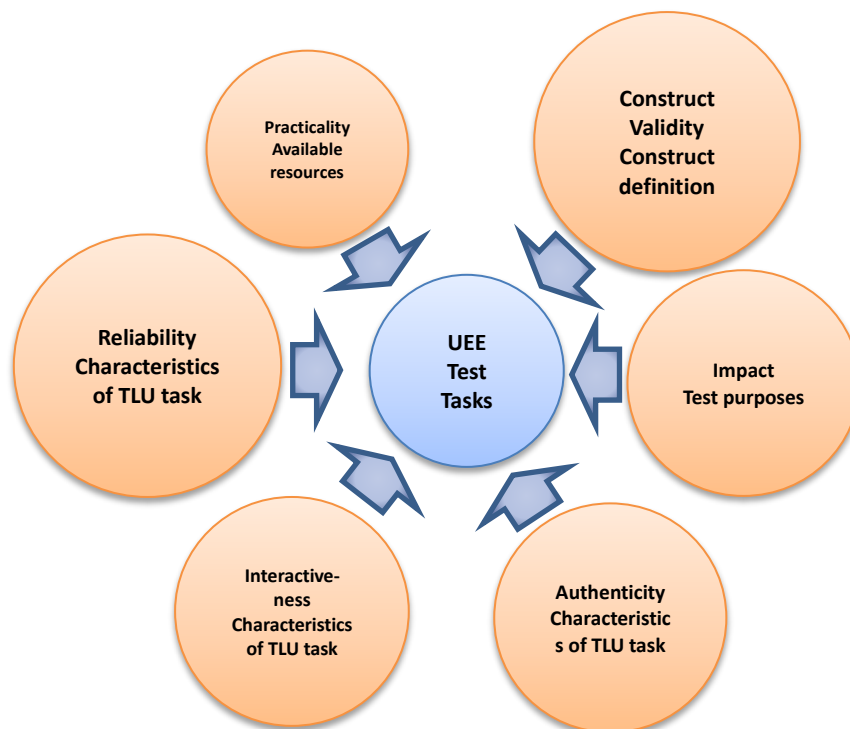


Figure 15: Considerations of test qualities in designing and developing the UEE test tasks

9.3 Major issues arising from the study

9.3.1 Target language use (TLU) domains

9.3.1.1 Purpose of the UEE

It is not possible to determine to which TLU domain the UEE is relevant without answering the question “what is the purpose of the University Entrance Exam?” The main purpose of the UEE is to select the best students – those who have learned their subject well according to the English curriculum – and the most talented students – those who will study academic subjects in universities.

It is useful to draw comparisons with Hong Kong and Singapore because there are some parallels in the circumstances of English language teaching in the two cities. In respect of Hong Kong, a new university entrance exam in Hong Kong was introduced in 2012, known as the Hong Kong Diploma of Secondary Education (HKDSE). English is one of the four core subjects in the HKDSE. The overall aim of the assessment is to evaluate test-takers in achieving their learning targets and objectives of the curriculum (Education Bureau, HKSARG, 2007). In addition, the main medium of instruction at universities is English. The content and format of the English Language Exam papers of Hong Kong focus on language skills, viz., reading, writing, speaking and listening, and the content of the test tasks is more related to university situations. English is one of the official languages other than Chinese. Hong Kong is also one of the major global financial centres in the world and considered to be an Asian World City. Therefore, students in Hong Kong have more opportunities to use English in their daily lives than students in mainland China.

Singapore provides another possible parallel. English is the medium of instruction in schools and one of the official languages. Bilingualism is a keystone of the Singapore education system and students learn both English and their mother tongue in schools (Ministry of Education, Singapore, 2010). High school students are expected to be able to communicate effectively in spoken and written English by the time they finish senior high schools. The UEE in Singapore is conducted to assess the suitability of selected test-takers for entering into undergraduate courses in which most subjects are taught in English. In comparison to China, there is a much higher level of English language fluency because Singapore is a multicultural society in which English is both an official and widely spoken second language. Similar to Hong Kong, the

content and format of the English Language Exam papers of Singapore focus not only on language skills, but also on language knowledge – grammar and vocabulary. The content of the test tasks are more related to university situations, for example, the articles are academic reading materials, and in the writing tasks, test-takers are required to write a letter to a university lecturer to express their ideas on particular contexts.

Compared to Hong Kong and to Singapore, the purpose of the UEE English Language Exam in China is more general and hardly relevant to English language skills and competence. One of the reasons is that Chinese, not English, is the teaching medium in most universities in China. That is why the purpose of the UEE English Language Exam is not to select potential candidates who are going to study subjects in English. In addition, the test users, such as university faculty members consider the English language score as just one of the UEE subjects like mathematics, but not as their English language ability for future study in the universities.

In the past 10 years, there have been more and more students going to study at overseas universities after they graduate from high school in China. The UEE scores become more important because they are used as a part of academic references for the candidates, one which is needed in order for test-takers to be considered eligible to apply to a post-secondary institution. At present, only a small number of overseas educational institutions accept the scores of the UEE in place of traditional entrance test scores. Yu Jihai, deputy director of the Division of International Education at the Chinese Ministry of Education, confirmed at a conference that more universities outside of China will soon accept the UEE for university entry and the Chinese Government is currently working on having overseas countries recognise China's UEE scores (Wang, 2015). This trend is becoming popular. Therefore, when UEE designers reconsider the purpose of the UEE English Language Exam, they should take into account that more Chinese students are going to study overseas and the purpose of the UEE is no longer to only select the best or most talented students study in Chinese universities, but also assess the suitability of selected students for entering into university programmes in which most subjects are taught in English.

9.3.1.2 TLU Domains: language instructional versus real-life

Achievement tests are designed to measure the extent to which learners have mastered the materials in a particular language course or programme. They are directly related to its purpose of how successful an individual learner has been in achieving the objectives of the

course or programme. In contrast to achievement tests, proficiency tests are designed to measure general ability in a language irrespective of any study in that language. The content of a proficiency test should be based on a specification of what test-takers are able to do in the language so as to be proficient.

Some people consider the UEE a proficiency test because it is designed to measure students' ability to use language for academic studies (Qi, 2007), while some say the UEE is an achievement test because it is directly related to the language course and its detailed course syllabus (Wang, 2007). However, the distinction between proficiency tests and achievement tests is becoming increasingly unclear (Brindley, 1991). If the UEE is used to determine whether a senior high school student's English is good enough to study at a Chinese university, then this test is measuring the level of English required to follow courses in particular programmes. This means that the test is measuring whether the student is able to have sufficient command of the language for a particular purpose. The particular purpose in this case is to measure whether students will be able to study academic subjects in universities and especially those taught in English. Therefore, the UEE should be considered a proficiency test. However, the UEE is still governed by the English Curriculum Standards and UEE Vocabulary Handbook for English Language. Based on the above reasons, the researcher thinks that the UEE should be considered partly a proficiency exam and partly an achievement exam. In fact, the UEE performs more like an achievement test but it is not a very adequate one.

At present, there are two public examinations in Shanghai – the University Entrance Exam (UEE) and High School Graduation Exam (HSGE); both exams have very similar formats and content. However, these two exams (the UEE and HSGE) carry out different functions. The UEE focuses on selecting the best students from the secondary schools to continue their studies in the universities, whereas the HSGE focuses on the measuring senior high school students' achievement of the English curriculum's learning objectives. This researcher suggests that the difference in function between these two exams should be more obvious: both format and content should be different so as to ensure that each achieves its prescribed functions effectively. The UEE should not be limited to any vocabulary level in the Vocabulary Handbook and the language skills and language knowledge of the ECS. The UEE format could be more flexible and it could also be made to examine more real language use, as in university; whilst the HSGE content should focus mainly on language knowledge and skills as required in the ECS.

There are two general types of target language use domains: real-life domains and language instructional domains (Bachman & Palmer, 1996). Real-life domains are areas in which language is used for communication purposes, whereas language instructional domains comprise circumstances in which language is used for language teaching and learning purposes. Test-takers in the UEE are students studying English as a foreign language in senior high schools. The test needs to be able to give feedback on the extent to which the students have mastered the English language in class. This places students in the language instructional domains because the purpose of students' going to school is to learn English. However, some people may argue that the ultimate goal of students going to school is to take the UEE and then enter university. Since the tasks test-takers engage in the language instructional domain relate closely to those in an equivalent real-life domain, test designers can use tasks in either domain or both as a basis for developing test tasks. As mentioned, the main purpose of the UEE is to select the best students for university study, so the UEE designers should construct the test tasks in an appropriate real-university-life domain.

Since there are two public examinations for senior high school students – the UEE and HSGE, this researcher suggests that these two exams focus on two different domains. The UEE could constitute a real-life domain and base test-tasks exclusively on real-life tasks, especially university-life related, whereas HSGE could constitute a language instructional domain that based test-tasks on more practice tasks. The HSGE could possibly be more authentic because the characteristics of the HSGE will be very similar to those of TLU tasks, that is, instructional tasks. Therefore, the HSGE would create a positive impact on test-takers and test-users in school. Meanwhile, the UEE designers could maximise authenticity and thereby have a positive impact on instruction by designing test tasks with more real-university-life characteristics. This researcher believes that teachers could construct both real-university-life and instructional domains in school for both UEE and HSGE respectively at the same time. As a consequence, students would benefit from both positive impact on teaching and learning in schools because of both exams with different domains.

9.3.2 *Assessing English Curriculum Standards*

9.3.2.1 Revisiting the English Curriculum Standards

Does the UEE realise the objective of the English Curriculum Standards (ECS)? The overall goal of the ECS is to develop the student's comprehensive language ability by fostering positive

values and attitudes, developing effective learning strategies, and enhancing cross-cultural awareness. These strategies should combine to make the students gradually become independent learners (Wang, 2005). But this study has shown that the UEE papers focused on language knowledge and language skills, which are only two out of the five learning outcomes of the ECS. The UEE does not focus on the other three, viz., affect and attitude, learning strategies, and cultural understanding.

To briefly review these learning outcomes:

- *Affect and attitude* inspire the English learning process, and the crucial tactic in mastering English is to maintain a positive learning attitude;
- *Learning strategies* help improve students' English, and therefore, teachers and students are encouraged to design their own strategic learning plans and set their own goals;
- *Cultural understanding* of English speaking countries is good for student's English learning, and teachers are encouraged to bring intercultural communication awareness and knowledge into the classroom through different methods.

These three aspects of English language learning are not easily assessed and as a result, very few items in the UEE address them. It is probable that because these aspects were not tested, teachers did not think they were important. Nonetheless, this is the area of assessment in which the UEE should be most effective.

If policymakers think these three aspects are important, they should incorporate them into the UEE assessment, thereby prompting teachers to include the relevant content. Meanwhile, policymakers should investigate how best to implement these three areas in the classroom.

In the discussion which follows, this researcher identifies ways in which the assessment of these elements in English language learning can be greatly improved.

First, this researcher suggests that projects or tasks like the following be incorporated into classroom assessment: 1: students be required to read a book, write a book report, and present it in class; 2: students be required to report to their teacher regularly and present individually on their progress in class, and 3: students be required to form groups and work on a topic related to world cultures; the groups would then be required to write a report and present their project in class. These tasks would stimulate student interest and cultivate an

awareness of world cultures, and would be more easily assessed in class than in an exam; therefore, these suggestions would be valuable in assessing these learning outcomes in the classroom.

Second, policymakers should also investigate how student performance in these areas can be assessed. Student's performance can be assessed through a range of practices, including teacher assessment, self-assessment, peer assessment, portfolio assessment. Assessment by teacher provides feedback to students to guide improvements in their study. Self-assessment is the process of students making evaluation of their own learning (Boud & Falchikov, 1989). Peer assessment provides learners with the opportunity to take responsibility for analysing, monitoring, and evaluating the learning process and outcome with their peers (Cheng & Warren, 2005). Portfolio assessment encourages learners to evaluate what they have learned in the whole phase of study (Hamp-Lyons & Condon, 2000).

Third, assessment practices involve different tasks and activities, group projects, and process writing. Different tasks and activities of these three aspects can be used to monitor learners' progress (Black & Wiliam, 1998, 2009). They include listening tasks, speaking tasks, reading tasks, writing tasks, and tasks involving an integration of skills. Regarding group projects, teachers should assess the process as well as the product by observations, discussing and reviewing learners' work, which includes content, structure and organisation, language use and team work in the projects. The writing process above is made up of stages: pre-writing, focusing ideas, evaluating, structuring, and editing (Graham, 2011). Teachers should give feedback on learners' drafts at appropriate stages in the writing process. In addition to teacher feedback, learners can also be encouraged to provide feedback on their own drafts (self-assessment) and those of others (peer assessment). Self-assessment stimulates learners' reflection on their own work, whilst peer assessment empowers learners to learn from one another.

Fourth, when teachers are first asked to introduce these three aspects – affect and attitude, learning strategies, and cultural understanding – in the classroom, they might lack knowledge or ideas. Extra resources should be provided to schools and teachers, such as seminars and training programmes presented by policymakers. Textbook publishers should include more information related to these areas in textbooks. And additional teaching materials should be given to the teachers. Communication between Curriculum developers and UEE designers is

the most important element in the whole initiative. The two sides should study how these areas, viz., affect and attitude, learning strategies, and cultural understanding, could be located in the assessment. For example, should they feature in reading, writing, listening or speaking? Should the two sides also study how these aspects can be actually assessed in schools?

9.3.2.2 School-based assessment

This is also an appropriate place to discuss school-based assessment (SBA), because the introduction of such a system would facilitate the assessment of affect and attitude, learning strategies, and cultural understanding. A similar system has been introduced in the Hong Kong assessment system; this is of obvious relevance to the Mainland Chinese situation. SBA is an assessment carried out by schools with students where the tasks are assessed by their own teacher (HKEAA, 2012). Compared with other forms of assessment, there are a number of distinctive characteristics (HKEAA, 2012):

- The SBA process is carried out in classrooms and lasts for a whole school year;
- Teachers fill an important role, which includes planning the SBA programme, developing suitable assessment tasks and marking the assessment;
- Students have to submit different assignments for assessment over a period of time and the assignments are assessed by their own teachers;
- The SBA involves students' active participation, and encourages students to communicate with their peers and teachers;
- Students receive constructive feedback from their teachers; SBA stimulates continuous appraisal and fine-tuning of the teaching and learning programme;
- SBA complements other forms of assessment, including internal and external examinations.

School-based assessment has been criticised as a non-fair assessment system compared with traditional external examinations, but in fact, it has advantages (HKEAA, 2012):

- The scope of tasks assessed is more diverse, compared to limited test tasks in public exams;
- Validity is enhanced by assessing factors that cannot be covered in public exams;

- Reliability is advanced by assessing more than one task by a teacher who is familiar with the students;
- The assessments are more realistic and authentic, for example, oral presentation of a project or writing a book report;
- Students are able to receive immediate and constructive feedback from their teachers, thereby enhance learning;
- Continuous assessment encourages students to work regularly and provides crucial information for both teachers and students about learning and teaching practice.

In summary, in order to facilitate the assessment of affect and attitude, learning strategies, and cultural understanding, school-based assessment should be introduced in the assessment system for university entry in China.

9.3.3 UEE reform

9.3.3.1 The latest UEE reform

In September 2014, the Chinese Ministry of Education announced the implementation concepts of the Chinese State Council on the deepening of Reform of the Examination and Enrolment System (In Chinese, 国务院关于深化考试招生制度改革的实施意见) to be carried out in 2017 (MOE, 2014). A comprehensive UEE Reform and a reform of the set examination subjects are included. The reform enhanced the linkage between the UEE and the HSGE. The total score is composed of the UEE scores for Chinese language, mathematics and English language (as a foreign language) and HSGE scores for three elective subjects (politics, history, geography, physics, chemistry, and biology). Students will have a second opportunity to take the English Language Exam; the higher mark of the two will go on the student's record. There are two breakthroughs in this announcement: 1: the UEE score for three elective subjects will be replaced by the HSGE, and 2: students will have two opportunities to take the English Language Exam.

Scheduling two exams each year has implications for test design, however. At present, the UEE is set up a month before the examination (in June) by a group of UEE designers. With two exams, two groups of the designers will be assembled at two different periods. This will induce a problem, that of the parallel-form reliability of the test items and tasks for the UEE. In this study, stability and parallel-form reliability were tested in a mock-exam activity involving two

UEE papers (2008 and 2011) taken by participants over a month. However, the correlations were not very satisfactory. The main reasons were that two UEE papers were designed by two different small groups of designers and the chosen materials were different.

In order to maintain a high internal consistency and stability of UEE test items, the Examination Centre should set up a test bank, which is a testing resource for ready-made test items and tasks. Professional designers, such as high school teachers and university professors, should be assembled regularly several times a year to design more test items and test tasks so that more question-items are ready for the examination. In order to standardise the test items, designers should also be offered more professional training. Regarding the security and confidentiality of the UEE, technical issues will need to be discussed and investigated.

9.3.3.2 Impact on tertiary education

The University Entrance Exam has a strong impact not only on the secondary education system but also on tertiary education. At present, English courses in the university focus mainly on grammar, vocabulary, translation, and reading comprehension, and so are very similar to English teaching in high schools. College English Test (CET) is the only national standardised and recognised English language test in China that measures language ability of non-English major university students (Jin, 2010). CET has a very strong negative washback on English courses in tertiary education; for example, teachers present strategies for doing well in the CET and teach to the exams in class (Gu, 2005). Most university students focus on the CET and learn nothing from the English courses that is related to their professional major.

In addition, the format and content of the CET is very similar to that of UEE. UEE reform accordingly will have a very strong impact not only on the CET exam but also on the English curriculum in the universities. If the UEE format and content was changed to take in more language use tasks, CET might have to change as well and the English curriculum at tertiary level would be required to adjust. This would enhance the efficacy of English language teaching throughout the education system in China.

Different subjects require different levels of English proficiency; for example, International Finance and Trade requires a very different level compared with Engineering or Science, the former needing English much more in their careers. The university should provide English programmes more relevant to likely future careers. That means courses that do much more than teach grammar and vocabulary.

This is not just a matter of preparing students for the workplace. It is also an increasingly important issue inside the university. Bilingual teaching in Chinese universities is a recent trend (Li, 2012). The teachers use both Chinese and English in class. This reform poses a challenge to students because they have to handle the subjects' knowledge well but also English. Students are required to have a good command of English. They must use English textbooks and academic articles, and sometimes express themselves in English in class. Sometimes, visiting scholars from overseas give some seminars in the universities; these will also be in English.

Moreover, a number of Chinese universities have organised joint programmes with overseas universities; students in the joint programmes can opt to either stay in China or study overseas in the third and fourth years. Students enrolled in such programmes may be required to read many English articles and books, write assignments and present projects in English. More real-life language tasks should be taught to adapt students to this life. The tertiary English curriculum should also be made more realistic.

In sum, when UEE is being reformed, tertiary education policymakers should consider reforming both the English curriculum and the CET exam.

9.3.3.3 The UEE goals

There are no details about the UEE reform in 2017 except that there will be two English Language Exams in a year. As yet, nobody knows what the 2017 test content and format are going to be. UEE designers have to devise test format and content that conforms to the reformed UEE, once they have received confirmation from the Ministry of Education. They also need to ask: whether the Examination Specifications for the UEE should follow the English Curriculum Standards and whether the UEE should follow the language ability theory model? That is, what construct should the UEE measure and what test content should it contain?

Construct validity denotes the appropriateness of inferences made on the basis of observations or measurements (often test scores) – specifically whether a test measures the intended construct. Do students who get high scores from the UEE expect to become proficient in English? A high score in UEE more likely indicates only that students have mastered grammar, vocabulary, translation, reading and listening comprehension. That is the traditional construct. Should the new UEE include more practical tasks that they normally did in school, or should it focus on the kinds of integrated language use tasks included in IELTS or TOEFL? Two

interviewees said after the mock exam activity that the IELTS sample test was more difficult than the UEE papers because IELTS included many more language use tasks than UEE, including practice tasks.

When Educational Test Service (ETS) in the U.S. carried out a major reform on the Test of English as a Foreign Language (TOEFL), the project called TOEFL Framework 2000, they set up new goals and directions to determine a work plan for the project (Jamieson, Eignor, & Kunnan, 2008). The directions, regarding what construct the test should measure, what the test should consist of, and what validation work would be needed, were identified as follows (Jamieson et al., 2008):

- Construct: The test should accurately reflect communicative competence, that is, the ability to put language knowledge into use in relevant contexts;
- Test content: Test tasks should require test-takers to perform tasks that resemble those that they will encounter in the real world and reflect current theories of communicative language use in the academic setting;
- Validation: The test should be subject to ongoing investigations to provide evidence about score meaning and the consequences of score use.

The major change for the TOEFL test was that the tasks for four skills (listening, reading, speaking and writing) were no longer tested separately; test-takers are now asked to perform tasks that combine more than one skill. In respect of the Speaking Section, integrated reading/speaking tasks ask test-takers to produce discourse in the form of a monologue by using information from various academic reading texts, whereas integrated listening/speaking tasks use materials from different interactions, conversations and lectures (Jamieson et al., 2008). Regarding the Writing Section, the integrated reading/writing tasks require test-takers to communicate their ideas by organising information from a reading text or selected materials, whilst in the integrated listening/writing tasks, test-takers are asked to describe problem-solving tasks and write summaries (Jamieson et al., 2008).

Similarly, when the UEE is being reformed, designers should think about goals and directions in three ways: 1: what construct should the UEE measure; 2: what should the UEE consist of, and 3: what is the justification for the UEE? They should also consider whether the UEE is measuring if the student is able to have sufficient command of language for a particular

purpose, or measuring the senior high school students' ability to achieve their learning objectives throughout the English curriculum. Should the UEE test students' grammar, vocabulary or translation or test them more on language use tasks?

9.3.4 *Future assessment system for university entry in China*

9.3.4.1 Assessment system for university entry in Hong Kong

A major education reform for senior secondary academic structure was implemented in September 2009 in Hong Kong (HKSARG Education Bureau, 2007). The senior secondary stage was changed from four years to three, and a new university entrance exam was introduced in 2012, the Hong Kong Diploma of Secondary Education (HKDSE). HKDSE provides a common end-of-school credential that gives access to university study. English is one of the four core subjects in the HKDSE framework.

The main purpose of the new English language curriculum is to provide more opportunities for English learners to broaden their knowledge and experience of foreign culture as well as opportunities for personal and intellectual development through study, leisure and work in English. The public assessment for the university entry in senior secondary school includes both public examinations and moderated school-based assessments (SBA), which are conducted by the Hong Kong Examination and Assessment Authority (HKEAA). The public examinations are more summative whereas the SBA in public assessment enhances formative assessment, within the context of HKDSE.

There are four papers in the public examination, viz., Reading, Writing, Listening and Integrated Skills, and Speaking. They account for 20 percent, 25 percent, 30 percent, and 10 percent respectively (HKSARG Education Bureau, 2007). Different types of items in different papers are used to measure test-takers' performance in a wide range of language skills and abilities. The question items include multiple-choice questions, short questions, extended opened responses and essays in Reading and Writing papers. In the Listening and Integrated Skills paper, test-takers are required to process information by using and organising various materials from spoken and written sources through listening and reading skills to accomplish different listening and writing tasks. In the Speaking paper, test-takers' performances are measured through individual and group interaction tasks.

The SBA accounts for the remaining 15 percent of the public assessment. There are two parts. Part A includes a reading/viewing programme where students are required to read/view four texts over the course of three years. They are required to write comments and personal reflections, and then participate in a group discussion or make an individual presentation on the texts they have read/viewed. Part B comprises either a group interaction or an individual presentation based on the elective modules of the curriculum. The SBA is intended to motivate students in engaging in extensive reading and viewing that helps develop their overall language ability, and to develop their speaking ability in discussing issues in depth and expressing their own concepts clearly and concisely (HKSARG Education Bureau, 2007).

9.3.4.2 Proposed assessment system for university entry in China

Having reached this point in the discussion, it is appropriate to present a proposed assessment system for university entry. The Chinese Ministry of Education recently proposed that the total score of English language for university entry should consist of three components: the University Entrance Exam (UEE), the High School Graduation Exam (HSGE), and the School-based Assessments (SBA), but the details have not yet been finalised. This researcher agrees with this proposal, and suggests the following weighting: 40 percent UEE, 40 percent HSGE, and 20 percent SBA. This percentage breakdown is tentative but reflects the relative significance of the three components in developing accuracy and fluency in English.

$$\text{Total score for English language} = \text{UEE (40\%)} + \text{HSGE (40\%)} + \text{SBA (20\%)}$$

The UEE should focus on preparing students for university study. The purpose of the UEE is to measure whether the student has sufficient language for a particular purpose – being capable of studying academic subjects in universities. It should therefore be considered a proficiency test. At present, the UEE has very low concurrent validity compared with the IELTS test sample, which means it does not function as a proficiency test. The UEE should focus more on integrated language use and communicative tasks, and the test format should be modified to be similar to international proficiency language tests such as IELTS and TOEFL. The UEE should have four papers: Reading, Writing, Listening and Integrated Skills, and Speaking. An Oral English Test should be included in the exam; it would have a positive impact on instruction and

teachers. Schools would expend more resources on speaking lessons and teachers will organise more group discussion and presentation activities in class to practise spoken English. Moreover, functional knowledge and sociolinguistic knowledge (as defined in Bachman and Palmer's model of language ability) can be assessed in the Speaking paper.

The HSGE should focus on an instructional domain and have a similar format as at present: Listening, Grammar and Vocabulary, Cloze and Reading Comprehension, Translation and Guided Writing. HSGE is intended to measure the extent to which learners have learned the materials in school; it is an achievement test. The materials in the textbook are based on the English Curriculum Standards, especially language skills and knowledge. The Vocabulary Handbook should still be kept for reference and to ensure that a certain number of words have to be learned by students in school. The test tasks in the HSGE should focus on the basics of grammar and vocabulary and on instructional tasks.

The SBA should include those areas in the ECS which are not easily tested in the examination – affect and attitude, learning strategies, and cultural understanding. The SBA is, and ideally would continue to be, administered and conducted in schools and marked by the students' own teachers. A reading programme could be set up over three years in the senior high school, where students were required to write comments, reports, and reflections on what they have read. The reading texts could be related to values and attitudes, learning strategies, and cross-cultural issues. In addition, students are recommended to have group discussions and oral presentations so as to complement monologue discourse in the Oral English Test. The SBA should be designed to engage students in extensive reading and speaking by writing reports and by having individual or group presentations and discussions. With the three forms of assessment combined in this way, students will be much better prepared for English study or study in English at university than is presently the case.

9.4 Chapter summary

This chapter gave a summary and evaluation of findings by using Bachman and Palmer's model of six test qualities and then discussed the major issues arising from the study. First, policymakers should re-visit the main purpose of setting up the English Language Exam. The debate about the UEE as an achievement test or a proficiency test, and whether UEE test-takers belong to a language instructional domain or a real-university-life domain, has been discussed. Next, the chapter discussed whether policymakers should incorporate three areas

– affect and attitude, learning strategies, and cultural understanding – into the UEE assessment. It also examined how best policymakers should implement these three areas in the classroom and how student performance can be assessed. After that, the chapter reviewed the latest UEE reform and discussed how the UEE reform might affect tertiary education. When the UEE is being reformed, tertiary education policymakers should consider the English curriculum and the CET exam, meanwhile UEE designers should think clearly about the goals and directions of the UEE. The last section of the chapter introduced the assessment system of English language for university entry in Hong Kong and presented a proposed assessment system for university entry in China.

Chapter 10: Conclusion

10.1 Introduction

This chapter presents the major findings of this study, followed by qualifying the significance of the research, and implications for different stakeholders. The limitations of the study are then considered. The chapter ends with recommendations for further research.

10.2 Major findings

The primary aim of this study was to validate the English Language Exam for university entry in Shanghai. There were two phases in this research: 1: a qualitative and content analysis, which compared the past 10 years' UEE papers with English Curriculum Standards and Bachman and Palmer's model of language ability respectively, and 2: empirical studies, which included mock exams, classroom observations and interviews, and think-aloud tasks. The major findings are as follows.

The overall objective of the English Curriculum Standards is to promote students' general ability in language use and whole person development. However, the study has revealed that while the UEE covered most of the objectives in the area of language skills and language knowledge and a small part of learning strategies, nothing was tested in the area of affect and attitude and cultural understanding. One possible reason is that the UEE designers did not familiarise themselves with the details of the ECS when they constructed the UEE papers. Ideally, the ECS should be directly related to the UEE, in which test items are supposed to reflect all the objectives listed in the ECS; however, the UEE was narrowed down by the objectives of the Examination Specifications, which was designed by a group of professionals under the Examination Centre (personal communication with the UEE designer). It may well also be a matter of deliberate choice based on their analysis of what constitutes the core competence that should be included.

When comparing the past 10 years' UEE papers with Bachman and Palmer's model of language ability, the findings showed that the content tested in the UEE mainly focused on organisational knowledge (grammatical and textual knowledge) and strategic competence, but not pragmatic knowledge (functional and sociolinguistic knowledge).

The mock exams provided support for the validity of the UEE in some ways but not others. The correlations between the subtests in both the 2008 and 2011 UEEs showed satisfactory construct validity. The high correlations between the subtests and total UEE scores revealed that the test components significantly contributed to the final total UEE scores. However, the total scores of the 2008 and 2011 UEEs had quite low correlations with the IELTS sample test; which showed that the concurrent validity was low for the UEE. Two interviewees after finishing the mock exams said that the UEE was able to test their language ability; that is, face validity. Regarding reliability, Spearman-Brown reliability coefficients, Cronbach coefficient alphas, and KR-21 reliability coefficients showed that internal consistency was quite high for both the 2008 and 2011 UEEs. However, the stability and parallel-form reliability was low when correlating between the 2008 and 2011 UEEs for different subtests.

The results from classroom observations and interviews confirmed that the UEE had an enormous impact on different stakeholders: teachers, students, textbook publishers, curriculum developers and UEE designers. These impacts were: using examination preparation materials in class that were closely linked to the UEE; pursuing UEE outcomes; publishers issuing extra materials suited to the UEE; the narrowing of instruction and of the curriculum; and focusing on language skills and knowledge. It seemed that those impacts were unintended consequences of the UEE. Moreover, the impact of test use operated at two levels: micro and macro levels. At the micro level, the students and teachers were the most influenced by the UEE. It influenced the content, beliefs and behaviour of both students and teachers. At the macro level, the potential impact of the test use on values and goals of their surrounding society and educational programme should be considered.

In contrast, the findings from the think-aloud tasks showed that the test items and tasks fulfilled the requirements of the Examination Specifications and the response validity of the English Language Exam was confirmed. Participants responded as anticipated by the UEE designers and their performance showed that they engaged with individual facets of the test items and tasks. Nevertheless, from the think-aloud tasks some items were found to be guessed at by participants. Possible reasons for this could be limitations in the participants' language ability, confusing multiple-choice options, and poorly designed test items. A further point: abilities other than language ability were tested in the UEE. This might undermine the construct validity of the English Language Exam.

10.3 Significance of research

This research used both quantitative and qualitative methods including a qualitative and content analysis of the past 10 years' UEE papers, and both quantitative and qualitative methods in the three empirical studies. Quantitative data was derived from measurement instruments that provided participants' scores from the mock exams, whereas qualitative data, which allowed the researcher to describe participants' beliefs and behaviour, was derived from class observations and interviews, and think-aloud tasks. The combination of qualitative and quantitative approaches ensured that the resultant data was both comprehensive and significant. This use of multiple methods enabled the researcher to not only obtain valuable data, but also to comprehend the study thoroughly from different angles. Independent checking of content analysis likewise enhanced reliability. Two independent coders were invited to check the items of the UEE papers in the content analysis. The inter-coder reliability checks ranged from 91% to 99% agreement with the findings of the researcher.

This research has theoretical significance. Little research has been carried out on whether or not these tests are of a quality suitable to such a high-stakes exam as the UEE. Regarding the UEE, very little information is publically available. This study has reported a comprehensive test validation of the UEE in different aspects – content validity, construct validity, concurrent validity, face validity, consequential validity and response validity – and evaluated the qualities of the UEE by using Bachman and Palmer's model. With reference to Bachman and Palmer's model of language ability, grammatical knowledge, textual knowledge, and strategic competence were substantially covered in the UEE. However, functional and sociolinguistic knowledge should also be assessed in a language test. More academic professionals should be encouraged to investigate these two areas of knowledge.

This research also has been of practical significance. The findings from both content analysis and the empirical studies will allow policymakers to make more informed decisions about this high-stakes exam and about language assessment policies. This study proposed an assessment system for university by combining the UEE, HGSE, and SBA. This research also has important practical implications for other stakeholders, particularly universities and schools. It also has implications for teachers and students, especially for teaching and learning English language.

10.4 Implications

10.4.1 Implications for Curriculum developers

What are the implications for Curriculum developers? The UEE bears a very uneven relationship to the ECS in that it covered most objectives in language skills and language knowledge, but very few in the other three areas (affect and attitude, learning strategies, and cultural understanding). Teachers also find it difficult to implement the theories propounded in the ECS when they lack practical examples and resources in the classroom. Therefore, Curriculum developers should provide extra resources, such as seminars and training programmes, to help schools and teachers implement these three areas in the classroom more easily. Curriculum developers should also communicate with UEE designers as to how these three areas could be incorporated in the assessment, and these two parties should study how these areas can be actually assessed in schools. Importantly, when designing and implementing the ECS, the Curriculum developers should communicate with other stakeholders, such as universities and senior high schools principals, explaining what they want to achieve in the ECS.

10.4.2 Implications for UEE designers

The UEE has always been considered the biggest obstacle for Curriculum reform (Wang & Chen, 2012). The format and content of the UEE needs to change in order to reflect the goals fostered by the ECS and the UEE should have a positive impact on classroom teaching and learning. In order to implement the formative assessment policy suggested in the ECS, UEE designers should strike a balance between the imperatives of a high-stakes exam and those of classroom assessment (Gu, 2014a). Further, the UEE designers should provide teachers with professional training programmes on how to carry out classroom assessment in school. When a test is developed, there are potential consequences to evaluate, and some can be predicted (Reckase, 1998). UEE designers should communicate their ideas on teaching and learning with test users so as to reduce any unintended consequences in school from the UEE. From time to time they should change the test-task format in order to make the UEE materials more practical and authentic. In addition, the UEE designers need to clarify their thinking on what and how the exam measures to the test users and test-takers and to explain how the exam scores can be correctly interpreted.

10.4.3 Implications for universities and schools

The UEE could constitute a real-university-life domain that bases test-tasks on real-university-life. If the UEE format and content was changed to take in more language use tasks, both the English exams and the curriculum in universities might have to be amended to be consistent with the UEE. On the other hand, school-based assessment should be introduced; it would facilitate the assessment of affect and attitude, learning strategies, and cultural understanding in schools. Schools might need to reallocate teaching resources between the public examinations and school assessments. Schools should also direct and support their English teachers and provide them with enough resources so that the teachers can execute their lesson plans to the benefit of the students.

10.4.4 Implications for teachers and students

This study revealed that content of teaching and learning, beliefs and behaviour of teachers and students are influenced by the UEE. The teacher observed in this study mainly focused on which was the most expedient means of preparing her students for the UEE and most students thought their English learning beliefs and behaviour would be very different if there was no UEE. Most students interviewed thought that the UEE should be extended to communicative English in the four skills of listening, reading, speaking and writing, and they believed Oral English was crucial and necessary for communication in their daily lives and future study and career.

10.4.5 Implication for textbook publishers

The school had a standard teaching textbook for students, *The Oxford English for Senior II*. This textbook, from Oxford Publishers, was approved by the Shanghai Municipal Education Commission. The content studied should follow the requirements of the ECS but most of the text types or activities from the textbook are neglected as they are irrelevant to the UEE. In order to implement the ECS in the classrooms, the textbook publishers should provide more resources and teachers' training and tell teachers how to use the resources and materials in the classroom so that the content from the ECS will be implemented properly. Textbook publishers should also stop publishing extra materials focusing on the UEE which have a negative impact on teaching and learning.

10.5 Limitations of the study

This study is an individual PhD research project and some limits to the project flow from this.

Although the content of teaching and learning, and the behaviour and beliefs of both teacher and students were observed to be driven by the impact of the UEE over two weeks of classroom observations and interviews, a longitudinal study over a school term would provide richer and fuller information and potential for insight into the impact of the UEE on the behaviour and beliefs of both teacher and students.

The most obvious limitation in this study was the small number of interviewees, a limitation that hinders the ability to make a clear generalised statement about the impact of the UEE. With a bigger sample, including more teachers and students from different senior high schools in Shanghai, the researcher would have been able to make a more robust evaluation of impact. Nevertheless, the small sample in the classroom observations and interviews did not contradict the result data.

The researcher did not have any contact with policy-makers and stakeholders in Shanghai, such as university faculty members, senior high school principals, curriculum designers and textbook publishers. It would have been more comprehensive to have collected information from different perspectives so as to gauge the seriousness of the impact on different stakeholders of the UEE.

As Phakiti (2010) recommends, the larger the number of participants, the more stable the data distribution. A minimum of 100 participants was expected to participate in this study in order to generate a better distribution of a range of scores for quantitative data analysis; however, only 60 participants took part in the mock exams. Further, these 60 participants were Shanghai first-year university students who had taken the UEE nine months before the mock exams. However, the ideal participants would be high senior III students who were ready to take the UEE around the same time; unfortunately they were very busy preparing for the UEE, to be held in three months and schools would not provide access to researchers at that time.

In this study, only the reading and listening sections were used in the mock exams. It would be ideal if all sections including writing and speaking had been taken by the participants so that the data would be complete. Such an endeavour required ample amounts of both time – to mark the writing papers and speaking tasks – and resources – professional examiners and

markers. Moreover, the marking of writing and speaking sections is a more subjective exercise than marking listening and reading sections, which mainly consist of multiple-choice items.

Eight students were divided into four groups in the think-aloud tasks: two groups worked on the 2008 UEE, and two on the 2011 UEE. Within each group, one member worked on grammar and vocabulary, cloze, and reading comprehension and the other member worked on listening comprehension and writing. Each group spent at least three hours on each task. If more time was allowed, stimulated recalls as an introspective method in the interviews could have been done, for example, asking the participants in depth and in detail why they had particular answers or strategies on particular tasks.

10.6 Recommendations for further research

1. As the UEE Designer mentioned in the interviews, he thought that think-aloud protocols would be a good means of checking the response validity of the UEE paper during the test development process. It would be valuable if a researcher had an opportunity to work with the Shanghai Examination Centre to this end;
2. In the mock exams, if time and resources allowed, the writing section and speaking section, including both 2008 and 2011 UEE papers and the IELTS test sample, should be included so a complete mock exam could be implemented by the participants for test validation. As a result, more information and results will be provided for investigating construct validity and other test qualities;
3. By correlating students' past UEE paper scores with their scores on other tests, the concurrent validity of the UEE can be gauged. In this study, an IELTS sample test was used as an external test for correlation. More external tests could also be used to explore the concurrent validity of the UEE; for example, High School Graduation Test or Test of English as Foreign Language (TOEFL). The Shanghai High School Graduation Examination is another public examination. Senior III students sit this exam before they graduate from high schools and are required to achieve the minimum requirements set by English Language Teaching Basic Requirements. TOEFL is another internationally recognised English proficiency test with a quite different test format from IELTS;
4. Predictive validity correlates the results with other measures some time after the test has been given in order to gauge how well a test predicts future success in the target

domain. The researcher could expand the study to students who are planning an English major or plan to study in a university using English as a medium of teaching. The researcher might then collect grade-point averages for each of these students after each successive year of university study, followed by a correlation of the UEE English subject scores with successive annual grade-point averages;

5. One of the major purposes of the English Language Exam for university entry is to select the most talented students for study in academic subjects. Interviews with and questionnaires addressed to university professors would establish whether selecting according to UEE scores delivered the best students. It would also be useful to interview university professors about their overall satisfaction with student proficiency in English;
6. As Roever (2011) says, testing of second language pragmatic knowledge is still under exploration and is a growing area of second language assessment. Findings from the study revealed that very few test items were tested on functional and sociolinguistic knowledge. Validating test of second language pragmatic knowledge will be a new challenging study area in language assessment.

The findings from this research have allowed various insights into the test validation of the English Language Exam for university entry. Other researchers are encouraged to explore more research in this field in the future to further advance knowledge on validating a language test. Limitations of this study have been discussed. If those limitations had been resolved, better and more comprehensive results would have been obtained. Further research on the test validations from different perspectives were recommended and will contribute to the field of language testing and assessment.

References

- Abeywickrama, P., & Brown, H. D. (2010). *Language assessment: Principles and classroom practices*. White Plains, NY: Pearson Education.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standard for educational psychological testing*. Washington, DC: American Psychological Association.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standard for educational psychological testing*. Washington, DC: American Psychological Association.
- Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.
- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13(3), 280–297. <http://doi.org/10.1177/026553229601300304>
- Alderson, J. C., & Wall, D. (1993). Does Washback Exist? *Applied Linguistics*, 14(2), 115–129. <http://doi.org/10.1093/applin/14.2.115>
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F. (2004). *Statistical analyses for language assessment*. Cambridge: Cambridge University Press.
- Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2(1), 1–34. http://doi.org/10.1207/s15434311laq0201_1
- Bachman, L. F., Lynch, B. K., & Mason, M. (1995). Investigating variability in tasks and rater judgements in a performance test of foreign language speaking. *Language Testing*, 12(2), 238–257. <http://doi.org/10.1177/026553229501200206>
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.

- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.
- Baghaei, P., & Amrahi, N. (2011). Validation of a multiple choice English vocabulary test with the Rasch model. *Journal of Language Teaching and Research*, 2(5), 1052–1060. <http://doi.org/10.4304/jltr.2.5.1052-1060>
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–148.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <http://doi.org/10.1007/s11092-008-9068-5>
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum.
- Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education*, 18(5), 529–549. <http://doi.org/10.1007/BF00138746>
- Brindley, G. (1991). Assessing achievement in a learner-centred curriculum. In J. C. Alderson & B. North (Eds.), *Language testing in the 1990s: The communicative legacy*. London: Macmillan.
- Bryman, J. (1992). Quantitative and qualitative research: Further reflections on their integration. In J. Brannen (Ed.), *Mixing methods: Qualitative and quantitative research* (pp. 57–78). Aldershot: Avebury.
- Buck, G. (1991). The testing of listening comprehension: An introspective study. *Language Testing*, 8(1), 67–91. <http://doi.org/10.1177/026553229100800105>
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Cambridge ESOL. (2014). *Cambridge English IELTS 9*. Cambridge: Cambridge University Press.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105. <http://doi.org/http://dx.doi.org/10.1037/h0046016>

- Canale, M. (1983a). From communicative competence to communicative language pedagogy. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 2–27). London: Longman.
- Canale, M. (1983b). On some dimensions of language proficiency. In J. W. Oller (Ed.), *Issues in language testing research* (pp. 333–342). Rowley, MA: Newbury House.
- Canale, M. (1987). The measurement of communicative competence. *Annual Review of Applied Linguistics*, 8, 67–84. <http://doi.org/10.1017/S0267190500001033>
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47.
- Carroll, J. B. (1961). Fundamental considerations in testing for English language proficiency of foreign students. In H. B. Allen & R. A. Campbell (Eds.), *Teaching English as a foreign language* (pp. 313–321). New York, NY: McGraw-Hill.
- Carroll, J. B. (1968). The psychology of language testing. In A. Davies (Ed.), *Language testing symposium* (pp. 46–69). London: Oxford University Press.
- Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. *Applied Linguistics*, 6(2), 5–35.
- Chapelle, C. A. (1999). Validity in language assessment. *Annual Review of Applied Linguistics*, 19, 254–272. <http://doi.org/10.1017/S0267190599190135>
- Chapelle, C. A., Enright, M. K., & Jamieson, J. (Eds.). (2008a). *Building a validity argument for the Test of English as a Foreign Language*. New York, NY: Routledge.
- Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2008b). Test score interpretation and use. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 1–25). New York, NY: Routledge.
- Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2010). Does an argument-based approach to validity make a difference? *Educational Measurement: Issues and Practice*, 29(1), 3–13. <http://doi.org/10.1111/j.1745-3992.2009.00165.x>

- Cheng, L., & Qi, L. (2006). Description and examination of the National Matriculation English Test. *Language Assessment Quarterly*, 3(1), 53–70.
http://doi.org/10.1207/s15434311laq0301_4
- Cheng, W., & Warren, M. (2005). Peer assessment of language proficiency. *Language Testing*, 22(1), 93–121. <http://doi.org/10.1191/0265532205lt298oa>
- Cheng, X. (2011). The “English Curriculum Standards” in China: Rationales and issues. In A. Feng (Ed.), *English language education across Greater China* (pp. 133–150). Bristol: Multilingual Matters.
- China Education Online. (2015). 2015 年全国各地高考报名人数 [The total number of candidates for the university entrance examination in China in 2015]. Retrieved from http://gaozhong.eol.cn/news/201506/t20150602_1266104.shtml
- Chinese Ministry of Education. (2001). 全日制义务教育普通高级中学英语课程标准 (实验稿) [English curriculum standards for full-time compulsory education and senior high schools (trial version)]. Beijing: Beijing Normal University Press.
- Chinese Ministry of Education. (2003). 普通高中英语课程标准 (实验稿) [English curriculum standards for senior high schools (trial version)]. Beijing: People’s Education Press.
- Chinese Ministry of Education. (2011). 义务教育英语课程标准 (2011 年版) [English curriculum standards for compulsory education (2011 version)]. Beijing: Beijing Normal University Press.
- Chinese Ministry of Education. (2014). 国务院关于深化考试招生制度改革的实施意见 [The implementation concepts of the Chinese State Council on the deepening of Reform of the Examination and Enrolment System]. Retrieved from http://www.gov.cn/zhengce/content/2014-09/04/content_9065.htm
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: M.I.T. Press.
- Cohen, A. D. (1984). On taking language tests: What the students report. *Language Testing*, 1(1), 70–81. <http://doi.org/10.1177/026553228400100106>
- Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 3–17). Hillsdale, NJ: Lawrence Erlbaum.

- Cumming, A. (1996). Introduction: The concept of validation in language testing. In A. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 1–14). Clevedon; Philadelphia, PA: Multilingual Matters.
- Cummins, J. P. (1983). Language proficiency and academic achievement. In J. W. Oller (Ed.), *Issues in language testing research* (pp. 108–129). Rowley, MA: Newbury House.
- Custer, C. (2015). *The Gaokao: China's national college entrance exam*. Retrieved from <http://chineseculture.about.com/od/historyofchina/fl/The-Gaokao.htm>
- Davey, G., Lian, C. D., & Higgins, L. (2007). The university entrance examination system in China. *Journal of Further and Higher Education*, 31(4), 385–396. <http://doi.org/10.1080/03098770701625761>
- Davies, A. (1990). *Principles of language testing*. Oxford: Basil Blackwell.
- Davies, A. (1997). Demands of being professional in language testing. *Language Testing*, 14(3), 328–339. <http://doi.org/10.1177/026553229701400309>
- Dello-Iacovo, B. (2009). Curriculum reform and “Quality Education” in China: An overview. *International Journal of Educational Development*, 29(3), 241–249. <http://doi.org/10.1016/j.ijedudev.2008.02.008>
- Denscombe, M. (2007). *The good research guide: for small-scale social research projects* (3rd ed). London: Open University Press.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.
- Dörnyei, Z., & Katona, L. (1992). Validation of the C-test amongst Hungarian EFL learners. *Language Testing*, 9(2), 187–206. <http://doi.org/10.1177/026553229200900206>
- Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge: Cambridge University Press.
- Ebel, R. L. (1979). *Essentials of educational measurement* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

- Education Bureau, Hong Kong Special Administrative Region Government. (2007). *English language education key learning area*. Retrieved from <http://www.edb.gov.hk/attachment/en/curriculum-development/kla/eng-edu/Curriculum%20Document/EngLangCAGuide.pdf>
- Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1), 107–115. <http://doi.org/10.1111/j.1365-2648.2007.04569.x>
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed). Cambridge, MA: MIT Press.
- Færch, C., & Kasper, G. (Eds.). (1983). *Strategies in interlanguage communication*. London: Longman.
- Fredrickson, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist*, 39, 193–202.
- Fulcher, G. (1997). An English language placement test: Issues in reliability and validity. *Language Testing*, 14(2), 113–139. <http://doi.org/10.1177/026553229701400201>
- Fulcher, G. (2010). *Practical language testing*. London: Hodder Education.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. London: Routledge.
- Gass, S. (2010). Experimental research. In B. Paltridge & A. Phakiti (Eds.), *Continuum companion to research methods in applied linguistics* (pp. 7–21). London: Continuum.
- Graham, S. (2011). *Approaches to process writing*. Retrieved from <http://www.teachingenglish.org.uk/article/approaches-process-writing>
- Green, A. (Ed.). (1998). *Verbal protocol analysis in language testing research: A handbook*. Cambridge: Cambridge University Press.
- Greene, L. (2011). *How many English learners in China?* Retrieved from <http://www.economist.com/blogs/johnson/2011/09/english-china>
- Gu, X. (2005). Positive or negative? An empirical study of CET washback on college English teaching and learning in China. *ILTA Online Newsletter* 2. Retrieved from <http://www.iltaonline.com/newsletter/02-2005oct/>

- Gu, Y. (2012). English curriculum and assessment for basic education in China. In J. Ruan & C. B. Leung (Eds.), *Perspectives on teaching and learning English literacy in China* (pp. 35–50). Dordrecht, MA: Springer.
- Gu, Y. (2014a). The unbearable lightness of the curriculum: What drives the assessment practices of a teacher of English as a Foreign Language in a Chinese secondary school? *Assessment in Education: Principles, Policy & Practice*, 21(3), 286–305. <http://doi.org/10.1080/0969594X.2013.836076>
- Gu, Y. (2014b). To code or not to code: Dilemmas in analysing think-aloud protocols in learning strategies research. *System*, 43, 74–81. <http://doi.org/10.1016/j.system.2013.12.011>
- Habermas, J. (1970). Towards a theory of communicative competence. In H. P. Dreitzel (Ed.), *Recent sociology*, No. 2 (pp. 115–148). London: Macmillan.
- Halliday, M. A. K. (1976). *Halliday: System and function in language: Selected papers*. London: Oxford University Press.
- Halliday, M. A. K. (1985). *An introduction to functional grammar*. London: Arnold.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Hamp-Lyons, L. (1997). Washback, impact and validity: Ethical concerns. *Language Testing*, 14(3), 295–303. <http://doi.org/10.1177/026553229701400306>
- Hamp-Lyons, L., & Condon, W. (2000). *Assessing the portfolio: Principles for practice, theory, and research*. Cresskill, NJ: Hampton Press.
- Hatch, E. M., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Boston, MA: Heinle & Heinle Publishers.
- Heaton, J. B. (1988). *Writing English language tests*. London; New York, NY: Longman.
- Henning, G. (1987). *A guide to language testing: Development, evaluation, research*. Cambridge, MA: Newberry House Publishers.
- Hong Kong Examination and Assessment Authority. (2012). *2012 HKDSE English language examination: Introduction to school-based assessment*. Retrieved from http://www.hkeaa.edu.hk/DocLibrary/SBA/HKDSE/Eng_DVD/index.html

- Hu, G. (2002). Recent important developments in secondary English-language teaching in the People's Republic of China. *Language, Culture and Curriculum*, 15(1), 30–49.
<http://doi.org/10.1080/07908310208666631>
- Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Hyland, K. (2010). Research writing. In B. Paltridge & A. Phakiti (Eds.), *Continuum companion to research methods in applied linguistics* (pp. 191–204). London: Continuum.
- Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269–293). Harmondsworth: Penguin.
- Ingram, E. (1977). Basic concepts in testing. In J. P. B. Allen & A. Davies (Eds.), *Testing and experimental methods*. Oxford: Oxford University Press.
- Ito, A. (2005). A validation study on the English language test in a Japanese Nationwide University Entrance Examination. *Asian EFL Journal*, 7(2), 1–26.
- Jamieson, J. M., Eignor, D., & Kunnan, A. J. (2008). Frameworks for a new TOEFL. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 55–95). New York, NY: Routledge.
- Jin, Y. (2010). The National College English Testing committee. In L. Cheng & A. Curtis (Eds.), *English language assessment and the Chinese learner* (pp. 44–59). New York, NY; London: Routledge.
- Jin, Y., & Fan, J. (2011). Test for English Majors (TEM) in China. *Language Testing*, 28(4), 589–596. <http://doi.org/http://dx.doi.org/10.1177/0265532211414852>
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527–535. <http://doi.org/http://dx.doi.org/10.1037/0033-2909.112.3.527>
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38(4), 319–342. <http://doi.org/10.1111/j.1745-3984.2001.tb01130.x>
- Kane, M. T. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practice*, 21(1), 31–41. <http://doi.org/10.1111/j.1745-3992.2002.tb00083.x>
- Kane, M. T. (2006). Validation. In *Educational measurement* (4th ed., pp. 17–64). Westport, CT: American Council on Education and Praeger.

- Kane, M. T. (2012). Validating score interpretations and uses. *Language Testing*, 29(1), 3–17.
<http://doi.org/10.1177/0265532211417210>
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73. <http://doi.org/10.1111/jedm.12000>
- Kane, M. T., Crooks, T., & Cohen, A. (1999). Validating Measures of Performance. *Educational Measurement: Issues and Practice*, 18(2), 5–17. <http://doi.org/10.1111/j.1745-3992.1999.tb00010.x>
- Kasper, G. (1998). Analysing verbal protocols. *TESOL Quarterly*, 32(2), 358–362.
<http://doi.org/10.2307/3587591>
- Kramsch, C. (1986). From language proficiency to interactional competence. *The Modern Language Journal*, 70(4), 366–372. <http://doi.org/10.1111/j.1540-4781.1986.tb05291.x>
- Krippendorff, K. (Ed.). (2013). *Content analysis: An introduction to its methodology* (3rd ed). Thousand Oaks, CA: SAGE Publications.
- Kunnan, A. J. (2004). Test fairness. In M. Milanovic, C. Weir, & S. Bolton (Eds.), *Europe language testing in a global context: Selected papers from the ALTE conference in Barcelona* (pp. 27–48). Cambridge: Cambridge University Press.
- Kuusela, H., & Paul, P. (2000). A comparison of concurrent and retrospective verbal protocol analysis. *The American Journal of Psychology*, 113(3), 387–404.
<http://doi.org/10.2307/1423365>
- Lado, R. (1961). *Language testing: The construction and use of foreign language tests; a teacher's book*. London: Longman.
- Lam, H. P. (1994). Methodology washback – an insider's view. In D. Nunan, R. Berry & V. Berry (Eds.), *Bringing About Change in Language Education* (pp.83-99). Hong Kong: The University of Hong Kong.
- Li, M. (2012). English curriculum in higher education in China for non-English majors. In J. Ruan & C. B. Leung (Eds.), *Perspectives on teaching and learning English literacy in China* (pp. 105–114). Dordrecht, MA: Springer.

- Linn, R. L. (1983). Testing and instruction: Links and distinctions. *Journal of Educational Measurement*, 20(2), 179–189. <http://doi.org/10.1111/j.1745-3984.1983.tb00198.x>
- Litwin, M. S. (1995). *How to measure survey reliability and validity*. Thousand Oaks, CA: SAGE Publications.
- Liu, H., & Wu, Q. (2006). Consequences of college entrance exams in China and the reform challenges. *KEDI Journal of Educational Policy*, 3(1), 7–21.
- Liu, Q. (2010). The National Education Examinations Authority and its English language tests. In L. Cheng & A. Curtis (Eds.), *English language assessment and the Chinese learner* (pp. 29–43). New York, NY; London: Routledge.
- Llurda, E. (2000). On competence, proficiency, and communicative language ability. *International Journal of Applied Linguistics*, 10(1), 85–96. <http://doi.org/10.1111/j.1473-4192.2000.tb00141.x>
- Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.
- Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum.
- Madaus, G. F. (1988). The influences of testing on the curriculum. In L. Travers (Ed.), *Critical issues in curriculum* (pp. 83–121). Chicago, IL: University of Chicago Press.
- Madsen, H. S. (1983). *Techniques in testing*. New York, NY: Oxford University Press.
- McCarthy, M. (1991). *Discourse analysis for language teachers*. Cambridge: Cambridge University Press.
- McDonough, J., & McDonough, S. H. (1997). *Research methods for English language teachers*. London: Arnold.
- McNamara, T. F. (1996). *Measuring second language performance: A new era in language testing*. New York, NY: Longman.
- McNamara, T. F. (2000). *Language testing*. Oxford: Oxford University Press.
- Messick, S. (1989). Validity. *ETS Research Report Series*, 1989(2), i–208. <http://doi.org/10.1002/j.2330-8516.1987.tb00244.x>
- Messick, S. (1990). *Validity of test interpretation and use*. Educational Testing Service.

- Messick, S. (1996). Validity and washback in language testing. *ETS Research Report Series*, 1996(1), i–18. <http://doi.org/10.1002/j.2333-8504.1996.tb01695.x>
- Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research*, 45(1-3), 35–44. <http://doi.org/10.1023/A:1006964925094>
- Meyers, J., Lytle, S., Palladino, D., Devenpeck, G., & Green, M. (1990). Think-aloud protocol analysis: An investigation of reading comprehension strategies in fourth- and fifth-grade students. *Journal of Psychoeducational Assessment*, 8(2), 112–127. <http://doi.org/10.1177/073428299000800201>
- Mickan, P., Slater, S., & Gibson, C. (2000). A study of response validity of the IELTS writing module. In http://www.ielts.org/researchers/research/volume_3.aspx. IELTS Australia Pty Limited. Retrieved from <https://digital.library.adelaide.edu.au/dspace/handle/2440/37029>
- Ministry of Education, Singapore. (2010). *English Language Syllabus 2010 Primary & Secondary (Express / Normal [Academic])*. Retrieved from <http://www.moe.gov.sg/education/syllabuses/english-language-and-literature/files/english-primary-secondary-express-normal-academic.pdf>
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). Design and analysis in task-based language assessment. *Language Testing*, 19(4), 477–496. <http://doi.org/10.1191/0265532202lt241oa>
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). Focus article: On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1(1), 3–62. http://doi.org/10.1207/S15366359MEA0101_02
- Morrow, K. (1977). *Techniques of evaluation for a notional syllabus*. London: Royal Society of Arts.
- Morrow, K. (1979). Communicative language testing: Revolution of evolution? In C. K. Brumfit & K. Johnson (Eds.), *The communicative approach to language teaching* (pp. 143–157). Oxford: Oxford University Press.
- Morrow, K. (1986). The evaluation of tests of communicative performance. In M. Portal (Ed.), *Innovations in language testing* (pp. 1–13). Windsor: NFER/Nelson.

- Morin, A. (2015). *Four writing styles kids will need to know*. Retrieved from <http://childparenting.about.com/od/schoollearning/a/4-types-of-writing-styles.htm>
- Munby, J. (1978). *Communicative syllabus design: A sociolinguistic model for defining the content of purpose-specific language programmes*. Cambridge: Cambridge University Press.
- Norris, S. P. (1991). Informal reasoning assessment: Using verbal reports of thinking to improve multiple-choice test validity. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), *Informal reasoning and education*. Hillsdale, NJ: Lawrence Erlbaum.
- Nunan, D. (1992). *Research methods in language learning*. Cambridge: Cambridge University Press.
- Nunan, D. (1993). *Introducing discourse analysis*. London: Penguin.
- OECD. (2010). *Shanghai and Hong Kong: Two distinct examples of education reform in China*. Retrieved from <http://www.oecd.org/dataoecd/34/45/46581016.pdf>
- Oller, J. W. (1979). *Language tests at school: A pragmatic approach*. London: Longman.
- O'Sullivan, B., & Weir, C. J. (2011). Test development and validation. In B. O'Sullivan (Ed.), *Language testing: theories and practices* (pp. 13–32). Basingstoke: Palgrave Macmillan.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed). Newbury Park, CA: SAGE Publications.
- Peason, I. (1988). Tests as levers for change. In D. Chamberlain & R. Baumgardner (Eds.), *ESP in the classroom: Practice and evaluation, ELT Documents* (pp. 98–107). London: Modern English Publications.
- Phakiti, A. (2010). Analysing quantitative data. In B. Paltridge & A. Phakiti (Eds.), *Continuum companion to research methods in applied linguistics* (pp. 39–49). London: Continuum.
- Popham, W. J. (1987). The merits of measurement-driven instruction. *The Phi Delta Kappan*, 68(9), 679–682.

- Popham, W. J. (1997). Consequential validity: Right concern-wrong concept. *Educational Measurement: Issues and Practice*, 16(2), 9–13. <http://doi.org/10.1111/j.1745-3992.1997.tb00586.x>
- Popham, W. J. (2013). *Classroom assessment: What teachers need to know* (7th ed.). Boston, NJ: Pearson.
- Purpura, J. E. (2004). *Assessing grammar*. Cambridge: Cambridge University Press.
- Qi, L. (2004). Has a high-stakes test produced the intended changes? In L. Cheng, J. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 171–190). Mahwah, NJ: Lawrence Erlbaum.
- Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing*, 22(2), 142–173.
<http://doi.org/10.1191/0265532205lt300oa>
- 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.
<http://doi.org/10.1080/09695940701272856>
- Qi, L. (2010). Should proofreading go? Examining the selection function and washback of the proofreading sub-test in the National Matriculation English Test. In L. Cheng & A. Curtis (Eds.), *English language assessment and the Chinese learner* (pp. 219–233). New York, NY; London: Routledge.
- Read, J. A. S. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Reckase, M. D. (1998). Consequential validity from the test developer's perspective. *Educational Measurement: Issues and Practice*, 17(2), 13–16.
<http://doi.org/10.1111/j.1745-3992.1998.tb00827.x>
- Roever, C. (2011). Testing of second language pragmatics: Past and future. *Language Testing*, 28(4), 463–481. <http://doi.org/10.1177/0265532210394633>
- Savignon, S. J. (1983). *Communicative competence: Theory and classroom practice: Texts and contexts in second language learning*. Reading, MA: Addison-Wesley.

- Shanghai Curriculum and Teaching Materials Reform Commission. (1998a). *全日制义务教育英语学科课程标准 (修订本)* [English curriculum for full-time nine-year compulsory education (revised version)]. Shanghai: Shanghai Education Press.
- Shanghai Curriculum and Teaching Materials Reform Commission. (1998b). *全日制高级中学英语学科课程标准 (修订本)* [English curriculum for full-time senior secondary schools (revised version)]. Shanghai: Shanghai Education Press.
- Shanghai Municipal Education Commission. (2004). *上海市中小学英语课程标准 (征求意见稿)* [English curriculum standards for Shanghai Municipal secondary and primary schools (draft version)]. Shanghai: Shanghai Educational Publishing House.
- Shanghai Municipal Education Commission. (2005). *上海市高级中学英语学科教学基本要求 (试验本)* [English language teaching basic requirement for Shanghai Municipal senior secondary schools (trial version)]. Shanghai: Shanghai Educational Publishing House.
- Shanghai Municipal Education Examination Centre. (2014a). *高考试题分析与评价 2014* [Analysis and evaluation of the university entrance examination 2014]. Shanghai: Shanghai Education Press.
- Shanghai Municipal Education Examination Centre. (2014b). *上海卷考试手册 2014* [Shanghai examination handbook 2014]. Shanghai: Shanghai Guxi Press.
- Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20(5), 8–11. <http://doi.org/10.3102/0013189X020005008>
- Spratt, M. (2005). Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, 9(1), 5–29. <http://doi.org/10.1191/1362168805lr152oa>
- Sudworth, J. (2012). *China's students take on tough Gaokao university entrance exam*. Retrieved from <http://www.economist.com/blogs/johnson/2011/09/english-china>
- Tan, C. (2012). The culture of education policy making: Curriculum reform in Shanghai. *Critical Studies in Education*, 53(2), 153–167. <http://doi.org/10.1080/17508487.2012.672333>
- Toulmin, S. E. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- Toulmin, S. E. (2003). *The uses of argument* (2nd ed.). Cambridge: Cambridge University Press.

- van Dijk, T. A. (1977). *Text and content: Explorations in the semantics and pragmatics of discourse*. London: Longman.
- van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think aloud method: A practical guide to modelling cognitive process*. London: Academic Press.
- Vollmer, H. J., & Sang, F. (1983). Competing hypotheses about second language ability: A plea for caution. In J. W. Oller (Ed.), *Issues in language testing research* (pp. 29–79). Rowley, MA: Newbury House.
- Wagner, E. (2010). Survey research. In B. Paltridge & A. Phakiti (Eds.), *Continuum companion to research methods in applied linguistics* (pp. 22–38). London: Continuum.
- Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan Impact Study. *Language Testing*, 10(1), 41–69. <http://doi.org/10.1177/026553229301000103>
- Wall, D., Clapham, C., & Alderson, J. C. (1994). Evaluating a placement test. *Language Testing*, 11(3), 321–344. <http://doi.org/10.1177/026553229401100305>
- Wang, H., Choi, I., Schmidgall, J., & Bachman, L. F. (2012). Review of Pearson Test of English Academic: Building an assessment use argument. *Language Testing*, 29(4), 603–619. <http://doi.org/10.1177/0265532212448619>
- Wang, Q. (2005). Primary English teachers' perceptions of learning-centre approach in English language teaching in China: A preliminary survey study. In Q. Wang (Ed.), *English teacher education in the 21st century: Research and development* (pp. 31–64). Beijing: Beijing Normal University Press.
- Wang, Q. (2007). The national curriculum changes and their effects on English language teaching in the People's Republic of China. In J. Cummins & C. Davison (Eds.), *International Handbook of English Language Teaching* (pp. 87–105). Norwell, MA: Springer.
- Wang, Q., & Chen, Z. (2012). Twenty-first century senior high school English curriculum reform in China. In J. Ruan & C. B. Leung (Eds.), *Perspectives on teaching and learning English literacy in China* (pp. 85–103). Dordrecht, MA: Springer.

- Wang, W., & Lam, A. S. L. (2009). The English language curriculum for senior secondary school in China its evolution from 1949. *RELC Journal*, 40(1), 65–82.
<http://doi.org/10.1177/0033688208101447>
- Wang, Z. (2015). *Chinese entrance exam needs to be recognised globally*. Retrieved from <http://www.telegraph.co.uk/sponsored/china-watch/society/11654256/chinese-entrance-exam-recognised-globally.html>
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Weir, C. J. (1990). *Communicative language testing*. New York, NY: Prentice Hall.
- Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Basingstoke: Palgrave Macmillan.
- Widdowson, H. G. (1978). *Teaching language as communication*. Oxford: Oxford University Press.
- Widdowson, H. G. (1983). *Learning purpose and language use*. Oxford: Oxford University Press.
- Xiao, Y., Sharpling, G., & Liu, H. (2011). Washback of National Matriculation English Test on students' learning in the Chinese secondary school context. *The Asian EFL Journal Quarterly*, 3(13), 103–129.
- Yang, H. Z. (2000). *Validation study of the College English Test*. Paper presented at the International Association of Applied Linguistics (AILA), Waseda, Japan.
- Yang, H. Z., & Weir, C. J. (1998). *The CET validation study*. Shanghai: Shanghai Foreign Language Education Press.
- Zeng, Y. (2010). The Computerized Oral English Test of the National Matriculation English Test. In L. Cheng & A. Curtis (Eds.), *English language assessment and the Chinese learner* (pp. 234–247). New York, NY; London: Routledge.
- Zou, S., Zhang, Y., & Zhou, Y. (2002). 阅读测试中题目类型、策略与分数的关系 – TEM4 考试阅读项目的答题效度研究 [The relationship between item types, reading strategies and test performance – a study of the response validity of TEM4 reading]. *Foreign Languages and Their Teaching*, 158(5), 19–22.

Appendices

Appendix A: Human Ethics Approval Letter



Phone 0-4-463 5676
Fax 0-4-463 5209
Email Allison.kirkman@vuw.ac.nz

MEMORANDUM

TO	Matthew Book
COPY TO	Peter Gu Averil Coxhead
FROM	Dr Allison Kirkman, Convener, Human Ethics Committee
DATE	10 August 2013
PAGES	1
SUBJECT	Ethics Approval: 19850 Validating the English Language Exam for university entry in Shanghai, China

Thank you for your application for ethical approval, which has now been considered by the Standing Committee of the Human Ethics Committee.

Your application has been approved from the above date and this approval continues until 30 August 2015. If your data collection is not completed by this date you should apply to the Human Ethics Committee for an extension to this approval.

Best wishes with the research.

Allison Kirkman
Human Ethics Committee

A handwritten signature in blue ink, appearing to read 'Allison Kirkman'.

Appendix B: Participants Information Sheets and Consent Forms



Teacher Information Sheet

Research project: Validating the English Language Exam for university entry in Shanghai, China
Researcher: Matthew Book, School of Linguistics and Applied Language Studies,
Victoria University of Wellington

Hello,

My name is Matthew Book, and I am a PhD student at the School of Linguistics and Applied Language Studies, Victoria University of Wellington, New Zealand. The topic of my research project is, "validating the English Language Exam for university entry in Shanghai, China". This study will be carried out *from mid-February to mid-June 2014*.

In this study I would like to observe your classes – one unit teaching (about 2 weeks). After class, I would like to interview you about your teaching – why you teach particular materials in class. The purpose of the study is to determine how serious the impact of the Exam is on teaching, and to rate the practicality of the Exam. The interviews will be in English and about 10 minutes each time. The interviews will be arranged at a convenient time for you in school.

In order to protect the anonymity of the participants and the school, the study will be reported using pseudonyms. Information collected through the interviews will be kept confidential. I will keep recorded data in a locked file on my computer. And paper data will be kept in a locker in my office. Only my supervisors and I will have the access to the data. The data will be kept for about two years after the completion of the PhD program and will then be destroyed. The data will be analysed and used for this research project and potential future publications only.

Human ethics approval of Victoria University of Wellington has been granted for this project. Participation in this research is entirely voluntary and you are free to withdraw from the study, including any information that you have provided, any time **before 1 March 2014**.

There will be no employment implications. The class observations and interviews are not related to your school teaching. Participation of this study (class observations and interviews) will not influence your teaching at school.

If you have any questions or would like to receive further information about this research, you can contact me or my supervisor Peter Gu via the following ways:

Matthew Book Room 306, Von Zedlitz Building Victoria University of Wellington, Wellington 6140, New Zealand Email: matthew.book@vuw.ac.nz Phone number in Shanghai: +86 13052293632	Dr Peter Gu Senior Lecturer Room 210, Von Zedlitz Building Victoria University of Wellington, Wellington 6140, New Zealand Email: peter.gu@vuw.ac.nz
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Thank you for your time.
Matthew



Students Information Sheet

Research project: The English Language Exam for university entry in Shanghai, China

Researcher: Matthew Book, School of Linguistics and Applied Language Studies,
Victoria University of Wellington

Hello,

My name is Matthew Book, and I am a PhD student at School of Linguistics and Applied Language Studies, Victoria University of Wellington, New Zealand. The topic of my research project is, "The English Language Exam for university entry in Shanghai, China". This study will be carried out *from mid-February to mid-June 2014*. Participation in this study will **NOT** influence your grades at school.

The study will include mock-trials, think-aloud tasks, and class observations & interviews.

First of all, I will invite all of you to participate in Mock-trials. You will take three of the past English Language Exam papers and a sample IELTS test over four months, each test will be taken in your school and will be about 2 hours long. The total time will be about 8 hours.

Second, **SIX** of you will be invited to participate in Think-aloud Tasks – you will be asked to tell me how you answer each task. The test will last for about 2 hours and you will take the test separately and sit in front of me. The test will be arranged at a convenient time for you in school.

Finally, I would like to observe your English classes (about 2 weeks). After class, I would like to interview **some** of you about what you will do after the class. The purpose of the study is to determine how serious the impact of the Exam is on learning, and to rate the practicality of the Exam. The interviews will be in Chinese and about 10 minutes each time. The interviews will be arranged at a convenient time for you in school.

In order to protect the anonymity of the participants and the school, the study will be reported using pseudonyms. Information collected through the interviews will be kept confidential. Only my supervisors and I will have the access to the data. The data will be kept for about two years after the completion of the PhD program and will then be destroyed. The data will be analysed and used for this research project and potential future publications only.

Human ethics approval of Victoria University of Wellington has been granted for this project. Participation in this research is entirely voluntary and you are free to withdraw from the study, including any information that you have provided, any time **before 1 March 2014**.

In return for your time, I will offer you free oral English lessons if your school approves. Details of the lessons will be discussed and arranged with your school.

If you have any questions or would like to receive further information about this research, you can contact me or my supervisor Peter Gu via the following ways:

Matthew Book Room 306, Von Zedlitz Building Victoria University of Wellington, Wellington 6140, New Zealand Email: matthew.book@vuw.ac.nz Phone number in Shanghai: +86 13052293632	Dr Peter Gu Senior Lecturer Room 210, Von Zedlitz Building Victoria University of Wellington, Wellington 6140, New Zealand Email: peter.gu@vuw.ac.nz
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Thank you for your time.
Matthew



Consent to participation (teacher) in the “validation of the English Language Exam for university entry in Shanghai, China” study

If you agree with the following statements, please sign your name at the bottom.

If you have any questions, please re-read the information sheet or consult with Matthew.

Participation in this research is entirely voluntary and you are free to withdraw from the study, including any information that you have provided, **any time before 1 March 2014**.

I have been provided with adequate information relating to the nature and objectives of this research project, and I have understood that information. I have been given the opportunity to seek further clarification or explanations.

I understand the information obtained in this study will be used for this research project and in academic papers on an anonymous basis.

I understand that there will be no employment implications, and the class observations and interviews are not related to my school teaching. Participation in this study (class observations and interviews) will not influence my teaching at school.

I understand that the information I have provided will not be used for any other purposes or released to others without my written consent.

I agree to participate in the class observations and interviews.

I would like to have a summary of my interviews returned to me at the conclusion of the project by email, my email address is _____.

I understand that the information obtained in the study will be destroyed about two years after the completion of the research project.

Name: _____

Signed: _____

Date: _____



Consent to participation (students) in the “English Language Exam for university entry in Shanghai, China” study

If you agree with the following statements, please sign your name at the bottom.

If you have any questions, please re-read the information sheet or consult with Matthew.

Participation in this research is entirely voluntary and you are free to withdraw from the study, including any information that you have provided, **any time before 1 March 2014**.

I have been provided with adequate information relating to the nature and objectives of this research project, and I have understood that information. I have been given the opportunity to seek further clarification or explanations.

I understand the information obtained in this study will be used for this research project and in academic papers on an anonymous basis.

I understand that this study is not related to my school studies. Participation in this study (the mock-trials, think-aloud task and interview) will **NOT** influence my grades at school.

I understand that the information I have provided will not be used for any other purposes or released to others without my written consent.

I agree to participate in this mock-trial, think-aloud task, and interview when I am invited.

I would like to have a summary of my interview returned to me at the conclusion of the project by email, my email address is _____.

I understand that the information obtained in the study will be destroyed about two years after the completion of the research project.

*Please note that if your age is below 16, please ask your parent or caregiver to sign the consent form for you.

Name: _____

Signed: _____

Signed: _____

(By your parent or caregiver if you are below 16.)

Date: _____

2008 年全国普通高等学校招生统一考试

上海 英语试卷

考生注意:

1. 考试时间 120 分钟, 试卷满分 150 分。
2. 本考试设试卷和答题纸两部分。试卷分为第 I 卷(第 1-11 页)和第 II 卷(第 12 页), 全卷共 11 页。所有答题必须涂(选择题)或写(非选择题)在答题纸上, 做在试卷上一律不得分。
3. 答题前, 务必在答题纸上填写准考证号和姓名, 并将核对后的条形码贴在指定位置上。

第 I 卷 (共 105 分)

I. Listening comprehension

Section A Short Conversations

Directions: In Section A, you will hear ten short conversations between two speakers. At the end of each conversation, a question will be asked about what was said. The conversations and the questions will be spoken only once. After you hear a conversation and the question about it, read the four possible answers on your paper, and decide which one is the best answer to the question you have heard.

1. A. This afternoon. B. Tomorrow. C. Next week. D. Next month.
2. A. She doesn't play tennis well. B. She likes other sports as well.
C. She is an enthusiastic tennis player. D. She is a professional athlete.
3. A. At a paint store. B. At an oil market.
C. At a science museum. D. At a gallery.
4. A. Work in the yard. B. Buy some wood.
C. Go to the bookstore. D. Take a walk.
5. A. A taxi driver. B. A passenger.
C. A car cleaner. D. A mechanic.
6. A. Call a repairman. B. Get out the paper stuck
C. Turn to her colleague for help. D. Restart the machine
7. A. There are not enough gardens. B. Parking areas are full before 10:00.
C. Parking areas are closed after 10:00. D. All classes begin at 10:00.
8. A. The presentation will begin at noon.
B. She'll present her work to the man.
C. She'd like to invite the man for lunch.
D. She suggests working on the presentation at 12:00.
9. A. The dormitory hours. B. The problem with the rules.
C. The door number of the dormitory. D. The time to open the dormitory.
10. A. The chairs didn't need to be painted.
B. He doesn't like the color of the chairs.
C. The park could have avoided the problem.
D. The woman should have been more careful.

Section B Passages

Directions: In Section B, you will hear two short passages, and you will be asked three questions on each of the passages. The passages will be read twice, but the questions will be spoken only once. When you hear a question, read the four possible answers on your paper and decide which one would be the best answer to the question you have heard.

Questions 11 through 13 are based on the following passage.

11. A. Worried. B. Surprised. C. Satisfied. D. Uninterested.
12. A. It spoiled Juana's reputation. B. It copied her ideas without permission.
 C. It bought Juana's dishwashers. D. It wanted to share the dishwasher market.
13. A. A successful business case. B. Juana's waterless laundry.
 C. A case against a global company. D. The worldwide dishwasher market.

Questions 14 through 16 are based on the following instructions.

14. A. Footprints. B. Food. C. Living insects. D. Orange seeds.
15. A. Don't touch animals under any circumstances.
 B. Don't take away any natural objects from the park.
 C. Don't leave litter in the park or throw any off the boat.
 D. Don't transport animals from one island to another.
16. A. To protect the guide's interest. B. To improve the unique environment.
 C. To ensure a trouble-free visit. D. To get rid of illegal behaviours.

Section C Longer Conversations

Directions: In Section C, you will hear two longer conversations. The conversations will be read twice. After you hear each conversation, you are required to fill in the numbered blanks with the information you have heard. Write your answers on your answer sheet.

Blanks 17 through 20 are based on the following conversation.

Complete the form. Write **ONE WORD** for each answer.

Customs Form	
Destination:	<u>17</u>
Contents in detail:	<u>18</u>
Value:	<u>19</u> dollars
Type of mail:	<u>20</u>
Weight:	1.5 pounds
Your item must not contain any dangerous articles prohibited by postal regulations.	

Blanks 21 through 24 are based on the following conversation.

Complete the form. Write **NO MORE THAN THREE WORDS** for each answer.

What is the man's oral report on?	An <u>21</u> woman.
What kinds of people are called refugees?	Those who <u>22</u> to escape war or other problems.
How many refugees are there worldwide?	About <u>23</u> .
What does the man think of Ms. Ogata?	She is successful both in <u>24</u> .

II. Grammar and Vocabulary

Section A

Directions: Beneath each of the following sentences there are four choices marked A, B, C and D. Choose the one answer that best completes the sentence.

25. The two sportsmen congratulated each other ____ winning the match by shaking hands.
A. with B. on C. in D. to
26. X: Do you want tea or coffee?
Y: ____. I really don't mind.
A. None B. Neither C. Either D. All
27. In my view, London's not as expensive in price as Tokyo but Tokyo is ____ in traffic.
A. the most organized B. more organized C. so organized as D. as organized as
28. X: Do you know if Terry will go camping this weekend?
Y: Terry? Never! She ____ tents and fresh air!
A. has hated B. hated C. will hate D. hates
29. According to the air traffic rules, you ____ switch off your mobile phone before boarding.
A. may B. can C. would D. should
30. My sister, an inexperienced rider, was found sitting on the bicycle ____ to balance it.
A. having tried B. trying C. to try D. tried
31. X: Are you ready for Spain?
Y: Yes. I want the girls to experience that ____ they are young.
A. while B. until C. if D. before
32. In recent years many football clubs ____ as business to make a profit.
A. have run B. have been run C. had been run D. will run
33. If there's a lot of work ____, I'm happy to just keep on until it is finished.
A. to do B. to be doing C. done D. doing
34. As his best friend, I can make accurate guesses about ____ he will do or think.
A. what B. which C. whom D. that
35. Something as simple as ____ some cold water may clear your mind and relieve pressure.
A. to drink B. drinking C. to be drinking D. drunk
36. It has been proved ____ eating vegetables in childhood helps to protect you against serious illnesses in later life.
A. if B. because C. when D. that
37. Ideally ____ for Broadway theatres and Fifth Avenue, the *York Park* hotel is a favorite with many guests.
A. locating B. being located C. having been located D. located
38. We went through a period ____ communications were very difficult in the rural areas.
A. which B. whose C. in which D. with which
39. So much of interest ____ that most visitors simply run out of time before seeing it all.
A. offers Beijing B. Beijing offers C. does Beijing offer D. Beijing does offer
40. ____ well prepared you are, you still need a lot of luck in mountain climbing.
A. However B. Whatever C. No matter D. Although

Section B

Directions: Complete the following passage by using the words in the box. Each word can only be used once. Note that there is one word more than you need.

A. honoured	B. set	C. historic	D. secretly	E. citizen
F. granted	G. route	H. briefly	I. restoration	J. leading

Frederick Douglass was an escaped slave in the movement that fought to end slavery in the United States. He became a 41 voice in the year before the Civil War.

A few weeks ago, the National Park Service (NPS) 42 Douglass's birth and Black History Month with the reopening of his home at Cedar Hill, a 43 site in Washington, D.C. The two-story house, which contains many of Douglass's personal possessions, had undergone a three-year 44. (Thanks to the NTS website, however, you don't have to live in the nation's capital to visit it. Take a tour online.)

He was born Frederick Augustus Washington Bailey to a slave mother and a white father he never knew. Douglass grew up to become the first black 45 to hold a government office - as US minister and *consul general* (总领事) to Haiti.

As a youth, he never gone to school. Educating slaves was illegal in the South, so he 46 taught himself to read and write. At 21 years old, he escaped from his slave owner to Massachusetts and changed his last name to Douglass, to hide his identity.

In the 1850s, Douglass was involved with the Underground Railroad, the system 47 up by antislavery groups to bring runaway slaves to the North and Canada. His home in Rochester, N.Y. was near the Canadian border. It became an important station on the 48, housing as many as 11 runaway slaves at a time.

He died in 1895. In his lifetime, Douglass witnessed the end of slavery in 1865 and the adoption of the 15th *Amendment to the US Constitution* (美国宪法修正案), which 49 African-Americans the right to vote.

III. Reading Comprehension

Section A

Directions: For each blank in the following passage there are four words or phrases marked A, B, C and D. Fill in each blank with the word or phrase that best fits the context.

People think children should play sports. Sports are fun, and playing with others. However, playing sports can have 50 effects on children. It may produce feelings of poor self-respect or aggressive behavior in some children. According to research on kids and sports, 40,000,000 kids play sports in the US. Of these, 18,000,000 say they have been 51 at or called names while playing sports. This leaves many children with a bad 52 of sports. They think sports are just too aggressive.

Many researchers believe adults, especially parents and coaches, are the main 53 of too much aggression in children's sports. They believe children 54 aggressive adult behavior. This behavior is then further strengthened through both positive and negative feedback. Parents and coaches are powerful teachers because children usually look up to them. Often these adults behave aggressively themselves, sending children the message that 55 is everything. Many parents go to children's sporting events and shout 56 at other players or cheer when their child behaves 57. As well, children are even taught that hurting other players is 58 or are pushed to continue playing even when they are injured. 59, the media makes violence seem exciting. Children watch adult sports games and see violent behavior replayed over and over on television.

As a society, we really need to 60 this problem and do something about it. Parents and coaches 61 should act as better examples for children. They also need to teach children better 62. They should not just cheer when children win or act aggressively. They should teach children to 63 themselves whether they win or not. Besides, children should not be allowed to continue to play when they are injured. If adults allow children to play when injured, this gives the message that 64 is not as important as winning.

- | | | | |
|----------------------|----------------|-----------------|-----------------|
| 50. A. restrictive | B. negative | C. active | D. instructive |
| 51. A. knocked | B. glanced | C. smiled | D. shouted |
| 52. A. impression | B. concept | C. taste | D. expectation |
| 53. A. resource | B. cause | C. course | D. consequence |
| 54. A. question | B. understand | C. copy | D. neglect |
| 55. A. winning | B. practising | C. fun | D. sport |
| 56. A. praises | B. orders | C. remarks | D. insults |
| 57. A. proudly | B. ambitiously | C. aggressively | D. bravely |
| 58. A. acceptable | B. impolite | C. possible | D. accessible |
| 59. A. By contrast | B. In addition | C. As a result | D. After all |
| 60. A. look up to | B. face up to | C. make up for | D. come up with |
| 61. A. in particular | B. in all | C. in return | D. in advance |
| 62. A. techniques | B. means | C. values | D. directions |
| 63. A. respect | B. relax | C. forgive | D. enjoy |
| 64. A. body | B. fame | C. health | D. spirit |

Section B

Directions: Read the following four passages. Each passage is followed by several questions or unfinished statements. For each of them there are four choices marked A, B, C and D. Choose the one that fits best according to the information given in the passage you have just read.

(A)

ASK LASKAS - YOU'VE GOT QUESTIONS. SHE'S GOT ANSWERS

Q My children go to a primary school where they are not allowed to play football in the playground for fear that a child might be hurt. Besides, now the school says there must be no homework because the local secondary school can't keep up with the amount of homework given in the primary school. Can the school do this?

A It can't if enough parents do something about it. It is not just schools. We live in a society which wishes to get rid of risk. However, schools should have a little common sense and courage. Children need risk if they are to grow up self-sufficient and confident. They need homework, too, if they are to fulfill their academic potential. Complain, loudly.

Q I have a beautiful teenage daughter who spends an hour making up her face in front of the mirror every day. I tell her to go easy. She just gets mad or bursts into tears. How can I make her understand she's beautiful the way she is.

A You can't. Your daughter is at the age when she's trying to look beautiful, trying on new masks. And if her friends all dress up as she does, you're in for an extra hard time. Support her and tell her she's beautiful - even if she looks ridiculous for now. Then invite her to join you for a day at a spa (健康美容中心). Let her try various looks until she's comfortable in her own skin.

65. Why are the children not allowed to play football in the playground?

- A. The school is afraid that children might be injured.
- B. The school is not sensible and confident.
- C. The children don't have enough time to do homework.
- D. The children may fail to fulfill their academic potential.

66. What disturbs Plain Mom is that her daughter_____.

- A. becomes mad
- B. cries a lot
- C. spends much time before the mirror
- D. is not beautiful enough

67. The solution to Plain Mom's problem is to_____.

- A. make her daughter look less ridiculous
- B. let her daughter dress up like her friends
- C. make her daughter go to a spa every week
- D. let her daughter feel herself what beauty is

(B)

Zoe Chambers was a successful PR (Public Relations) consultant and life was going well - she had a great job, a beautiful flat and a busy social life in London. Then one evening in June last year, she received a text message telling her she was out of work. The first two weeks were the most difficult to live through," she said. "After everything I'd done for the company, they dismissed me by text! I was so angry and I just didn't feel like looking for another job. I hated everything about the city and my life."

Then, Zoe received an invitation from an old school friend, Kathy, to come and stay. Kathy and her husband, Haw, had just bought a farm in north-west Wales. Zoe jumped at the chance to spend a weekend away from London, and now, ten months later, she is still on the farm.

"The moment I arrived at Kathy's farm, I loved it and I knew I wanted to stay," said Zoe. "Everything about my past life suddenly seemed meaningless."

Zoe has been working on the farm since October of last year and says she has no regrets. "It's a hard life, physically very tiring," she says. "In London I was stressed and often mentally exhausted. But this is a good, healthy tiredness. Here, all I need to put me in a good mood is a hot bath and one of Kathy's wonderful dinners."

Zoe says she has never felt bored on the farm. Every day brings a new experience. Kathy has been teaching her how to ride a horse and she has learnt to drive a tractor. Since Christmas, she has been helping with the lambing - watching a lamb being born is unbelievable, she says, "It's one of the most moving experiences I've ever had. I could never go back to city life now."

68. When working as a PR consultant in London, Zoe thought she lived a _____ life.
A. satisfying B. tough C. meaningless D. boring
69. The most important reason why Zoe went to visit Kathy's farm is that _____.
A. Zoe lost her job as a PR consultant B. Kathy persuaded her to do so
C. Zoe got tired of the city life D. Zoe loved Wales more than London
70. How does Zoe feel about the country life according to the passage?
A. Tiresome and troublesome. B. Romantic and peaceful
C. Mentally exhausting but healthy D. Physically tiring but rewarding.
71. Which of the following is closest to the main idea of the passage?
A. A friend in need is a friend indeed. B. Where there is a will, there is a way.
C. A misfortune may turn out a blessing. D. Kill two birds with one stone.

(c)

A study involving 8,500 teenagers from all social backgrounds found that most of them are ignorant when it comes to money. The findings, the first in a series of reports from NatWest that has started a five-year research project into teenagers and money, are particularly worrying as this generation of young people is likely to be burdened with greater debts than any before.

University *tuition fees* (学费) are currently capped at £3,000 annually, but this will be reviewed next year and the Government is under enormous pressure **to raise the ceiling**.

In the research, the teenagers were presented with the terms of four different loans but 76 per cent failed to identify the cheapest. The young people also predicted that they would be earning on average £ 31,000 by the age of 25, although the average salary for those aged 22 to 29 is just £ 17,815. The teenagers expected to be in debt when they finished university or training, although half said that they assumed the debts would be less than £ 10,000. Average debts for graduates are £ 12,363.

Stephen Moir, head of community investment at the Royal Bank of Scotland Group which owns NatWest, said. "The more exposed young people are to financial issues, and the younger they become aware of them, the more likely they are to become responsible, forward-planning adults who manage their finances confidently and effectively."

Ministers are deeply concerned about the financial pressures on teenagers and young people because of student loans and rising housing costs. They have just introduced new lessons in how to manage debts. Nikki Fairweather, aged 15 from St Helens, said that she had benefited from lessons on personal finance, but admitted that she still had a lot to learn about money.

72. Which of the following can be found from the five-year research project?
- A. Students understand personal finances differently.
 - B. University tuition fees in England have been rising.
 - C. Teenagers tend to overestimate their future earnings.
 - D. The students' payback ability has become a major issue.
73. The phrase "to raise the ceiling" in paragraph 2 probably means "_____".
- A. to raise the student loans
 - B. to improve the school facilities
 - C. to increase the upper limit of the tuition
 - D. to lift the school building roofs
74. According to Stephen Moir, students_____.
- A. are too young to be exposed to financial issues
 - B. should learn to manage their finances well
 - C. should maintain a positive attitude when facing loans
 - D. benefit a lot from lessons on personal finance
75. What can we learn from the passage?
- A. Many British teenagers do not know money matters well
 - B. Teenagers in Britain are heavily burdened with debts.
 - C. Financial planning is a required course at college.
 - D. Young people should become responsible adults.

(D)

The world economy has run into a brick wall. Despite countless warnings in recent years about the need to address a potential hunger crisis in poor countries and an energy crisis worldwide, world leaders failed to think ahead. The result is a global food crisis. Wheat, corn and rice prices have more than doubled in the past two years. And oil prices have increased more than three times since the start of 2004. These food-price increases combined with increasing energy costs, will slow if not stop economic growth in many parts of the world and will even affect political stability. Practical solutions to these problems do exist, but we'll have to start thinking ahead and acting globally.

Here are three steps to ease the current food crisis and avoid the potential for a global crisis. The first is to promote the dramatic success of Malawi, a country in southern Africa, which three years ago established a special fund to help its farmers get fertilizer and seeds with high productivity. Malawi's harvest doubled after just one year. An international fund based on the Malawi model would cost a mere \$10 per person annually in the rich world, or \$10 billion altogether.

Second, the U.S. and Europe should abandon their policies of paying partly for the change of food into biofuels. The U.S. government gives farmers a taxpayer-financed payment of 51 cents per gallon of *ethanol* (乙醇) changed from corn. There may be a case for biofuels produced on lands that do not produce foods - tree crops, grasses and wood products - but there's no case for the government to pay to put the world's dinner into the gas tank.

Third, we urgently need to weather-proof the world's crops as soon and as effectively as possible. For a poor farmer, sometimes something as simple as a farm pond - which collects rainwater to be used in dry weather - can make the difference between a good harvest and a bad one. The world has already committed to establishing a Climate Adaptation fund to help poor regions climate-proof vital economic activities such as food production and health care but has not yet acted upon the promise.

76. An international fund based on the Malawi model would _____.
A. cost each of the developed countries \$10 billion per year
B. aim to double the harvest in southern African countries in a year
C. decrease the food prices as well as the energy prices
D. give poor farmers access to fertilizer and highly productive seeds
77. With the second step, the author expresses the idea that _____.
A. it is not wise to change food crops into gas
B. it is misleading to put tree crops into the gas tank
C. we should get alternative forms of fuel in any way
D. biofuels should be developed on a large scale
78. Which of the following is true according to the passage?
A. A rain-collecting pond is a simple safeguard against dry well
B. A Climate Adaptation Fund has been established to help poor
C. The world has made a serious promise to build farm ponds.
D. It makes a great difference whether we develop wood products or not.

79. In the passage, the author calls on us to _____.
A. slow down but not to stop economic
B. develop tree crops, grasses and wood products
C. achieve economic growth and political stability
D. act now so as to relieve the global food shortage

Section C

Directions: Read the following text and choose the most suitable heading from A-F for each paragraph. There is one extra heading which you do not need.

- A. Maintain a Balanced Diet.
B. Keep a Healthy Self-image.
C. Difficulty in Curing Eating Disorders
D. Best Prevention of Eating Disorders
E. What Factors Bring about an Eating Disorder?
F. How Does One Develop an Eating Disorder?

80.

The causes of eating disorders are not clear. There may be genetic or biochemical factors in some cases. There may be psychological problems from early childhood or the present (such as school or family conflicts) that trigger the problem. Often, there is the double pressure to enjoy life through food and yet remain ultra-slim. Society all around us encourages eating and drinking as main ways to enjoy life. To be popular, you are supposed to eat, eat, eat.

81.

But in real life most people, if they keep consuming like that, just keep pulling on more and more weight. They then find themselves in conflict with another dominant pressure in society - to stay slim and trim. Models in ads, even those shown eating fairy foods, are usually physically lit and quite thin. All of these pressures put teens in a terrible situation. The "solution" that some teens choose is to deny themselves all the time so as to keep temptation away. But soon, hunger and food boredom lead to overeating. Over time they develop anorexia or bulimia.

82.

Once an eating disorder has become firmly established, there is no easy cure. Someone who's never suffered anorexia and never known an anorexic might be tempted to think. "It's simple - just tell them to eat more!" Unfortunately, it's not that simple. Even when anorexics have been brought into the hospital and are receiving physical care, nutritional therapy, and psychiatric care, many don't improve much.

83.

Our society's obsession with thinness, together with a constant emphasis on the theme that rich food means pleasure, puts many teens in a difficult situation. Being drawn to overeating on the one hand and self-denial on the other can bring about anorexia or bulimia - or both. Anorexics seem to feel that no matter how much weight they lose, they are still too fat. Therefore, it is important for each person in our society to try to maintain a healthy and realistic self-image. Don't compare yourself with the models and actors in the media. Set your sights more realistically by comparing yourself with family and friends, if anyone.

84.

Self-denial is the typical first step toward an eating disorder. Eventually, desires for what you've denied yourself become unbearable, and you react with either a binge or overly fierce self-control. The answer is not of course. To eat all the snacks and treats you fancy. But to avoid that first step that leads to overeating, strive for a good and healthy diet. Such a diet is based primarily on grain products, fruits, and vegetables, with moderate amounts of meat and dairy products and with small amounts of snacks and desserts. Research demonstrates that this kind of diet leaves you more alert and energetic.

第 II 卷 （共 45 分）

I. Translation

Directions: Translate the following sentences into English, using the words given in the brackets.

1. 我们打篮球的时间到了。 (time)
2. 他设法把游客及时送到了机场。 (manage)
3. 你今晚能来参加我的生日聚会? (possible)
4. 应该鼓励年轻人按照自己的特长选择职业。 (encourage)
5. 我对学生所谈的电子产品一无所知，我发现自己落伍了。 (ignorant)
6. 尽管遭受如此严重的自然灾害，但只要不灰心，我们终会克服暂时的困难。
(Although...)

II. Guided Writing

Directions: Write an English composition in 120 - 150 words according to the instructions given below in Chinese.

你班将组队参加学校组织的集体舞比赛 (group dancing competition)，班长希望大家积极参加。对此谈谈你的看法。

你的文章必须包括以下内容：

- 你是否会参加比赛；
- 你做出该决定的具体理由。

2011 年全国普通高等学校招生统一考试

上海 英语试卷

考生注意：

1. 考试时间 120 分钟，试卷满分 150 分。
2. 本考试设试卷和答题纸两部分。试卷分为第 I 卷（第 1-11 页）和第 II 卷（第 12 页），全卷共 11 页。所有答题必须涂（选择题）或写（非选择题）在答题纸上，做在试卷上一律不得分。
3. 答题前，务必在答题纸上填写准考证号和姓名，并将核对后的条形码贴在指定位置上。

第 I 卷 （共 105 分）

I. Listening comprehension:

Section A

Directions: In Section A; you will hear ten short conversations between two speakers. At the end of each conversation, a question will be asked about what was said. The conversations and the questions will be spoken only once. After you hear a conversation and the question about it, read the four possible answers on your paper, and decide which one is the best answer to the question you have heard.

1. A. At a train station. B. At an airport C. At a travel agency D. At a bus station.
2. A. \$5. B. \$10. C. \$15. D. \$50.
3. A. Receptionist and guest. B. Salesperson and customer.
C. Doctor and patient. D. Waiter and diner.
4. A. Excited. B. Dissatisfied. C. Bored. D. Exhausted.
5. A. Her hair has changed. B. She isn't satisfied with her hair style.
C. She prefers to wear long hair. D. The man has changed his hair style.
6. A. It is too early to watch the Talent Show.
B. He will go to bed in five minutes.
C. He would rather watch TV than go to bed.
D. He is old enough to stay up.
7. A. She has got everything ready. B. She never hesitates over what to take.
C. She hates packing by herself. D. She needs more time for packing.
8. A. They should wait for John for a while.
B. They should stay here for the night.
C. They should start the meeting right away.
D. They should call John at once.
9. A. She is unwilling to move into a new flat.
B. Her neighbors get along well with her.
C. She can't tell the man why she is moving.
D. Her neighbors usually play their TV loud.
10. A. Ask for directions. B. Try a different route.
C. Go back for the map. D. Cancel their trip.

Section B

Directions: In Section B, you will hear two short passages, and you will be asked to questions on each of the passages. The passages will be read twice, but the questions will be spoken only once. When you hear a question, read the four possible answers on your paper decide which one would be the best answer to the question you have heard.

Questions 11 through 13 are based on the following passage.

11. A. A political system. B. Religion. C. Working language. D. Race.
12. A. Discuss current issues. B. Join in a writing competition.
 C. Attend an arts and crafts competition. D. Celebrate their friendship.
13. A. The Commonwealth Games. B. An important holiday.
 C. The Commonwealth members. D. An international association.

Questions 14 through 16 are based on the following speech.

14. A. Equipping students with knowledge. B. Qualifying students for certain jobs.
 C. Developing students' habits of mind. D. Helping students to go to graduate school.
15. A. The ability to have critical analysis. B. Creative use of leisure time.
 C. Logical use of information. D. Willingness to accept uncertainty.
16. A. Goals to reach in a college education. B. Roles of knowledge in students' growth.
 C. Qualifications needed for a job. D. Importance of after-class activities.

Section C

Directions: In Section C, you will hear two longer conversations. The conversations will be read twice. After you hear each conversation, you are required to fill in the numbered blanks with the information you have heard. Write your answers on your answer sheet.

Blanks 17 through 20 are based on the following conversation.

Complete the form. Write **ONE WORD** for each answer.

Complaint	Form
Caller:	Mary White
Phone No.:	<u>17</u> .
Location of Problem:	A <u>18</u> restaurant, 449 Shanghai Street
Details:	It dumps its <u>19</u> on the street. It doesn't put bottles and cans in <u>20</u> bins.

Blanks 21 through 24 are based on the following conversation.

Complete the form. Write **NO MORE THAN THREE WORDS** for each answer.

How long does short memory last?	It lasts only <u>21</u> .
What is an example of medium term memory?	Buying bread, a sort of <u>22</u> of things to do.
What is long term memory concerned with?	<u>23</u> that happen in your life such as your wedding.
How is long term memory different from the others?	It <u>24</u> .

II. Grammar and Vocabulary

Section A

Directions: Beneath each of the following sentences there are four choices marked A, B, C and D. Choose the one answer that best completes the sentence.

25. Graduation is a good time to thank those who have helped you _____ the tough years.
A. through B. up C. with D. from
26. To stay awake, he finished a cup of coffee and ordered _____.
A. the other B. other C. the others D. another
27. It's no use _____ without taking action.
A. complain B. complaining C. being complained D. to be complained
28. I _____ worry about my weekend - I always have my plans ready before it comes.
A. can't B. mustn't C. daren't D. needn't
29. When Mom looked back on the early days of their marriage, she wondered how they had managed with _____ money.
A. so few B. such few C. so little D. such little
30. It doesn't matter if they want to come to your party, _____?
A. doesn't it B. does it C. don't they D. do they
31. After getting lost in a storm, a member of the navy team _____ four days later.
A. rescued B. was rescued C. has rescued D. had been rescued
32. The rare fish, _____ from the cooking pot, has been returned to the sea.
A. saved B. saving C. to be saved D. having saved
33. At one point I made up my mind to talk to Uncle Sam. Then I changed my mind, _____ that he could do nothing to help.
A. to realize B. realized C. realizing D. being realized
34. Did you predict that many students _____ up for the dance competition?
A. would sign B. signed C. have signed D. had signed
35. There is clear evidence _____ the most difficult feeling of all to interpret is bodily pain.
A. what B. if C. how D. that
36. If a lot of people say a film is not good, I won't bother to see it, or I'll wait _____ it comes out on DVD.
A. whether B. after C. though D. until
37. The police officers in our city work hard _____ the rest of us can live a safe life.
A. in case B. as if C. in order that D. only if
38. The message you intend to convey through words may be the exact opposite of _____ others actually understand.
A. why B. that C. which D. what
39. You'll find taxis waiting at the bus station _____ you can hire to reach your host family.
A. which B. where C. when D. as
40. Today we have chat rooms, text messaging, emailing, but we seem _____ the art of communicating face-to-face.
A. losing B. to be losing C. to be lost D. having lost

Section B

Directions: Complete the following passage by using the words in the box. Each word can only be used once. Note that there is one word more than you need.

A. launched	B. unpleasant	C. applying	D. technically	E. impact
F. coating	G. fixed	H. miraculously	I. superior	J. advances

How would you like to wear the same underwear (内衣裤) for weeks? Owing to the work that has gone into developing intelligent materials, this may not be as 41 as it sounds. Self-cleaning clothes have now been created, and these new materials provide 42 resistance to dirt as well as water. As a result, they require much less cleaning than traditional materials.

The creation of self-cleaning clothes provides an example of how nature helps scientists develop better products. This self-cleaning nature is known as the “lotus effect”. The name comes, of course, from the lotus leaves, which are famous for growing in muddy lakes and rivers while remaining almost 43 clean. By observing nature, scientists are 44 the qualities of the lotus leaves to the materials they have engineered. Because of this, some remarkable new products have been 45. Among them are special windows that are resistant to dirt and water. A special 46 on these windows not only prevents dirt from sticking to their surfaces, but also allows dust to be easily washed off by the rain. In fact, these new windows have already been 47 to some cars. Even when traveling at high speed through rain, these cars never have to use their windshield wipers (雨刮器).

Although we have already seen some practical applications, even more dramatic 48 will be made in the future, and they will, perhaps, change our world completely. Undoubtedly, technology is an important development, and it will have an even bigger 49 on our lives.

III. Reading Comprehension

Section A

Directions: For each blank in the following passage there are four words or phrases marked A, B, C and D. Fill in each blank with the word or phrase that fits the context.

Everyone in business has been told that success is all about attracting and retaining (留住) customers. It sounds simple and achievable. But, 50, words of wisdom are soon forgotten. Once companies have attracted customers they often 51 the second half of the story. In the excitement of beating off the competition, negotiating prices, securing orders, and delivering the product, managers tend to become carried away. They forget what they regard as the boring side of business - 52 that the customer remains a customer.

53 to concentrate on retaining as well as attracting customers costs business huge amounts of money annually. It has been estimated that the average company loses between 10 and 30 per cent of its customers every years. In constantly changing 54, this is not surprising. What is surprising is the fact that few companies have any idea how many customers they have lost.

Only now are organizations beginning to wake up to those lost opportunities and calculate the 55 implications. Cutting down the number of customers a company loses can make a big 56 in its performance. Research in the US found that a five per cent decrease in the number of defecting (流失的) customers led to 57 increases of between 25 and 85 per cent.

In the US, Domino's Pizza estimates that a regular customer is worth more than \$5,000 over ten years. A customer who receives a poor quality product or service on their first visit and 58 never returns, is losing the company thousands of dollars in 59 profits (more if you consider how many people they are likely to tell about their bad experience).

The logic behind cultivating customer 60 is impossible to deny. "In practice most companies' marketing effort is focused on getting customers, with little attention paid to 61 them", says Adrian Payne of Cornfield University' School of Management. "Research suggests that there is a close relationship between retaining customers and making profits. 62 customers tend to buy more, are predictable and usually cost less to service than new customers. Furthermore, they tend to be less price 63, and may provide free word-of-mouth advertising. Retaining customers also makes it 64 for competitors to enter a market or increase their share of a market.

- | | | | |
|----------------------|------------------|------------------|--------------------|
| 50. A. in particular | B. in reality | C. at least | D. first of all |
| 51. A. emphasize | B. doubt | C. overlook | D. believe |
| 52. A. denying | B. ensuring | C. arguing | D. proving |
| 53. A. Moving | B. Hoping | C. Starting | D. Failing |
| 54. A. markets | B. tastes | C. prices | D. expenses |
| 55. A. culture | B. social | C. financial | D. economical |
| 56. A. promise | B. plan | C. mistake | D. difference |
| 57. A. cost | B. opportunity | C. profit | D. budget |
| 58. A. as a result | B. on the whole | C. in conclusion | D. on the contrary |
| 59. A. huge | B. potential | C. extra | D. reasonable |
| 60. A. beliefs | B. loyalty | C. habits | D. interest |
| 61. A. altering | B. understanding | C. keeping | D. attracting |
| 62. A. Assumed | B. Respected | C. Established | D. Unexpected |
| 63. A. agreeable | B. flexible | C. friendly | D. sensitive |
| 64. A. unfair | B. difficult | C. essential | D. convenient |

Section B

Directions: Read the following four passages. Each passage is followed by several questions or unfinished statements. For each of them there are four choices marked A, B, C and D. Choose the one that fits best according to the information given in the passage you have just read.

(A)

The teacher who did the most to encourage me was, as it happened, my aunt. She was Myrtle C. Manigault, the wife of my mother's brother Bill. She taught in second grade at all-black Summer School in Camden, New Jersey.

During my childhood and youth, Aunt Myrtle encouraged me to develop every aspect of my potential, without regard for what was considered practical or possible for black females. I liked to sing; she listened to my voice and pronounced it good. I couldn't dance; she taught me the basic dancing steps. She took me to the theatre - not just children's theatre but adult comedies and dramas - and her faith that I could appreciate adult plays was not disappointed.

My aunt also took down books from her extensive library and shared them with me. I had books at home, but they were all serious classics. Even as a child I had a strong liking for humor, and I'll never forget the joy of discovering Don Marquis's Archy & Mehitabel through her.

Most important, perhaps, Aunt Myrtle provided my first opportunity to write for publication. A writer herself for one of the black newspapers, she suggested my name to the editor as a "youth columnist". My column, begun when I was fourteen, was supposed to cover teenage social activities - and it did - but it also gave me the freedom to write on many other subjects as well as the habit of gathering material, the discipline of meeting deadlines, and, after graduation from college six years later, a solid collection of published material that carried my name and was my passport to a series of writing jobs.

Today Aunt Myrtle is still an enthusiastic supporter of her "favorite niece". Like a diamond, she has reflected a bright, multifaceted(多面的) image of possibilities to every pupil who has crossed her path.

65. Which of the following did Aunt Myrtle do to the author during her childhood and youth?

- A. She lent her some serious classics.
- B. She cultivated her taste for music.
- C. She discovered her talent for dancing.
- D. She introduced her to adult plays.

66. What does Archy & Mehitabel in Paragraph 3 probably refer to?

- A. A book of great fun.
- B. A writer of high fame.
- C. A serious masterpiece.
- D. A heartbreaking play.

67. Aunt Myrtle recommended the author to a newspaper editor mainly to ____.

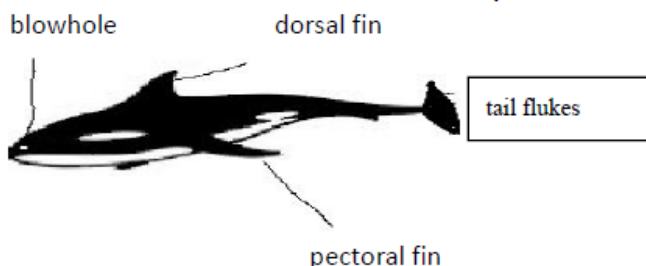
- A. develop her capabilities for writing.
- B. give her a chance to collect material
- C. involve her in teenage social activities
- D. offer her a series of writing jobs.

68. We can conclude from the passage that Aunt Myrtle was a teacher who ____.

- A. trained pupils to be diligent and well-disciplined
- B. gave pupils confidence in exploiting their potential
- C. emphasized what was practical or possible for pupils
- D. helped pupils overcome difficulties in learning

(B)

Humpback Whales



Humpback whales are sometimes called performers of the ocean. This is because they can make impressive movements when they dive. The name “humpback”, which is the common name for this whale, refers to the typical curve shape the whale’s back forms as it dives.

Sometimes the humpback will dive with a fantastic movement known as a **breach**. During breaching, the whale uses its powerful tail flukes to lift nearly two-thirds of its body out of the water in a giant leap. A breach might also include a sideways twist with fins stretched out like wings, as the whale reaches the height of the breach.

A humpback whale breathes air at the surface of the Water through two blowholes which are located near the top of the head. It blows a double stream of water that can Rise up to 4 meters above the water.

The humpback has a small dorsal fin located towards the tail flukes about two-thirds of the way down its back.

Other distinguishing features include large pectoral fins, which may be up to a third of the body length, and unique black and white spots on the underside of the tail flukes. These markings are like fingerprints: no two are the same. Humpback whales live in large groups. They communicate with each other through complex “songs”.

Quick Facts

Size: 14m - 18m in length;

30 - 50 tons in weight.

Living: Open ocean and shallow
coastline waters.

Environment: From warm tropical (热带的)
waters, where they breed,
to cold polar waters, where they eat.

Diet: Shellfish, plants and fish of small size.

Hunting: Sometimes in groups, in which
several whales form a circle under the
water, blowing bubbles that form a
“net” around a school of fish. The fish
are then forced up to the surface in a
concentrated mass.

Current state: endangered; it is estimated that
there are about 5000-7500 humpback whales
worldwide.

69. According to Quick Facts, a humpback whale ____.

- A. cannot survive in waters near the shore
- B. doesn't live in the same waters all the time
- C. lives mainly on underwater plants
- D. prefers to work alone when hunting food

70. To make a **breach**, a humpback whale must ____.

- A. use its tail flukes to leap out of the water
- B. twist its body sideways to jump high
- C. blow two streams of water
- D. communicate with a group of humpbacks

71. From the passage we can learn that a humpback whale ____.

- A. has its unique markings on its tail flukes
- B. has black and white fingerprints
- C. gets its name from the way it hunts
- D. is a great performer due to its songs

(C)

Human remains of ancient settlements will be reburied and lost to science under a law that threatens research into the history of humans in Britain, a group of leading archaeologists (考古学家) says. A letter addressed to the justice secretary, Ken Clarke, 40 archaeologists write of their “deep and widespread concern” about the issue. It centers on the law introduced by the Ministry of Justice in 2008 which requires all human remains unearthed in England and Wales to be reburied within two years, regardless of their age. The decision means scientists have too little time to study bones and other human remains of national and cultural significance.

“Your current requirement that all archaeologically unearthed human remains should be reburied, whether after a standard period of two years or further special extension, is contrary to basic principles of archaeological and scientific research and of museum practice,” they write.

The law applies to any pieces of bone uncovered at around 400 dig sites, including the remains of 60 or so bodies found at Stonehenge in 2008 that date back to 3,000 BC. Archaeologists have been granted a temporary extension to give them more time, but eventually the bones will have to be returned to the ground.

The arrangements may result in the waste of future discoveries at sites such as Happisburgh in Norfolk, where digging is continuing after the discovery of stone tools made by early humans 950,000 years ago. If human remains were found at Happisburgh, they would be the oldest in northern Europe and the first indication of what this species was. Under the current practice of the law those remains would have to be reburied and effectively destroyed.

Before 2008, guidelines allowed for the proper preservation and study of bones of sufficient age and historical interest, while the Burial Act 1857 applied to more recent remains. The Ministry of Justice assured archaeologists two years ago that the law was temporary, but has so far failed to revise it.

Mike Parker Pearson, an archaeologist at Sheffield University, said: “Archaeologists have been extremely patient because we were led to believe the ministry was sorting out this problem, but we feel that we cannot wait any longer.”

The ministry has no guidelines on where or how remains should be reburied, or on what records should be kept.

72. According to the passage, scientists are unhappy with the law mainly because

-
- A. it is only a temporary measure on the human remains
 - B. it is unreasonable and thus destructive to scientific research
 - C. it was introduced by the government without their knowledge
 - D. it is vague about where and how to rebury human remains

73. Which of the following statements is true according to the passage?

- A. Temporary extension of two years will guarantee scientists enough time.
- B. Human remains of the oldest species were dug out at Happisburgh.
- C. Human remains will have to be reburied despite the extension of time.
- D. Scientists have been warned that the law can hardly be changed.

74. What can be inferred about the British law governing human remains?

- A. The Ministry of Justice did not intend it to protect human remains.
- B. The Burial Act 1857 only applied to remains uncovered before 1857.
- C. The law on human remains hasn't changed in recent decades.
- D. The Ministry of Justice has not done enough about the law.

75. Which of the following might be the best title of the passage?

- A. New discoveries should be reburied, the government demands.
- B. Research time should be extended, scientists require.
- C. Law on human remains needs thorough discussion, authorities say.
- D. Law could bury ancient secrets for ever, archaeologists warn.

Section C

Directions: Read the following text and choose the most suitable heading from A - F for each paragraph. There is one extra heading which you do not need.

- A. Manufacturing industry in information economy
- B. News in the age of information
- C. Argument about individual accounts and their reliability
- D. Be your own investigative journalist
- E. Don't believe everything you read in the newspaper
- F. Information is presented in an interesting way

76.

With the arrival of the age of "information economy", intellectual work is becoming a more important source of wealth than manufacturing. Organizations in all walks of life are doing more to spread their information. So people of the Public Relations are hired to speak for them. A lot of our news is actually collected from press releases and reports of events intentionally staged for journalists. In the information age, journalists spend their time, not investigating, but passing on the words of a spokesperson.

77.

There is a joke in the novel *Scoop* about the newspaper's owner, Lord Copper. The editors can never disagree with him. When he is right about something they answer "definitely", and when he's wrong, they say "to some extent, Lord Copper". It seems reasonable to suppose that, in the real world, the opinions of such powerful people still influence the journalists and editors who work for them.

78.

In countries where the news is not officially controlled, it may be provided by commercial organizations which depend on advertising. The news has to attract viewers and maintain its audience ratings. I suspect that some stories get air-time just because there happen to be exciting pictures to show. In Britain, we have the tabloid newspapers which millions of people read simply for entertainment. There is progressively less room for historical background, or statistics, which are harder to present as a sensational story.

79.

There is an argument that with spreading access to the internet and cheap technology for recording sound and images we will all be able to find exactly the information we want. People around the world will be able to publish their own eye-witness accounts and compete with the widely-accepted news-gatherers on equal terms. But what it will mean also is that we'll be subjected to a still greater amount of nonsense and lies. Any web log may contain the least information of the year, or equally, a made-up story that you will never be able to check.

80.

Maybe the time has come to do something about it, and I don't just mean changing your choice of TV channel or newspaper. In a world where everyone wants you to listen to their version, you only have two choices: switch off altogether or start looking for sources you can trust. The investigative journalist of the future is everyone who wants to know the truth.

Section D

Directions: Read the passage carefully. Then answer the questions or complete the statements in the fewest possible words.

Sociologists have long recognized that organizations of less than 200 individuals can operate through the free flow of information among the members. Once their size goes beyond this figure, the organizations are getting less flexible. So it seems necessary to prevent total disorder resulting from failures of communication.

One solution to this problem would, of course, be to structure large organizations into smaller units of a size that can act as a group. By allowing these groups to build reliance on each other, larger organizations can be built up. However, merely having groups of, say, 150 will never of itself be a complete solution to the problems of the organization. Something else is needed: the people involved must be able to build direct personal relationships. To allow free flow of information, they have to be able to communicate with each other in a casual way. Maintaining too formal a structure of relationships inevitably presents the way a system works.

The importance of this was drawn to my attention two years ago by the case of a TV station. Whether by chance or by design, it so happened that there were almost exactly 150 people in the station. The whole process worked very smoothly as an organization for many years until they were moved into purpose-built accommodation. Then, for no apparent reason, the work seemed to be more difficult to do, not to say less satisfying.

It was some time before they worked out what the problem was. It turned out that, when the architects were designing the new building, they decided that the coffee room where everyone ate their sandwiches at lunch times was an unnecessary luxury and so did away with it. And with that, they accidentally destroyed the close social networks that strengthened the whole organization. What had apparently been happening was that, as people gathered informally over their sandwiches in the coffee room, useful information was casually being exchanged.

(Note: Answer the questions or complete the statements in **NO MORE THAN EIGHT WORDS**)

81. What size of an organization may lead to communication failures?
82. What are the two solutions to the communication problem within a large organization?
83. After the TV station moved into new accommodation, its operation _____.
84. From the case of the TV station, we can conclude it is _____ that make(s) an organization more successful.

第 II 卷 （共 45 分）

I. Translation

Directions: Translate the following sentences into English, using the words given in the brackets.

1. 你为什么不在网上订票呢？（Why）
2. 我常把王海误认为他的双胞胎弟弟，因为他们长得太像了。（mistake）
3. 对父母而言，没有什么能与孩子的身心健康相比。（compare）
4. 自从出国留学后，她就不再和我们保持联系了。（No longer）
5. 如果能找到任何适合你的学习方法，你的学习效率就可能明显提高。（whatever）

II. Guided Writing

Directions: Write an English composition in 120 - 150 words according to the instructions given below in Chinese.

假如你是启明中学（Ming Qi Middle School）的李明，想申请一个扶贫项目，帮助贫困地区的儿童。根据以下启事，写一封申请信（信中不能提到真实姓名和学校）。

启事

国际儿童基金会将资助中学生开展扶贫项目，以帮助贫困地区的儿童。申请成功者将获得项目经费 2000 元。

有意者请来信告知：

- 1) 你个人的基本情况；
- 2) 你对申请项目的基本设想；
- 3) 项目经费的使用计划。

联系方式: 2011hope@icf.org

2011

Appendix D: Marking Standards for Paper II Writing Section

Part I: Translation

1. For item 1 – 3, four marks each; for item 4 – 5, five marks each;
2. For each item, one mark is deducted for two errors of either spelling, vocabulary, use of punctuation, or use of capital or small letters;
3. One mark is deducted for each grammatical error;
4. One mark is deducted if test-taker does not use the given vocabulary.

Part II: Guided Writing

Band / marks	Content	Use of vocabulary and grammar	Organisation and structure
A	9-10	9-10	4-5
B	7-8	7-8	3
C	5-6	5-6	2
D	3-4	3-4	1
E	0-2	0-2	0

1. The total mark in the guided writing section is 25: 10 marks for content, 10 for the use of vocabulary and grammar, and 5 for organisation and structure;
2. Marks for each section are determined by different bands, which are described below in detail.
3. If the total number of words is fewer than 70, the mark for this section cannot exceed more than 10 marks (out of 25).

*Marking requirements for each **band** in the **content** section:*

- A. Express excellent ideas in the composition;
- B. Express good ideas in the composition;
- C. Express moderately good ideas in the composition;
- D. Express fair ideas in the composition, and some content is irrelevant;
- E. Obvious omission of the principal content in the composition.

*Marking requirements for each **band** in the **use of vocabulary and grammar** section:*

- A.** Excellent level of accuracy of form, meaning and use of vocabulary and grammar throughout the composition;
- B.** Good level of accuracy of vocabulary and grammar with only a few minor errors in the composition;
- C.** Reasonable control of vocabulary and grammar but with a few errors in the composition;
- D.** Fair level of vocabulary and grammar with some errors, sentence structures are simple and range of vocabulary is narrow;
- E.** Limited control of simple vocabulary, grammar and sentence structures with a lot of errors.

*Marking requirements for each **band** in the **organisation and structure** section:*

- A.** Excellent level of cohesion and coherence in the composition;
- B.** Good level of cohesion and coherence in the composition;
- C.** Moderate level of cohesion and coherence using simple sentence structures with a few inconsistencies in the composition;
- D.** Fair level of cohesion and coherence with some inconsistencies in the composition;
- E.** Poor level of cohesion and coherence with a lot of inconsistencies in the composition.

Appendix E: Interview Questions

Interview questions with teacher

- a. What do you expect your students will do after today class?
- b. Do you follow every task or activity provided from the text book? Why? If not, which tasks / activities in class you used were not from the text book? Why did you use such tasks / activities?
- c. Did you refer to any other textbooks or references for today's teaching? If yes, which books? Why do you think you need the references? If not, why not?
- d. Is [what you taught today] part of the English Curriculum Standards (ECS) or Exam Specifications?
- e. To what extent did your teaching today relate to the ECS?
- f. How much of the ECS or Exam Specifications did you follow in your teaching?
- g. What books or information such as the ECS do you read or refer to before you prepare the lessons?
- h. How much do you think what you taught today is related to the UEE?
- i. Would you change your teaching content or teaching styles as the UEE draws closer? Why or why not? If, yes, how would you do that?
- j. During the class you referred to "something related to the UEE"! Why do you think that particular element is very important in the UEE? Why did you deliberately mention it?
- k. I saw that you were using some multi-media materials. How do those materials relate to the ECS or the UEE?
- l. Before you prepare your lesson plan, do you always set up the goals and objectives for every lesson?
- m. What would you think if the content of the UEE was aligned with the goals and objectives of instruction (teaching) and with instructional activities?

Interview questions with students

- a. What kind of work will you do on English after school? How?
- b. Will you review exactly what your teacher taught today? Why? Which part of today's class will you revise?
- c. Do you think materials the teacher teaches in class are enough for the UEE?

- d. What parts of today's class do you think are directly related to the UEE, and how?
- e. Do you think the teacher's focus changes as the exams draw nearer? How does that change manifest itself? What kinds of changes?
- f. Have you heard of "the English Curriculum Standards" or "Exam Specifications"? If yes, what are they?
- g. Do you attend any private English tutorial outside the class? Why?
- h. What are the differences between what the teachers teach in the tutorial and what is/was taught in your school? Can you tell me more about what you learn from the tutorial?
- i. Are there similarities between the school teacher on the one hand and private tutors in a private tutorial on the other?
- j. If the UEE were cancelled, how would you go about learning English? How would your attitude to learning change?

Interview questions with UEE designer

- a. How do the UEE designers set up the UEE papers? In other words, what is the UEE development process?
- b. After the UEE questions are set, do any native English speakers proof-read the questions?
- c. How about the test piloting? Is the format the same every year?
- d. When and how did the Shanghai Exam Centre set up the Exam Specifications and Requirements?
- e. Do you think the UEE is a proficiency test or an achievement test?
- f. What do you think the future direction of the UEE should be or is likely to be?
- g. Have you ever heard of wash-back? Do you think the UEE designers consider the concept of wash-back when they construct the UEE papers?
- h. Do you think the UEE has more positive or negative impact on school?
- i. In China, do students learn grammar when they study Chinese? When do students start to learn English grammar?
- j. Do you think the UEE is a fair exam?