

**IMPACT OF TOEIC AS A UNIVERSITY EXIT TEST:
A VIETNAMESE PERSPECTIVE**

By

NGUYEN THI THANH HA

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
(2015)

Abstract

The Test of English for International Communication (TOEIC) has recently become a language exit test in many Vietnamese universities. Despite the high stakes of the test for a large number of Vietnamese students, no robust research has been done to explore its impact on teaching and learning yet. Also, although the powerful influence of tests on teaching and learning has been well documented, the way test impact occurs is very complicated and varied because of a myriad of context-specific factors. Language testers, therefore, have called for more research in different contexts.

In response to the needs mentioned above, this study examined the impact of TOEIC on teaching and learning in Vietnamese universities with an emphasis on discovering the variations of the test impact due to school settings. To meet the research objectives, a mixed-methods study was conducted at two different locations in Vietnam. The perspectives from three groups of stakeholders, namely educational policymakers, teachers, and students were sought by means of semi-structured interviews, questionnaires, and case study. The study was conducted in two phases. Phase 1 aimed at understanding policymakers' perspectives on the TOEIC-as-exit-test policy and identifying the general patterns of the actual test impact from the teachers' and students' perspectives. Phase 2 had the purpose of explaining the patterns found in Phase 1 through a case study. Both the questionnaire and case study participants came from the same three large universities.

The study gave interesting insight into the nature of the impact of the TOEIC test in the Vietnamese context. It showed a clear tendency of teaching and learning to the test since its introduction. However, the test impact varied greatly in form and intensity across different groups of participants although, contrary to expectations, school settings did not seem to be a major factor that caused this variability. On the one hand, the findings confirmed the indirect nature of the relationship between a test and its impact found in many other studies. On the other hand, they shed light on specific features of the mechanism of the impact of the TOEIC-as-exit-test, which might be attributed to the socio-cultural and educational context in Vietnam. The study suggested that the test impact on teaching and learning operated differently. While the influence of the TOEIC test on teaching was likely to be first filtered through the course factors, the latter was through students' abilities. In both cases, beliefs about communicative language teaching and

learning and some other person factors only played subservient roles in shaping the test impact. The study also had important implications for stakeholders, especially policymakers. It proved certain values of the TOEIC-as-exit-test policy as well as revealed negative issues associated with it. It suggested measures that need to be taken to modify the policy.

Acknowledgements

I would like to take this opportunity to express my deep gratitude to those who have given their invaluable support and assistance. My first special thanks go to my supervisors, Dr Peter Gu and Dr Angela Joe, for their warmest support and indispensable professional guidance. I feel fortunate to have learned a great deal from their research experience and knowledge. Their encouragement and direction made me more confident in what I was doing.

I also wish to thank all the other PhD candidates and staff at the School of Linguistics and Applied Language Studies for the fantastic, friendly, and supportive environment that they contributed to.

I would like to thank all my participants for their kindness, generosity, and enthusiasm and the valuable information that they provided, without which this thesis would not have been completed. I especially appreciate the long hours that the case study participants spent sharing their stories with me and their willing involvement in member-checking, which made this research more valid.

Further acknowledgement goes to all my friends and colleagues in New Zealand and Vietnam who gave me great spiritual support; many of whom gave me useful comments and suggestions for the development, piloting, and translation of my questionnaires.

I would also like to thank Beth Thomas for spending her precious Christmas time to proofread this thesis.

I am indebted to my family for their love and support. I am especially grateful to my parents, who always back me in all my academic and career goals. A big thank you goes to my husband Thang, my son Tuan, and my daughter Lan for their sharing of the ups and downs of my PhD journey and for their sacrifices and patience while accompanying me on this journey.

Last but not least, I acknowledge the New Zealand Aid Programme for granting me a doctoral scholarship, without which I would have never been able to pursue my PhD study in New Zealand, and the Faculty of Humanities and Social Sciences, Victoria University of Wellington, for its research grant, which enabled to me to collect data in my home country and disseminate my research findings at international conferences.

Table of Contents

ABSTRACT	i
ACKNOWLEDGEMENTS.....	iii
TABLE OF CONTENTS	v
LIST OF TABLES.....	xiii
LIST OF FIGURES	xvii
Chapter 1: INTRODUCTION	1
1.1 Rationales and Background	1
1.1.1 Rationales for study.	1
1.1.2 Vietnamese context and TOEIC-as-exit-test policy.	3
1.1.3 The TOEIC test.....	4
1.2 Purpose of Study and Research Design	9
1.3 Structure of Thesis	10
Chapter 2: LITERATURE REVIEW	13
2.1 Definitions of Washback and Impact	13
2.2 Dimensions of Washback	13
2.3 Effects of Tests on Teaching and Learning	15
2.3.1 Effects of tests on teachers’ and learners’ feelings and attitudes.	15
2.3.2 Effects of tests on teaching.	17
2.3.3 Effects of tests on learning.	23
2.3.4 Impact of TOEIC.	30
2.3.5 Influencing factors and models of test impact on teaching and learning.	35

2.4 Measurement-Driven Instruction	42
2.5 Chapter Summary	44
Chapter 3: METHODOLOGY.....	47
3.1 Research Questions.....	47
3.2 Overall Research Design	47
3.3 Participants and Research Sites	49
3.3.1 Policymakers.....	49
3.3.2 Teachers and students.	49
3.4 Reliability and Validity.....	57
3.4.1 Reliability and validity in quantitative research.	57
3.4.2 Reliability and validity in qualitative research.	59
3.5 Ethical Considerations	61
3.6 Phase 1 Methodology.....	61
3.6.1 Teacher and student questionnaires.	62
3.6.2 Interviews with policymakers.....	81
3.7 Phase 2 Methodology.....	84
3.7.1 Research design.	84
3.7.2 Data collection methods.....	85
3.7.3 Transformation of data and member checking.	88
3.7.4 Data analysis.	90
3.8 Chapter Summary	91
<i>PHASE 1 FINDINGS</i>	93
Chapter 4: FINDINGS FROM INTERVIEWS WITH POLICYMAKERS.....	95

4.1 Demographic Information	95
4.2 Purposes of Language Teaching at University	96
4.3 Rationales for Choosing TOEIC as a University Exit Test	97
4.3.1 First rationale – TOEIC is suitable.	98
4.3.2 Second rationale – TOEIC is prestigious.....	99
4.3.3 Third rationale – TOEIC has good quality.	100
4.3.4 Fourth rationale – Poor testing and assessment practice in Vietnam.	101
4.3.5 Fifth rationale – TOEIC is useful.	103
4.3.6 Sixth rationale – TOEIC is an independent assessment.	103
4.3.7 Seventh rationale – TOEIC is students’ preference.....	104
4.3.8 Other.	104
4.4 Intended Impact of the TOEIC-as-exit-test Policy	105
4.4.1 Intended impact on learning.	105
4.4.2 Intended impact on teaching.	106
4.4.3 Intended impact on language programmes.	107
4.5 Chapter Summary and Discussion.....	109
Chapter 5: TEACHER QUESTIONNAIRE FINDINGS	111
5.1 Demographic Information	111
5.2 Perceived Impact of TOEIC on Teaching	112
5.3 Factors Potentially Involved in Forming the Impact of TOEIC on Teaching	114
5.3.1 Teachers’ beliefs about language teaching and learning.	114
5.3.2 Teachers’ perceptions of TOEIC.	115
5.3.3 Teachers’ perceptions of context factors.	116

5.4 Involvement of Test Factors, Person Factors, and Context Factors in Forming the Impact of TOEIC on Teaching	116
5.4.1 Person factors and perceived impact of TOEIC on teaching.....	117
5.4.2 Perceptions of TOEIC and its perceived impact on teaching.	120
5.4.3 Perceptions of context factors and perceived impact of TOEIC on teaching ...	122
5.5 Differences between the Large City Universities and the Provincial University in the Perceived Impact of TOEIC on Teaching.....	124
5.6 Differences between the Large City Universities and the Provincial University in Factors Potentially Involved in Forming the Impact of TOEIC on Teaching	126
5.6.1 Differences between the large city universities and the provincial university in teachers' perceptions of TOEIC.	126
5.6.2 Differences between the large city universities and the provincial university in teachers' beliefs about language learning and teaching.....	127
5.6.3 Differences between the large city universities and the provincial university in teachers' perceptions of context factors.....	128
5.7 Chapter Summary and Discussion.....	128
Chapter 6: STUDENT QUESTIONNAIRE FINDINGS.....	131
6.1 Demographic Information.....	131
6.2 Perceived Impact of TOEIC on Student Learning.....	132
6.3 Factors Potentially Involved in Forming the Impact of TOEIC on Student Learning	134
6.3.1 Students' perceptions of TOEIC.....	134
6.3.2 Students' beliefs about language learning.	135
6.3.3 Students' perceptions of context factors.	136

6.4 Involvement of Test Factors, Person Factors, and Context Factors in Forming the Impact of TOEIC on Student Learning	136
6.4.1 Person factors and perceived impact of TOEIC on student learning.	137
6.4.2 Perceptions of TOEIC and its perceived impact of on student learning.....	140
6.4.3 Perceptions of context factors and perceived impact of TOEIC on student learning.	143
6.5 Differences between the Large City Universities and the Provincial University in the Perceived Impact of TOEIC on Student Learning.....	145
6.6 Differences between the Large City Universities and the Provincial Universities in Factors Potentially Involved in Forming the Impact of TOEIC on Student Learning ..	147
6.6.1 Differences between the large city universities and the provincial university in students' perceptions of TOEIC.	147
6.6.2 Differences between the large city universities and the provincial university in students' beliefs about language learning.....	149
6.6.3 Differences between the large city universities and the provincial university in students' perceptions of context factors.	151
6.7 Chapter Summary and Discussion.....	153
<i>PHASE 2 FINDINGS</i>	157
Chapter 7: TEACHER CASE STUDY.....	161
7.1 Demographic Information	161
7.2 A General Analysis of Teacher Cases	162
7.3 Narratives of Teacher Cases	164
7.3.1 Case 1: Phuong.	165
7.3.2 Case 2: Lan.	173
7.3.3 Case 3: Phong.	184

7.4 Chapter Summary and Discussion	194
Chapter 8: STUDENT CASE STUDY	199
8.1 Demographic Information.....	199
8.2 A Quantitative Analysis of Student Cases	201
8.3 Narratives of Student Cases	204
8.3.1 Case 1: Khoi.....	204
8.3.2 Case 2: Thien.	212
8.3.3 Case 4: Toan.	221
8.4 Chapter Summary and Discussion	231
Chapter 9: DISCUSSION	235
9.1 Summary of Major Findings.....	235
9.1.1 Policymakers' perspectives on the TOEIC-as-exit-test policy.	235
9.1.2 Actual impact of TOEIC on teaching and factors involved in test impact generation.....	236
9.1.3 Actual impact of TOEIC on learning and factors involved in test impact generation.....	237
9.2 Discussion of Major Issues	239
9.2.1 Intended and actual impact, positive and negative impact.	239
9.2.2 Discrepancy in perceptions of TOEIC validity between policymaking level and implementation level.....	244
9.2.3 Differences in test impact between large city universities and provincial university.	246
9.2.4 Mechanism of test impact on teaching and learning in the Vietnamese context.	252

9.2.5 A critical look at the TOEIC-as-exit-test policy.....	258
9.3 Chapter Summary	262
Chapter 10: CONCLUSION.....	263
10.1 Purpose, Methods, Findings, and Significance.....	263
10.2 Limitations.....	266
10.3 Implications	269
10.3.1 Implications for test developers.....	269
10.3.2 Implications for policymakers.	269
10.3.3 Implications for teachers and learners.	271
10.4 Recommendations for Further Research	272
BIBLIOGRAPHY	273
APPENDICES	291

List of Tables

Table 1.1 Original and Redesigned TOEIC Listening and Reading Test Format	6
Table 1.2 TOEIC Speaking and Writing Test Format	7
Table 3.1 Policymaker Participants	49
Table 3.2 Participants in Questionnaire Piloting	50
Table 3.3 Research Sites for Main Study	52
Table 3.4 Number of Teachers Participating in the Main Study – Phase 1	53
Table 3.5 Number of Students Participating in the Main Study – Phase 1	54
Table 3.6 Teacher Case Study Participants (N=7)	56
Table 3.7 Student Case Study Participants (N=12)	57
Table 3.8 Participants Involved in Questionnaire Piloting and Translation Checking.....	69
Table 3.9 Description of Teacher Questionnaire	70
Table 3.10 Description of Student Questionnaire.....	71
Table 3.11 Results of Factor Analysis of Student Questionnaire	77
Table 3.12 Results of Factor Analysis of Teacher Questionnaire	79
Table 3.13 Results of Member Checking with Policymaker Participants	83
Table 3.14 Description of Journal Questions	88
Table 3.15 Information on Classroom Observations	89
Table 4.1 Policymakers’ Opinions Regarding the Purpose of Language Teaching at University (N=14).....	96
Table 4.2 Rationales for Choosing TOEIC as a University Exit Test (N=14)	98
Table 4.3 Reasons Why TOEIC is Suitable (N=14).....	99

Table 4.4 Intended Impact of TOEIC	105
Table 5.1 Demographic Information of Teacher Participants – Part 1 (N=81)	112
Table 5.2 Demographic Information of Teacher Participants – Part 2 (N=81)	112
Table 5.3 Descriptive Statistics on Perceived Impact of TOEIC on Teacher Teaching (N=81).....	113
Table 5.4 Descriptive Statistics on Teachers’ Beliefs about Language Teaching and Learning (N=81)	115
Table 5.5 Descriptive Statistics on Teachers’ Perceptions of TOEIC (N=81)	115
Table 5.6 Descriptive Statistics on Teachers’ Perceptions of Context Factors (N=81)	116
Table 5.7 Statistics on Correlations between Person Factors and Perceived Impacts of TOEIC on Teaching (N=81)	119
Table 5.8 Statistics on Correlations between Perceptions of TOEIC and Its Perceived Impacts on Teaching (N=81)	121
Table 5.9 Statistics on Correlations between Perceptions of Context Factors and Perceived Impacts of TOEIC on Teaching (N=81)	124
Table 5.10 Statistics on Differences in Perceived Impact of TOEIC on Teaching between Large City Universities and Provincial University	125
Table 5.11 Statistics on Differences in Teachers’ Perceptions of TOEIC between Large City Universities and Provincial University	126
Table 5.12 Statistics on Differences in Teachers’ Beliefs about Language Teaching and Learning between Large City Universities and Provincial University	127
Table 5.13 Statistics on Differences in Teachers’ Perceptions of Context Factors between the Large City Universities and Provincial University	128
Table 6.1 Demographic Information of Student Participants (N=1445)	132
Table 6.2 Descriptive Statistics on Students’ Learning Experience (N=1445)	132

Table 6.3 Descriptive Statistics on Perceived Impact of TOEIC on Student Learning (N=1445)	133
Table 6.4 Descriptive Statistics on Students' Perceptions of TOEIC (N=1445).....	134
Table 6.5 Descriptive Statistics on Students' Beliefs about Language Learning (N=1445)	136
Table 6.6 Descriptive Statistics on Students' Perceptions of Context Factors (N=1445).	136
Table 6.7 Statistics on Correlations between Person Factors and Perceived Impacts of TOEIC on Student Learning (N=1445)	139
Table 6.8 Statistics on Correlations between Perceptions of TOEIC and Its Perceived Impacts on Learning (N=1445)	142
Table 6.9 Statistics on Correlations between Perceptions of Context Factors and Perceived Impacts of TOEIC on Student Learning (N=1445)	144
Table 6.10 Statistics on Differences in Perceived Impact of TOEIC on Student Learning between Large City Universities and provincial University.....	146
Table 6.11 Statistics on Differences in Students' Perceptions of TOEIC between Large City Universities and Provincial University	148
Table 6.12 Statistics on Differences in Students' Beliefs about Language Learning between Large City Universities and Provincial University	150
Table 6.13 Statistics on Differences in Students' Perceptions of Context Factors between Large City University and Provincial University	152
Table 7.1 Demographic Information of Teacher Cases.....	161
Table 8.1 Demographic Information of Student Cases (N= 12).....	200
Table 8.2 Time Khoi Spent on Language Skills.....	211
Table 8.3 Time Thien Spent on Language Skills	220
Table 8.4 Time Toan Spent on Language Skills.....	230

List of Figures

Figure 1 Washback on the Processes and Products of Learner Test Activity (Booth, 2012, p. 292).....	34
Figure 2 A basic Model of Washback (Bailey, 1996, p. 264)	38
Figure 3 Green’s Model of Washback Direction, Variability, and Intensity (Green, 2006, p. 340).....	39
Figure 4 A Washback Model of Students’ Learning (Shih, 2007, p. 151).....	41
Figure 5. Overall Research Design.....	49
Figure 6 Relationship among the TOEIC-as-exit-test, Course Assessment, Textbooks, and Teaching	197

Chapter 1: INTRODUCTION

1.1 Rationales and Background

1.1.1 Rationales for study.

In her discussion of the power of tests, Shohamy (2001b) wrote: “Tests can affect and manipulate educational and political systems, control curricula and redefine knowledge of communities” (p. 18). It is because of this power to change the behaviour of individuals and educational systems that high-stakes tests have been increasingly used by central authorities for educational improvement. This is clearly evident in many countries around the world (Andrews, Fullilove, & Wong, 2002; Au, 2007; Brindley, 2001; D. W. Chapman & Snyder Jr, 2000; Shih, 2007). However, there is fierce debate about the effectiveness of the use of high-stakes tests to drive instruction. The opponents of high-stakes testing policies point out the great pressure on teachers and students, the narrowing of curriculum, the detrimental impact on teaching methods as a result of this use of tests, and the use of tests for unfair purposes (Madaus, 1985; Shohamy, 1997). On the contrary, the advocates of the policies believe that it is possible to use tests to bring about beneficial effects if appropriately implemented. Some even consider using high-stakes tests to drive education as the most cost-effective way to improve instruction (Popham, 1987). Therefore, together with the increasing use of high-stakes tests as levers for change, there has been a growing interest in research on test impact on teaching and learning or washback, a term often used in language education. In accordance with this research trend, this study examined the impact of the TOEIC used as a language exit test in many universities in Vietnam.

There are several rationales for the study. Research on test impact on teaching and learning is a relatively new area in language education. Empirical evidence of this phenomenon started to accumulate only in the 1990s, especially after Alderson and Wall (1993) posed a question about its existence and suggested an agenda for test impact research. Research has now not only proved the existence of test impact, but also revealed areas of teaching and learning that might be affected by testing. Research has also revealed the factors that influence the form and intensity of test impact. However, “its existence raises more questions than it answers” (Alderson, 2004, p. xii). Findings from the research literature are varied from context to context, and the ways tests influence teaching and learning are not simple to explain. Some studies have found strong effects of tests on teachers’ and

students' feelings and attitudes and on the content and methods of their teaching/learning (Cruz & Brown, 2010; Shohamy, 1993; M. L. Smith, 1991). Others have shown that these effects are only minimal (Cheng, 1998; Shih, 2007). Also, test impact might be strong on certain areas of teaching and learning, but weak on others (Cheng, 1997; Wall & Alderson, 1993). In addition, these effects could be either positive or negative, but it is wrong to assume that a good test will produce beneficial effects, and that a bad test will produce detrimental effects (Wall, 2000). This variability and elusiveness of test impact has been attributed to the fact that a myriad of context-specific factors are involved in its formation. Influencing factors might include test stakes, test status, test design, teacher's beliefs, teacher's teaching styles, students' ability, and school settings. As test impact is context-specific, the list of factors does not seem exhaustive. In each context, different factors come into play and/or contribute differently to the shaping of test impact on teaching and learning. The way they interact with each other, therefore, also differs. For this reason, many language testers (Spratt, 2005; Tsagari, 2007) have called for more studies on test impact to understand how it operates in different contexts.

Also, while effects of tests on teaching have received a lot of attention from researchers (Alderson & Hamp-Lyons, 1996; Manjarrés, 2005; Watanabe, 1996), their impact on learning has still been very little explored. Research has shown that test impact on learning is also complex and involves numerous factors like test impact on teaching (Shih, 2007), but not many studies have investigated the mechanism by which it operates. The current study is also an attempt to help fill in this gap in research literature.

Additionally, TOEIC is one of the major English tests which is being used by over 10,000 companies, government agencies, and English Language Learning programmes in 120 countries (ETS, n.d.-b), but its impact has been explored by only a surprisingly limited number of studies (Apichatrojanakul, 2011; Newfields, 2005; Nhan, 2013) compared to TOEFL or IELTS. With a design very different from the TOEFL or IELTS, TOEIC is expected to have very different impact on teaching and learning. Therefore, it is vital to examine the impact of the test to get an insight into how differently different tests might affect teachers and learners.

Furthermore, there are important practical rationales for a study on the effects of the TOEIC test on English teaching and learning in the Vietnamese context. The test has recently been used as a language exit test in many universities to drive teaching and

learning (Hồng Ninh, 2010). However, concerns have been raised about this use of the test (Nhu Thuận, 2008). Some experts and teachers doubt its suitability for Vietnamese universities because the TOEIC Listening and Reading that universities are using has been designed based on an outdated theory of language ability and assesses only the two receptive skills – listening and reading. They are worried that teachers will only show students how to get high scores rather than teach the four skills, and the content of teaching will also be narrowed. They also say that Vietnam should learn a lesson from Japan and Korea, where a huge amount of money has been spent on the TOEIC test, but people's English is still not good enough especially when compared to Singapore and Malaysia. Despite the ambition of, and controversy over, the TOEIC-as-exit-test policy, or TOEIC policy for short, no robust research has been done to examine its actual impact on Vietnamese education and society. Given the high stakes of the policy for language education, research is urgently needed to evaluate its impact in the teaching and learning context. Findings from such research would have important practical implications for stakeholders including policymakers, teachers, students, and the wider society. This lack of research has become a major motive for the current study.

1.1.2 Vietnamese context and TOEIC-as-exit-test policy.

This study takes place in Vietnam, a developing country with a fast growing economy. Due to the globalization and integration of economies, the demand for foreign language competence, especially English competence, has become increasingly high in this country. English has a special status in society because of the economic advantages it can give to individuals and the country. English teaching, therefore, has been given top priority in the government's educational policy. This is clearly manifested in the National Foreign Languages 2020 Project (also called the 2020 Project) approved by the Prime Minister in 2008 (Government, 2008). The goal of the project is to comprehensively renovate foreign language teaching and learning in the national education system so that

by 2020 most Vietnamese youths graduating from vocational schools, colleges, and universities will have been able to use a foreign language independently and confidently to communicate, study, and work in an integrated, multilingual, and multicultural environment, making foreign languages an advantage for Vietnamese people in the cause of industrialization and modernization of the country. (p. 1)

(Translated by the researcher).

The language exit test policy at the tertiary education level partially reflects this ambition. A few universities introduced the TOEIC test to their language programmes around 2006; more and more universities adopted the same policy later, especially around 2010, when the Ministry of Education and Training (MOET) required universities to publicize their learning outcomes for all the subjects they taught. For most of them, the policy was initially a response to MOET's requirement concerning their learning outcomes. MOET did not require universities to use a language exit test or specify any tests for learning outcomes assessment (although in 2005, they mentioned a guiding principle for using TOEIC to establish benchmarks for graduates' English ability (Thanh Hà, 2005)). However, with a concern about their graduates' low language ability, many universities started to use a language test for gate keeping purposes. The majority of them chose TOEIC. On top of passing the universities' English course assessment, students are required to take the TOEIC test and achieve a certain TOEIC score called TOEIC benchmark before the end of their studies. If they fail to obtain the required score, they will not be able to graduate. Generally, those students who do not meet the language exit requirement have to continue preparing for the test themselves and submit the required score within two years of the end of their studies. If they cannot submit it within that time period, their bachelor's degrees, in principle, will be cancelled. Motivating students and making them meet the requirements of the labour market were the most common rationales for the TOEIC-as-exit-test policy mentioned by the policymakers who were interviewed by the researcher. As a bottom-up initiative, the policy varies a lot across universities. Some universities require their students to pass the international TOEIC test designed and administered by Educational Testing Service. Others just administer a simulated TOEIC test, which is usually called the internal TOEIC, to their students (Minh Giảng, 2012). The internal TOEIC test is compiled by universities using test preparation materials available on the market. The TOEIC benchmark also differs among universities, but it is usually in the range from 400 to 550.

1.1.3 The TOEIC test.

The TOEIC test was developed by Educational Testing Service (ETS) in response to the request of the Japanese Ministry of Trade and Industry and first launched in Japan in 1979 (M. Chapman, 2004). According to ETS, the test aims at assessing English-language skills in the workplace (ETS, n.d.-a); and it can be specifically used “to recruit, place and promote the most qualified employees”; “identify job-training requirements”; “assign

employees to positions overseas” (ETS, n.d.-c). For approximately the first 30 years of its existence, the test assessed only the two receptive language skills – listening and reading. The TOEIC Listening and Reading is a two-hour multiple choice test. It includes two sections – listening and reading, each section containing 100 questions. The listening section is composed of four parts and lasts approximately 45 minutes; the reading section consists of three parts and is approximately 75 minutes long. The scaled score for each section ranges from 5 to 495. The following are the question types for each part of the original TOEIC test.

Part 1 – Photographs: For each question, test-takers hear four statements about a picture in their test book and then select one statement that best describes the picture.

Part 2 – Question-Response: For each question, test-takers hear a question or a statement and three responses, and they have to select the best response to the question or the statement.

Part 3 – Short Conversations: For each question, test-takers hear a conversation between two people and then answer a question about what the people say in the conversation.

Part 4 – Talks: Test-takers hear some talks given by a single speaker, and they have to answer two to four questions about what the speaker says in each talk.

Part 5 – Incomplete Sentences: For each question, test-takers read a sentence with a missing word or phrase, and they have to select the best answer among the four answer choices given below each sentence to complete it.

Part 6 – Error Recognition: For each question, test-takers read a sentence with four words or phrases underlined, and they have to select the word or phrase that is incorrect.

Part 7 – Reading Comprehension: Test-takers read a number of texts including magazine and newspaper articles, letters, and advertisements, each of which is followed by several questions. Test-takers have to select the best answer to each question.

Table 1.1 Original and Redesigned TOEIC Listening and Reading Test Format

ORIGINAL TOEIC	REDESIGNED TOEIC
Section 1: Listening Comprehension: 100 items (45 minutes)	Section 1: Listening Comprehension: 100 items (45 minutes)
Part 1: Photographs: 20 questions	Part 1: Photographs: 10 questions
Part 2: Question-response: 30 questions	Part 2: Question-response: 30 questions
Part 3: Short Conversations: 30 questions; 30 conversations with one question each	Part 3: Short Conversations: 30 questions; 10 conversations with 3 questions each
Part 4: Talks: 20 questions	Part 4: Talks: 30 questions; 10 talks with 3 questions each
ORIGINAL TOEIC	REDESIGNED TOEIC
Section 2: Reading Comprehension: 100 items (75 minutes)	Section 2: Reading Comprehension: 100 items (75 minutes)
Part 5: Incomplete Sentences: 40 questions	Part 5: Incomplete Sentences: 40 questions
Part 6: Error Recognition: 20 questions	Part 6: Text Completion: 12 questions
Part 7: Reading Comprehension: 40 questions	Part 7: Reading Comprehension: - <i>Single Passages</i> : 28 questions; 7-10 reading texts with 2-5 questions each - <i>Double Passages</i> : 20 questions; 4 pairs of reading texts with 5 questions per pair

Source: TOEIC Examinee Handbook (ETS, 2012b)

In 2006, the TOEIC Listening and Reading underwent some changes to be “align[ed] more closely with theories of communicative competence” (Powers, Kim, & Weng, 2008, p. 1). Powers et al. (2008) detailed these changes in their report on the correlations between the redesigned TOEIC test scores and test-taker perceptions of proficiency in English. The most significant changes in the listening section include the use of a variety of English accents such as North American, British, Australian, and New Zealand accents to reflect international business communication styles and the increase of sets of questions in Part 3. As for the reading section, error recognition questions were removed, and text completion questions were added. Interrelated passages that require test-takers to connect information in order to be able to answer questions were also introduced to the test. (For the differences between the original TOEIC Listening and Reading and the redesigned one, see Table 1.1.) With these modifications, TOEIC seems to be moving towards an assessment of the textual level of language and becoming a more authentic test. However, it still has the same multiple choice format, and most of the question types remain the same. Therefore, it

is still being criticized for being “based on the structuralist, behaviorist model of language learning and testing that informed discrete-point testing”; not assessing essential aspects of real-life listening comprehension such as speech acts and pragmatic implications; and lacking natural hesitations, fast speech, phonological shifts, and negotiations of meaning between interlocutors (M. Chapman, 2005, p. 1192).

Also in 2006, the TOEIC Speaking and Writing tests were introduced to complement the TOEIC Listening and Reading test. However, these components are only optional. They are delivered through the computer, and test-takers’ responses are sent to ETS’s Online Scoring Network to be scored. The speaking test consists of 20 questions (six task types) and takes approximately 20 minutes to complete. The writing test has eight questions (three task types) and lasts approximately one hour. (For the speaking and writing task types, see Table 1.2.) The speaking and writing test scores are reported separately. The scaled score for each of the tests can be in the range from 0 to 200.

Table 1.2 TOEIC Speaking and Writing Test Format

TOEIC Speaking Test (20 minutes)		TOEIC Writing Test (1 hour)	
Question	Task	Question	Task
1-2	Read a text aloud	1-5	Write a sentence based on a picture
3	Describe a picture	6-7	Respond to a written request
4-6	Response to questions	8	Write an essay
7-9	Response to questions using information provided		
10	Propose a solution		
11	Express an opinion		

Source: Examinee Handbook – Speaking and Writing (ETS, 2012a)

Although the TOEIC test has become widely used in the world, not much research has been done on its quality as compared to TOEFL and IELTS. Some early research reports, which were published on ETS website, found that the TOEIC test was valid and reliable. One of these is Woodford’s (1982). Examining the data from a group of Japanese test-takers, who sat the test in 1979, Woodford reported that the TOEIC test had very high reliability (the reliability was above .9 for the Listening Comprehension Section, Reading

Section, and the whole test). In the same paper, Woodford also reported his study on the validity of TOEIC. He found high correlations between the TOEIC test scores and TOEFL test scores as well as between the TOEIC test scores and direct measures of the listening, reading, speaking, and writing skills. He concluded that “TOEIC provides a good indication of candidates’ language ability in English” (p. 16). Similarly, Wilson (1989) found strong concurrent correlations between TOEIC test scores and the ratings of oral English proficiency (using the Language Proficiency Interview). He also concluded that valid inferences could be made about levels of oral English proficiency. These studies seemed to suggest that it was not necessary to use direct measures to assess test-takers’ overall language proficiency. However, some independent researchers found results which were somewhat different from what Woodford (1982) and Wilson (1989) suggested. Hirai (2002), for example, showed that the overall correlation coefficient between the TOEIC scores of the intermediate and advanced students and the Hitachi’s interview test scores was fairly good (0.78), but when the two groups of students were examined separately, the correlations between the TOEIC total scores and the interview scores dropped significantly (to 0.49) for the intermediate group. Also, TOEIC scores had a correlation of only 0.66 with the BULATS Writing Test, a well-established business language test. This correlation was lower than ETS’s findings. Hirai explained the differences in the findings between his and ETS’s studies by the different nature of the writing tasks used. He suggested that “TOEIC scores be interpreted cautiously by businesses. To assess business writing skills, an exam designed with a typical business environment in mind such as the BULATS Writing Test is recommended.” Another researcher, Cunningham (2002), examined the relationship between TOEIC test scores and the scores from a direct assessment of reading, listening, and writing. She found no correlation between the TOEIC test scores and communicative abilities. Cunningham’s findings should be received with caution, though, because communicative abilities were measured by her own test rather than a test with well-established reliability and validity. Nevertheless, the most recent studies published on ETS website after the TOEIC Speaking and Writing tests were introduced tended to support Hirai’s (2002) and Cunningham’s (2002) findings. Liao, Qu, and Morgan (2010), who examined the relationship among the skills measured by the TOEIC Listening and Reading and TOEIC Speaking and Writing, found that the test scores only moderately correlated with one another and came to a conclusion that “each test measures distinct aspects of English language proficiency that cannot be adequately assessed by the other

tests” (p.13.11). A similar study carried out later by Liu and Costanzo (2013) also yielded the same results. ETS now advises that test-takers should take all four skill tests in order to have an accurate, comprehensive understanding of their language ability.

Although the TOEIC Speaking and Writing is now available, universities in Vietnam are still using the TOEIC Listening and Reading, which is referred to in this thesis as TOEIC or the TOEIC test, as their language exit test. Being used for a high-stakes purpose in a social setting where English enjoys a high status, the TOEIC test might bring about substantial changes to English teaching and learning in Vietnamese universities.

1.2 Purpose of Study and Research Design

In response to language testers’ calls for research on test impact in new contexts to better understand its complexity and elusiveness, this study has the purpose of exploring the impact of the TOEIC test used as a language exit test in Vietnamese universities. It examines the effects of the test on both teaching and learning taking into account the perspectives of three groups of stakeholders: policymakers, teachers, and students. Teachers and students are included into the study because they are directly involved in the teaching and learning process. Policymakers including MOET officials and Heads of Department (HODs) are also considered key stakeholders in the research context because they have important roles in the adoption and implementation of the policy. From the media (Thanh Hà, 2005; Trung Toàn, 2008), it is clear that MOET supports and encourages universities’ TOEIC policy. Heads of Department are always responsible for academic affairs, i.e., they decide on issues closely related to teaching such as syllabuses, textbooks, and assessment. Also, having been a university teacher in Vietnam for a long time, the researcher believes that in order for the TOEIC policy to be adopted, there has been a consensus between the universities’ top leaders and the HODs. Therefore, HODs are supposed to be able to provide the researcher with valuable information on the TOEIC policy. Knowing MOET officials’ and HODs’ perspectives would help the researcher to better explain the impact of the TOEIC test on teaching and learning and evaluate whether the TOEIC policy has achieved its intended impact or not, and whether there is any unintended impact.

As school settings have shown to be an important factor in shaping test impact (McCarthy, 2008), but still have been little explored, the study also puts an emphasis on discovering

the differences in test impact on teaching and learning due to school settings. Expecting big differences between large cities and provinces in teaching and learning resources and teacher and student population, the study is conducted at two different locations, a large city and a province, in order to better understand the mechanism of test impact.

The study is conducted in two phases at three large universities in Vietnam. Two of them are located in a large city, and the other is situated in a province. To meet the research objectives, a mixed-methods research design is utilized. Phase 1 of the study aims at identifying general patterns of the influence of the TOEIC test on teaching and learning, factors involved in the generation of the test impact, and the policymakers' perspectives on the TOEIC-as-exit-test policy. Two sets of questionnaires are administered to teachers and students at the three universities mentioned above, and semi-structured interviews are carried out with two MOET officials and twelve Heads of Department at various universities in Vietnam. The purpose of Phase 2 is to explain the patterns found in Phase 1 and to gain deeper insight into the mechanism of test impact. A case study is conducted with seven teachers and twelve students at the same three universities.

1.3 Structure of Thesis

This chapter has provided the background to this study, the rationales for it, its purpose, its research design, and its significance.

Chapter 2 provides a framework for the study by reviewing research literature on test impact on teaching and learning. It begins with the definitions of the common terms used in research literature to refer to the effects of tests in the teaching and learning context and beyond this context. Next, it presents the five dimensions of test impact. Then, it reviews research findings about the effects of tests on different areas of teaching and learning and the mechanism of test impact. Finally, it gives an overview of the use of high-stakes tests to drive instruction.

Chapter 3 explains the methods to achieve the research purpose. It starts with the research questions guiding the study. Then it presents the overall research design as well as the rationales for it. Next, it describes the participants and research sites. After that, it discusses reliability and validity in both quantitative and qualitative research, and reports on the strategies used in this study to ensure these important research qualities. This is followed by a description of Phase 1 and Phase 2 methodology.

Chapter 4 presents the results of the analysis of the interviews with the policymakers with regard to their perspectives on the purpose of language teaching and learning at university, their rationales for choosing TOEIC as a university language exit test, and their intentions behind the TOEIC-as-exit-test policy.

Chapters 5 and 6 report findings from the teacher and student questionnaires, respectively. These chapters have the same structure. They begin with the findings on the patterns of test impact on teaching/learning. Then they report the measurements of factors potentially involved in test impact generation. After that, they present the results of the bivariate correlation analysis between the test impacts and the potentially influencing factors, which showed the extent to which each factor was responsible for the shaping of the test impact. Finally, they describe the differences in test impact and factors involved between the large city universities and the provincial one.

Chapters 7 and 8 report the findings from the teacher and student case studies, respectively. They also have the same structure. They first give a summary of the quantitative analysis of the interviews with participants, then report the findings from the classroom observations (in the case of the teachers) or journal sheets (in the case of the students), and finally present the narratives of three selected teachers and students.

Chapter 9 summarises the research findings and discusses the major issues emerging from the findings. The issues include the intended and actual impact and positive and negative impact of the TOEIC test on teaching and learning, the discrepancy in the perceptions of the test validity between the policymaking level and the implementation levels (teachers and students), the differences in test impact between the large city and the province, the mechanism of test impact on teaching and learning in the Vietnamese context, and an evaluation of the TOEIC-as-exit-test policy.

Chapter 10 summarises the study, discusses its limitations, and presents the implications it has for different groups of stakeholders. It also makes recommendations for future research.

Chapter 2: LITERATURE REVIEW

2.1 Definitions of Washback and Impact

When referring to the effects of tests, language testers usually use two different terms: *washback* and *impact*. Washback is commonly understood as the influence of testing on teaching and learning (Alderson & Wall, 1993; Bailey, 1996; Cheng, Watanabe, & Curtis, 2004; Hamp-Lyons, 1997; Messick, 1996; Tsagari, 2007). However, there is not a unanimous understanding of impact. For some language testers, impact is a broader construct referring to “any of the effects that tests may have on individuals, policies or practices, within the classroom, the school, the educational system, or society as a whole” (Tsagari, 2007, p. 4), and thus washback is only one of its dimensions. Other language testers distinguish washback and impact as micro and macro effects of testing within society with the term washback referring to the effects of tests in the teaching/learning context and the term impact referring to the effects beyond the teaching and learning context (Taylor, 2005). In general education, however, only the term impact is used. In this study, impact is used as an umbrella term including washback.

2.2 Dimensions of Washback

The term *dimensions* was used by Watanabe (2004) to refer to aspects of washback. He described the nature of washback in five dimensions: specificity, intensity, lengths, intentionality, and value.

In terms of *specificity*, washback can be general or specific. “General washback means a type of effect that may be produced by any test” (Watanabe, 2004, p. 20). In accordance with this understanding, Alderson’s and Wall’s hypothesis (1993) that a test will influence what teachers teach can be considered as referring to general washback. On the other hand, specific washback “refers to a type of washback that relates to only one specific aspect of a test or one specific test type” (Watanabe, 2004, p. 20). Thus, a finding that a multiple-choice test does not encourage learners to learn productive language skills, for instance, relates to specific washback.

Intensity, which is synonymous with the term *extent* used by Bachman and Palmer (1996), describes the strength of washback. It was first coined by Cheng (1997) “to refer to the

degree of washback effect in an area or a number of areas of teaching and learning affected by an examination” (p. 43). Watanabe (2004) later elaborated on this term as follows:

If the test has a strong effect, then it will determine everything that happens in the classroom, and lead all teachers to teach in the same way toward the exams. On the other hand, if a test has a weak effect, then it will affect only a part of the classroom events, or only some teachers and students, but not others (p. 20).

Thus, the impact of the EFL test and the ASL test in Shohamy et al’s study (1996) described below in Section 2.3 can be classified as strong and weak, respectively.

Length is used to describe how long the washback of a test lasts (Watanabe, 2004). Again, the examples from Shohamy et al’s study (1996) can perfectly illustrate this aspect of washback. The effect of the ASL existed only before it was first administered, so this effect is short-term. However, the washback of the EFL exam still persisted many years after its introduction, so this is long-term washback.

Washback can be intended or unintended. To the best of my knowledge, language testers, including Watanabe, have not given an official definition for the term *intentionality* yet. However, a number of them have acknowledged and discussed this aspect of washback in their papers (Andrews, 2004; Messick, 1996; Tsagari, 2007). For example, arguing that validity of performance tests could not be taken for granted, many authors (Linn, Baker, & Dunbar, 1991; McNamara, 1996; Messick, 1994, 1996) stressed the need to examine both their intended and unintended consequences on teaching and learning.

Watanabe’s term *value* is equivalent to *direction*, which is used by many other authors (Alderson & Wall, 1993; Bailey & Masuhara, 2013; Green, 2007a; Tsagari, 2007). Washback may be positive or negative. According to Bailey and Masuhara (2013), the judgement of washback as positive or negative depends “on our view of the desirable outcomes of language learning” (p. 304). They illustrated their point with an example of a test with an emphasis on recognition and interpretation of vocabulary items in reading and listening passages. Suppose that this test encourages students to learn academic vocabulary. If we think it is especially important for university-bound students to have a large capacity of receptive academic English vocabulary, then the washback effects of the test are positive. On the contrary, if we believe our goal of language teaching is students’ oral communicative ability, then we might see the influence of the test as negative. For the same reason, Watanabe (2004) also suggested that the judgement of the value of a test

should be made with reference to specific audience because “one type of outcome may be evaluated as being positive by teachers, whereas the same outcome may be judged to be negative by school principals” (p. 21).

According to Weigle and Jensen (as cited in Bailey & Masuhara, 2013), in order for a test to have positive washback on teaching and learning, there should be a match between instructional goals and activities and assessment procedures. Similarly, Green (2006) pointed out the “congruence between test design and skills developed by a curriculum or required in a target language use domain” as a necessary condition for positive washback (p. 339).

Watanabe (2004) also discussed the association between intentionality and value of washback. He argued that because test writers certainly do not intend to cause negative washback, “intended washback may normally be associated with positive washback, while unintended washback is related to both negative and positive washback” (p. 21).

2.3 Effects of Tests on Teaching and Learning

Test impact on teaching and learning has been a research interest in general education for a long time, however, in language education, empirical evidence of the phenomenon only started to flourish in the 1990s, especially after Alderson and Wall (1993) posed their famous question “Does washback exist?” and proposed an agenda for future research. Since then a considerable number of studies in language education have been done to seek empirical evidence for the widespread belief that tests have impact on teaching and learning. In this part of the review, I summarize the most important research findings in both language education and general education with regard to the areas affected by washback, namely teachers’ and learners’ feelings and attitudes, their behaviours, and the consequences of their behaviours.

2.3.1 Effects of tests on teachers’ and learners’ feelings and attitudes.

It seems that researchers have found more negative effects of tests on teachers’ and learners’ feelings and attitudes than positive ones. Many studies (Alderson & Hamp-Lyons, 1996; Choi, 2008; Cruz & Brown, 2010; McCarthey, 2008; Shohamy et al., 1996; M. L. Smith, 1991) show that teachers and students experience enormous pressure from high-stakes testing and think unfavourably of tests. However, some studies (Cheng, 1997; Marriott, 2009; McCarthey, 2008; Parke, Lane, & Stone, 2006; Read & Hayes, 2003;

Shohamy et al., 1996) report no impact or even positive impact on teachers' and students' feelings and attitudes.

There are some interpretations of these differences in the research results. It is very interesting to examine Shohamy et al.'s and McCarthy's findings because they showed contradictory pictures within the same study. In Shohamy et al.'s study (1996), the teachers did not attach any importance to the low-stakes Arabic as a second language (ASL) test. Although many students thought the test was important, they were not reported to experience anxiety due to the test. In contrast, both the teachers and students suffered from high anxiety and fear of the results of the high-stakes English as a foreign language (EFL) test. Teachers "feel that the success or failure of the students reflects on them and they speak of the pressure to cover the material for the exam" (pp. 309–310). Most of the students believed the test results could affect their overall matriculation score and their success in their future studies. Clearly, the stakes of the test and the status of the language tested make the difference. In McCarthy's study (2008), the teachers in high-income schools generally "felt safe in their school settings and confident that their students would succeed, almost without their intervention" (p. 485), whereas the teachers in low-income schools experienced tremendous pressure to raise their students' test scores; otherwise, the schools would be closed and they would lose their jobs. Here different school settings resulted in different 'de facto' stakes of the test for the two groups of teachers, which eventually caused the variability in test impact on them. It is interesting to note that similar variability in test impact were found in the two studies, but it was produced in different ways. In the former, test stakes seemed to have more direct influence on the participants' behaviour, but in the latter the effects of stakes were moderated by the school settings.

Another factor that causes variability in teachers' and learners' attitudes is probably the degree of congruence between the test characteristics and teachers' and learners' beliefs about what teaching, learning and testing should be (Alderson & Hamp-Lyons, 1996; Choi, 2008; Parke et al., 2006; Read & Hayes, 2003; Shohamy, 2001a). The teachers in Alderson and Hamp-Lyons' study (1996) had negative attitudes towards TOEFL because "they considered it inauthentic and noncommunicative" (p. 285). They also disliked teaching TOEFL because it was boring and fragmentary. In contrast, the teachers in Read and Hayes' (2003) study had favourable attitudes towards IELTS (compared to TOEFL) because they thought it included the kind of tasks and language that learners would need

for their academic studies, and IELTS scores well reflected test-takers' ability to cope with the academic language requirements in New Zealand. Students' age can also be a factor. Scott's study (2007) showed that the difference in the students' age led to different perceptions of test consequences and therefore different perceptions of pressure of tests. It is clear from the review above that tests can have various emotional impacts on teachers and students depending on different factors involved.

2.3.2 Effects of tests on teaching.

2.3.2.1 Effects of tests on content of teaching.

The content of teaching is probably an area most susceptible to change. A lot of studies (Cheng, 1997; Chen, 2002b; Diamond, 2007; Green, 2006; Saif, 2006; Wall & Alderson, 1993) have revealed that tests influence the content of teaching more than teaching methods. Effects of tests on the content of teaching include the changes in teaching materials and curriculums – official and de facto. Generally, a change in examinations was often accompanied by a change in the official curriculum and textbooks (Cheng, 1997; Chen, 2002b; Saif, 2006) (In some cases, examinations were changed later to match the new textbooks (Wall & Alderson, 1993)). In Cheng's study (1997), for example, following the introduction of the revised HKCEE, new materials were quickly published and adopted: "By the time the examination syllabus affected teaching in Hong Kong secondary schools..., nearly every school had changed their textbooks for the students. Almost all textbooks are labelled specifically 'For The New Certificate Syllabus'... Publishers in Hong Kong worked really hard and quickly to get textbooks ready for the schools" (p. 50). A lot of commercial test preparation materials were also published and used by teachers to prepare students for the new tests (Cheng, 1999; Shohamy et al., 1996; Tsagari, 2011). Many studies (Cheng, 1997; Chen, 2002b) show that teachers relied heavily on new textbooks. They also made an extensive use of exam-related materials (Andrews et al., 2002; Hwang, 2003; Nunan, Berry, & Berry, 1994; O, 2001; Read & Hayes, 2003; Tsagari, 2011). Investigating the impact of the EU oral on English teaching in Hong Kong secondary schools, Andrews (1994) discovered that "an estimated two thirds of class-time is spent working with exam-focused published materials" (p. 78). Through a national survey of language schools in New Zealand in 2000, Read and Hayes (2003) found that 90% of the 60 schools that offered IELTS preparation courses usually used commercial test

preparation materials compared to only about 30% of the schools (19) that usually used materials locally and specifically written for their own programmes.

A new test can also create what Shohamy (1997, 1998) calls ‘de facto curriculum’. When tests are used as levers for change or for accountability purposes, they make teachers “teach to the test” and consequently create a narrower curriculum overriding the official one. This influence of tests on the content of teaching is obvious in a large number of test impact studies (Bekis, 2008; Collins, Reiss, & Stobart, 2010; Cruz & Brown, 2010; Faulkner & Cook, 2006; Green, 2006; Holme, Richards, Jimerson, & Cohen, 2010; Hwang, 2003; O, 2001; Quesada Inces, 2001). Hwang (2003), for instance, found negative impact of CSAT, a university entrance exam, on teaching and learning in secondary schools in South Korea. Both general and vocational school teachers in his study “skip over most of the speaking and writing activities in the textbook such as ‘speak’ ‘role-play’, and ‘writing practice’ sections, for they think these parts are not relevant to the CSAT” (p. 77). Both groups, however, covered the listening and reading parts, although to different extents. Hwang argued that this was a result of the mismatch between CSAT and the textbooks and the official curriculum. Similarly, although Costa Rica’s curriculum advocated communicative language teaching, the Bachillerato English test, which assessed only the reading skill, made teachers and students focus overwhelmingly on it while ignoring the other skills. Thus, teachers and students started studying topics with reading, their readings only covered the topics found in the test, and classroom activities aimed at developing reading comprehension skills with explicit attention to grammar and vocabulary only (Quesada Inces, 2001).

A narrowed-down curriculum could result even from an exam which embraces a communicative language teaching philosophy which incorporates all four language skills. The impact of the FCE on English teaching in Greek schools is an example (Tsagari, 2011). On the one hand, the teachers in the study put a lot of emphasis on grammar and vocabulary because they thought they were vital for success in the exam. They paid special attention to the reading and writing tasks that were tested by the exam. On the other hand, they spent less time on listening and speaking and neglected pronunciation. According to Tsagari, this type of narrowing of curriculum could be in part explained by the test design, e.g., the light weight of listening, the nature and marking of the speaking skill. It could also be a result of teachers’ lack of understanding of the underlying principles of the exams,

their lack of training, and the teaching practices in the mainstream Greek schools, which involved teaching grammar, vocabulary and dictation. Again, the relationship between the test design/nature and the content of teaching is not always straightforward because it was moderated by other factors.

However, test impact on the content of teaching does not always occur. Although teachers' efforts to align their own curriculums with the test design have been evident in a great deal of research, some studies found no such effects on teaching. These contrasting pictures can be vividly seen in Shohamy et al's study (1996). The impact of two national tests – Arabic as a second language (ASL) and English as a foreign language (EFL), was re-examined after they had been in effect for several years. While the impact of the EFL oral test significantly increased, "the impact of the ASL test has decreased over the years to the point where it has no effect" (p. 312). Teachers no longer gave students long lists of words to memorize, taught to the test, reviewed the materials to be tested, prepared special materials for the test, and used past tests to drill their students and get them familiar with the test format. They did not even inform students of the upcoming test. In fact, there were no test preparation activities at all. The big differences in the impact of the two tests were explained by their stakes and the status of the language they assessed. The EFL was a high-stakes test, whose results would affect students' graduation from high school and entrance to tertiary institutions. The ASL, by contrast, was a low-stakes one. Over the years, teachers had realized that its results had no personal consequences on them, so they had become "relaxed and fearless" (p. 314). Also, English enjoyed a high status in Israeli society. It is considered "the most honoured linguistic resource besides Hebrew," "a power asset and a boundary marker," "a means of gaining socioeconomic status" (p. 315). The Arabic language had a much lowered status compared to English. It was the language of the lower classes in Israeli society and the countries that Israel was at war with, and it was not useful for gaining employment, academic achievements, and social status. Watanabe's (2000) study also showed that teachers did not necessarily taught to the target exam (TE). For example, they did not teach writing even though the TE covered this skill because they did not know the scoring methods or the scoring criteria. They intentionally avoided teaching test-taking skills because they believed the actual language skills would enable students to pass the exam. On the other hand, those who believed in the effectiveness of translation still used it even though it was not included in the TE, and those who believed

in the necessity of grammar still frequently referred to formal aspects of English and used metalanguage irrespective of the TE.

Again, the variability of washback on the content of teaching can be explained by a number of factors. In both Cheng's (1997) and Chen's (2002b) studies, the teaching content changes were due to the fact that teachers tend to rely on widely available new textbooks. Teachers in Taiwan taught listening and speaking because these skills were included in the new textbook issued by the Ministry of Education, which they often adopted as a curriculum (Chen, 2002b). The teachers in Shohamy's study used no exam-related materials at all, certainly because of the low stakes of the ASL test. In Watanabe's (2000) study, the lack of either positive or negative washback was explained by a number of factors such as the lack of test specifications (scoring methods, keys, etc.), teachers' sense of guilt about test coaching, and teachers' beliefs about the best teaching method.

2.3.2.2 Effects of tests on teaching methods.

The effects of tests on teaching methods must be the most complex and most difficult to explain. Findings vary on a continuum from "no effects" (Qi, 2007; Shohamy et al., 1996; Wall & Alderson, 1993), "superficial effects" (Cheng, 1997, 1999, 2003; Chen, 2002a) to "strong effects" (Gerwin & Visone, 2006; Manjarrés, 2005; Quesada Inces, 2001; Shohamy et al., 1996; Vogler et al., 2007). In addition, some effects are positive (Muñoz & Álvarez, 2010); others are negative (Au, 2007; Bickham, Burns, & Monahan, 2001; Thomas, 2004).

As mentioned above, a lot of research has shown that while the introduction of a new test quickly affects the content of teaching, its impact on teaching methods may not be so obvious or may take some time to happen. Wall and Alderson (1993) found no evidence of the impact of the O-level exam in English as a Second Language on Sri Lankan secondary school teachers' methodology. While the exam, with long and complex reading passages, required students to develop skimming and scanning abilities to cope with it, many teachers still believed that "students need to understand all the words and the grammar of every passage" (p. 14). They used a procedure more suitable for lower level classes, i.e., giving background to the topic, explaining vocabulary and structures, reading the passage aloud, reading silently, reading aloud again, etc. before answering questions. They gave students excessive support by "explaining all the difficult words," "dissecting passages sentence by sentence," and "doing considerable amounts of explaining in the first language"

(p. 14), “paraphrasing or translating difficult sentences” (p. 17). There was no evidence of test impact on teaching writing, either. Similar to teaching reading, teachers inclined to provide students with excessive support, not leaving them to collaborate with and help each other as suggested by methodology. The researchers attributed the incongruence between the aim of the exam and teachers’ methodology to teachers’ ignorance of the exam principles and appropriate teaching methods as well as the inadequate guidance from the exam support materials and Teachers’ Guides to the textbooks.

Other research has shown only a superficial influence of tests on teaching methodology. Cheng (1997, 1999, 2003) found that teachers did incorporate role-play and group discussion in their lessons to match the revised HKCEE format, but their interaction pattern remained almost the same: Lessons were “highly teacher-controlled,” and teacher talking time “dominated the majority of the class time” (Cheng, 1999, p. 267). She concluded that test impact on teaching methods occurs “slowly and reluctantly and with difficulties” (Cheng, 1997, p. 38).

Some studies, however, found substantial test impact on teachers’ teaching methods. In Quesada Inces’ study (2001), the standardized Bachillero test, which had a multiple choice format and assessed only the reading skill, made teachers and students focus exclusively on the reading skill only. This teaching to the test led to a complete neglect of the features of the communicative approach. Thus, the teachers ignored the goal of oral production, they did not use any authentic materials or life-like tasks, and they either discarded the materials that were designed to be taught communicatively or altered them to teach to the test. However, the researcher cautioned that the way the teachers taught might not be determined by the nature of the test only. Other influencing factors might be the teachers’ ignorance and erroneous concept of communicative language teaching, their negative attitudes to their teaching jobs, their misconceptions of the evaluative system and the evaluation system itself.

Shohamy et al’s (1996) study, again, showed that the ASL test and the revised EFL test caused contrasting effects on teaching methods, too. While there was almost no evidence of the impact of the former, the latter produced huge changes in the way teachers taught. In the upper level classes, where students were to take the test soon, teachers focused exclusively on developing the oral skills to be tested. There were numerous oral teaching activities, and teachers used a lot of oral teaching techniques to prepare students for the

test, including brainstorming, simulating authentic situations, pair work, group work, jigsaw activities, debates, and discussion. However, the study also revealed that the effects of the EFL test varied depending on the time. The teachers of the lower level classes reported that they focused less on the tested skills, and they used more creative types of oral activities. The reasons for the differences in the impact of the ASL and the EFL tests, as mentioned previously, were the stakes of the tests and the status of the language being tested.

Other studies found variability in the impact of the same test on different teachers. Alderson and Hamp-Lyons (1996) observed and interviewed two teachers, who were teaching both TOEFL preparation classes and language proficiency classes. They found that the two teachers shared some common patterns of differences between TOEFL and non-TOEFL classes. For example, there were more digressions and laughter in non-TOEFL classes; teachers talked more in TOEFL classes; there was less pair work in TOEFL classes. However, the teachers differed a lot from each other. Teacher 1, for instance, used more testing in TOEFL classes (which might be an influence of the test), but teacher 2 did not use class time for testing at all in both types of classes. Alderson and Wall concluded that the test had impact on both teachers, but to different extents, and the difference between TOEFL and non-TOEFL alone could not explain the ways teachers taught. In their opinion, other factors might be involved. These could be the large class size of the TOEFL classes, or materials writers, or the teachers' themselves.

This complexity of washback on teaching methods can be attributed to the interaction of numerous factors. Again, the stakes of the test are cited as one reason. The low-stake ASL test created no impact at all, while the high-stake EFL test created major impact on teaching activities and time devoted to test preparation. Another reason is the context, for example, school settings and the influence of colleagues. In McCarthy's (2008) study, many teachers in high-income schools could enjoy the freedom to teach the way they wanted, while the teachers in low-income schools desperately taught to the test to increase their students' test scores. And the teacher, who is unique with his educational background, teaching experience, belief about the best teaching method, belief about students' belief, and understanding of the nature and rationale of the test plays a central role (Alderson & Hamp-Lyons, 1996; Fournier-Kowaleski, 2005; McCarthy, 2008; O, 2001; Watanabe, 1996). In Alderson's and Hamp-Lyons' (1996) study, for example, the dichotomy between

TOEFL and non-TOEFL could not explain the differences in teachers' teaching. Their individual styles determined the nature of classes observed to a great extent.

2.3.3 Effects of tests on learning.

To date, there have been very few studies on test impact on learning compared to those on teaching. The research into test impact on learning has also focused on the influence of the test on learners' behaviours and their learning outcomes. The following is a review of research findings of these two areas.

2.3.3.1 Effects of tests on learners' motivation.

Some studies show positive effects of language tests on the majority of students' motivation. For example, Li (1990) found positive changes in students' learning after the addition of the language use component to the Matriculation English Test in China. These changes most clearly manifested through "a new enthusiasm for after-class learning of English" with increased extra learning materials and activities, i.e., reading readers and journals, listening to the radio, and watching TV (p. 401). Similarly, Tanabe (2006), who investigated the impact of a TOEIC course on 12 students in Tokyo Polytechnic University, discovered that the majority of them acknowledged positive motivational changes at the end of the course. The students said they were motivated, they were happy with their improvement, they started to like English, they wanted to speak to foreigners, and so on. Li attributed the positive changes to the change in the test design, i.e., the weight put on the use of language component of the test. Tanabe did not give an explanation for the enhanced motivation, but, students' comments seemed to show that one of the reasons for their motivation was their achievement after the course. In Shohamy's (1993) study, 62% of students claimed that the ASL test had positive impact on their learning, including their motivation.

However, other studies (Cheng, 1998; Pan & Newfields, 2012; Shih, 2007) found only minimal test impact on students' motivation. Examining the impact of the revised HKCEE on students' English learning, Cheng (1998) administered the same questionnaire to two different cohorts of students over a two year period. The first cohort had to take the old version of the HKCEE, and the second one had to take the new integrated and task-based version of the test. Comparing the data, she found that "the changes in students' motivation and learning strategies remained minimal" (p. 280). She measured students'

motivation by asking them about their reasons for learning English. Only three out of the nine reasons given showed significant change after two years: *meeting the requirements of the society and watching English movies and listening to English programmes* became more motivating, while *fulfilling parents' expectations* was less motivating. Cheng considered the changes in students' motivation, to some extent, the washback effect of the new test on student learning because "these types of motivation were also related to the requirements of the new 1996 HKCEE" (p. 295).

Investigating the washback of the EFL proficiency graduation requirements (EGR) on university students' learning in Taiwan, Pan and Newfields (2012) also found only small differences in students' motivation. They carried out a survey with two groups of students – one at schools with such requirements and the other at schools without them. To determine the impact of EGR on motivation, they also compared the two groups' responses to questions about their reasons for learning English, many of which were borrowed from Cheng's (1998) questionnaire. They found that only three out of twelve reasons (*to earn certificates, to pass the test to graduate, and to improve their English for further education*) had statistically significant differences, and the effect sizes for these differences were only small. Pan and Newfields associated these changes with the pressure of the EGR on students. They also found that EGR appeared to somewhat motivate some EGR students, but impede low ability students. It is not easy to explain why the findings from Li's (1990), Tanabe's (2006) and Shohamy's (1993) studies were different from those from Cheng's (1998) and Pan's and Newfield's (2012). However, one of the reasons could be the way motivation was measured. The former made inferences about students' motivational changes from their behaviours, attitudes, or direct statements about the change, but the latter measured students' motivation by the changes in students' reasons for learning English.

Some studies have offered explanations for the extent of test impact on students' motivation. The first factors were, of course, the test stakes and test status. Stoneman (2006) compared the effects of the Graduating Students' Language Proficiency Assessment (GSLPA) and the IELTS used as Hong Kong's territory-wide Common English Proficiency Assessment (IELTS–CEPAS) on students at Hong Kong Polytechnic University. She found that a significantly higher percentage of the respondents who sat the IELTS–CEPAS (74.9%) were motivated to prepare for it than the percentage of the

respondents who sat the GSLPA (18.8%) was. The reason was that the IELTS had a higher status than the GSLPA did. Also, 81.8% of the IELTS–CEPAS students did not put as much effort into preparing for the test as they did into other public exams they had taken. This less effort was due to the fact that the IELTS–CEPAS was essentially low stakes. It was only voluntary and students could choose not to include the test results in their transcripts. Shih (2007), however, pointed out that the stakes of the English exit exam (GEPT) for universities in Taiwan alone could not explain its little, and sometimes detrimental, impact on student learning. In his opinion, the de-motivation might be caused by the incongruence of GEPT and the teaching and curricula at institutions of higher education in Taiwan and other extrinsic (e.g., part-time jobs, personal affairs, etc.) and intrinsic factors (e.g., learning attitudes, motivation, etc.)

Research has also revealed some features of test impact on students' motivation. One of them is that motivation can decrease after the test. Berwick and Ross (1989), found that “the overall intensity of motivation of the college students was low” (p. 206). In their opinion, “the intensity of motivation to ‘learn English’ hits a peak in the last year of high school”; and “once the university examinations are over, there is very little to sustain this kind of motivation, so the student appears in freshmen classrooms as a kind of timid, exam-worn survivor with no apparent academic purpose at university” (p. 206). These authors attributed this de-motivation to the fierce competition for a place at university and the grammar and translation focus of the exam.

Another feature of test impact on motivation is that the same test may have a differential impact on different groups of students (Ferman, 2004; Shohamy, 1993). The most important factor that caused the variability on students' motivations seemed to be students' ability. The EFL National Oral Matriculation Test in Israel produced different effects on the three different ability levels: “The lower the students' ability level, the higher the intensity of learning” to the test (Ferman, 2004, p. 199). Gan et al (2004, p. 238) also found that successful students were more willing to prepare for and take tests.

2.3.3.2 Effects of tests on learners' behaviours.

Due to a small number of studies on the effects of tests on learner behaviours, the empirical evidence is still far from conclusive. Some studies found only superficial influence of tests on learners' behaviours (Andrews et al., 2002; Cheng, 1998; Pan & Newfields, 2011, 2012; Shih, 2007). Cheng (1998) found that only four out of the eleven

preferred learning strategies were significantly changed during the two years. In particular, note-taking, a skill required by the new HKCEE, was still the least preferred, and the slight increase in the mean score for this item was not statistically significant. Similarly, Pan and Newfields (2012) also discovered only minimal washback of the EFL proficiency graduation requirements (EGR) on university students' learning in Taiwan. Their data showed that the test requirements "did not lead to a noteworthy amount of 'studying for the test'" (p. 118). To a small extent, EGR students might have adopted communicatively oriented and test-preparation approaches, which resulted in a slight increase in their productive skills. However, most of the students in both groups still embraced traditional methods of learning, i.e., text reading, rote memorization, and practicing grammar exercises. In line with other researchers, Pan and Newfields attributed students' traditional methods of learning to their teachers' traditional methods of teaching. However, Pan and Newfields' conclusion about the impact of EGR based on the comparison of the two EGR and non-EGR groups of students should be taken cautiously because EGR may not be the only factor that was responsible for the differences between them. The compositions of the two groups of students were very different in terms of majors. The majority of the non-EGR group (67%) were Business/management students, while more than half of the EGR group (52.2%) were engineering students. Engineering students could be very different from Business/management students in their motivation and ability to learn English.

Shih's (2007) study also revealed that the English exit exam (GEPT) for university students in Taiwan generated little washback on their learning although it was supposed to have important consequences for them. Arguing that the low-stakes and high-stakes dichotomy could not explain the minimal impact of the test on his participants' learning, Shih suggested a number of other factors that could have mediated the test impact. These included the absence of immediate importance of the test; students' major; their proficiency, which was higher than required for the test; their learning attitudes; their laziness, their unavailability prior to the test; their resistance to learning for the test, and the university's use of a make-up examination. Shih grouped these into three sets of influencing factors which he called extrinsic factors, intrinsic factors, and test factors, and proposed a tentative model of washback of students' learning which involved the interaction of these factors (see Section 2.3.5). Clearly, washback on learner behaviours also seems to be complex and involves as many mediating factors as that on teacher behaviours.

2.3.3.3 Effects of tests on learners' achievement.

In language education, most studies on learners' achievement focus on the relationship between direct test preparation and students' test scores. These studies often used the same method/approach of comparing test scores of groups of students who were involved in direct test preparation to different extents. Findings also showed various degrees of the influence of direct test preparation on score gains.

No clear relationship

Green (2007b) and Read and Hayes (2003) both studied the effect of direct IELTS preparation on score gains of students who were preparing for their academic/tertiary studies in an English speaking country and found no clear evidence of test impact. Green examined the influence of direct test preparation on writing scores. His study involved three different groups of international students in the UK. The first group (85 students) participated in IELTS preparation courses, the second one (331 students) – in English for Academic Purposes courses (EAP) (with no IELTS component), and the third one (60 students) – in combination courses (EAP courses with IELTS preparation strands). All the participants completed IELTS writing tests at the course entry and exit. They also responded to a questionnaire about factors that might account for the differences in the mean score gains. The results showed that the groups with IELTS preparation did not improve their scores more than the EAP group. They also suggested that, in the context of the study, the individual learner's response to the demand of the test affected their outcomes more than the content of their classes.

Read and Hayes (2003) compared score gains on the IELTS listening, reading, and writing tests of international students participating in two IELTS courses at different language schools in New Zealand. Course A mainly aimed at test preparation. Course B used a topic-based approach and focused not only on test tasks, but also on the development of language knowledge and academic skills. Retired versions of IELTS were used for pre- and post-testing. Paired t-tests showed no statistically significant differences in the overall pre- and post-test scores of both courses. There was only a statistically significant difference in the listening scores in course A, which might have resulted from a large amount of class time spent on listening tests and exercises.

Gan (2009) also looked at the influence of IELTS preparation on score gains. However, in his study, IELTS was used as an exit language test for university students in Hong Kong. Comparing the exit IELTS test scores of students who had taken an IELTS preparation course before the test and those who had not, Gan found no statistically significant difference between the groups. However, the findings also showed that the two groups were significantly different in their university English entrance exam scores, and that students with lower entrance scores were more likely to take a test preparation course before the exit test. According to Gan, this suggested that university English learning, IELTS preparation in particular, made important contribution to narrowing down the gap between the two groups in the entrance exam scores. Thus, the relationship between test preparation and score gains in Gan's study seemed more positive than those of Green and Read and Hayes.

Weak relationship

Some other researchers found only a small influence of test preparation on score gains. Robb and Ercranbrack (1999) investigated the effect of TOEIC preparation on score gains. They used two samples of students – English major and non-major students. They assigned each sample to three different treatments: TOEIC preparation, Business English, and General (four-skills) English. The results showed only statistically significant gains for non-majors' reading component. However, the researchers were cautious about the results, saying that they were "by no means conclusive" because the non-major students had much lower pre-test scores than the English majors did, and so they could benefit from test preparation more than higher ability students.

Strong relationship

A few studies, however, reported positive effects of tests on learning outcomes. Saif's (2006) study of the impact of a needs-based test of spoken language proficiency on teaching and learning outcomes was an example. The test was developed for international teaching assistants (ITAs) based on an analysis of their needs. The subjects of the study were then divided into two groups. The control group continued with the old orientation programme for ITAs. The experimental group took a new course, which was taught by an ESL teacher who was involved in the preliminary administration and scoring of the test. The results showed that the experimental group "performed significantly better than the

control group” (p. 26). Saif attributed the superior performance of the experimental group to the improved teaching which was aligned with the needs-based test in terms of teaching objectives, content and activities, and, therefore, suggested a cause-effect relationship between the test and the learning outcomes.

Another experiment carried out by Munoz and Alvarez (2010) to assess the impact of the oral assessment system (OAS) on teaching and learning outcomes in EFL classrooms in Columbia also showed positive effects of the test on learning outcomes. The experimental groups improved their scores in some areas, including communicative effectiveness, grammar, and pronunciation. This improvement may be due to the fact that the teachers of the experimental group put more emphasis on assessing “not only linguistic competence, but also communicative competence as a whole.” Other factors might have also played a role. Those are the ongoing training on assessment practices provided to the teachers, students’ awareness of the test objectives, criteria, procedures, and techniques.

It is noticeable that in both of the studies above the positive impact of the test was a result of stake-holders’ clear understanding of the test they were aiming for and the alignment of teaching/learning with the test. This supported other language testers’ opinions that an introduction of a new test alone is not sufficient to make positive changes to teaching and learning.

In general education, many studies aim at discovering the impact of high-stakes testing in American schools, a result of No Child Left Behind Act (NCLB), and use large scale national data to compare students’ achievement across states and years. There have been mixed results, too, however. Many studies (Amrein & Berliner, 2002; Carnoy, 2005; Grodsky, Warren, & Kalogrides, 2009; Holme et al., 2010; Nichols, Glass, & Berliner, 2006) did not find improvement in students’ achievement. Nevertheless, there are a few studies that found positive effects such as Roderick, Jacob, and Bryk’s (2002) and Braun’s (2004). The dispute over the effects of high-stakes testing thus remains unresolved.

All the above studies examined the effects of tests on students’ achievements quantitatively, i.e., through their test score gains. Andrews et al (2002), however, investigated this aspect of test impact not only “quantitatively,” but also “qualitatively.” The researchers looked at the influence of the Use of English (UE) oral exam on students’ performance in spoken English. A spoken test was administered to three different cohorts

of Secondary 7 students in three consecutive years 1993 (before the introduction of the exam), 1994, and 1995. A comparison of the mean scores showed an apparent, although not statistically significant, improvement in students' test scores between 1993 and 1995, suggesting a positive effect of the EU exam on students' performance. The analysis of students' recorded performances also revealed the changes in their organizational and language features. For example, students stopped introducing themselves when starting group discussion, probably, because they had become familiar with the exam format. Another example is students' use of a wider range of language to introduce their presentation topic. Sometimes the changes in the content of students' performance reflected the influence of published materials, which acted as a mediating factor in the relationship between the test and students' performance. Andrews et al also noted that the test impact was delayed, with the differences being more obvious in the second year of the test.

In brief, research findings on the three areas reviewed above have not only confirmed the existence of washback, but also revealed how complex the phenomenon is. Having got the positive answer to his question "Does washback exit?" after approximately ten years of empirical research on the topic, Alderson, however, noted that the existence of washback "raises more questions than it answers" (Cheng et al., 2004, p. xii). As seen above, a myriad of factors come into play and interact with each other during the washback process. However, they can be classified into three categories: test factors, person factors, and context factors. Test factors include the stakes of the tests, its purpose, content, format, interpretation, skills tested, and the status of the language tested. Person factors include stakeholders' beliefs, their attitudes, their educational background and experience, their personalities, and age. Context factors include macro-context (society) and micro-context (school settings, other stakeholders, etc.).

2.3.4 Impact of TOEIC.

This section reviews empirical research on the impact of the TOEIC on all the areas reviewed above. Surprisingly, up to now, there have been very few studies on the influence of the TOEIC test compared to those on the effects of other major international English tests such as TOEFL and IELTS. Many of these studies are only preliminary in nature. They explore different aspects related to TOEIC and its impact. Some reports, for example Apichatrojanakul (2011) and Nhan (2013), did not describe the research designs,

instruments, and procedures in detail. Therefore, it is unknown whether the findings can be generalized. However, they raised some interesting issues relating to the use of the TOEIC test for various purposes and in various contexts. Apichatrojanakul (2011) reported his own experience of teaching TOEIC preparation courses in a business school in Thailand and the perceptions of some teachers and students that he interviewed. The themes were built around the two approaches of teaching TOEIC courses that the school used. In his first year of teaching TOEIC, the concept of 4-mat model was adopted. It incorporated a lot of communicative activities such as pair work, group work, and presentation as well as grammar exercises and quizzes, but it did not include sample TOEIC questions. However, at the end of the year, most of the students did not achieve the minimum requirement of TOEIC 150 and 200. In the second year, the school decided to abandon the student-centred teaching style. It started to adopt a teacher-centred, grammar translation style. Their study sheets included six grammar topics often assessed in the TOEIC test and sample TOEIC questions. Teachers also used new materials and new techniques such as reading comprehension, fill-in-the blanks, and synonym-antonym. As a result, most of the students' test scores improved. The researcher attributed students' poor test results in the first year to the mismatch between the TOEIC test and the teaching/learning styles and materials. The interviews with some teachers and students at the school supported his opinion. Most of the interviewees believed that the teacher-centred, grammar-translation style was more effective and faster for the purpose of boosting students' scores because it was more straightforward, faster and easier for explaining grammar points. However, many of them missed the enjoyment and opportunity for discussions and interactions in the first year. Some teachers felt guilty not teaching their students to speak and write. It is ironic that the grammar-translation style and direct test preparation seemed to be more effective for preparing students for a test of communication than the communicative approach.

Nhan (2013) wrote about the issues related to the use of the TOEIC test as an exit requirement in tertiary institutions in Danang city, Vietnam. Her paper was based on "selected comments" from an unspecified number of teachers, students, and administrators who responded to her questionnaire. It discussed the challenges that teachers and students experienced with the adoption of the TOEIC policy and slightly touched on some effects of the test on the tertiary sector and on the wider society. Thus, there was a rather big gap between students' English levels at entry and the required levels at exit, but the contact hours were not enough to fill the gap. Also, students with low English ability, who usually

came from rural areas or studied another language at school, were at a disadvantage and the consequences of failing to meet the benchmarks could be serious. Besides, learners from low-income backgrounds experienced financial problems because of the high cost of test preparation courses. As for teachers, they had to deal with the tension to meet both course objectives and TOEIC preparation. Teachers also experienced difficulty preparing their students for the test due to a lack of training or professional development. Universities had difficulty implementing and evaluating the policy. As for the wider society, Nhan mentioned the appearance of numerous language centres providing TOEIC preparation courses and the flourishing of the publishing sectors, testing franchises and authorized TOEIC consultation offices. In brief, Nhan warned of the social inequality that the policy might create and raised the issues of its implementation. However, she believed that using an [international] English test as an exit requirement was necessary because companies preferred international certificates – something that has not been proved yet.

Newfields' pilot study (2005) explored teachers' perceptions of some washback aspects related to the use of the TOEIC test as a placement tool and a curricular component in a Japanese university. Through a questionnaire and an interview, he asked 24 English teachers questions about the usefulness of TOEIC as a placement tool, its difficulty for their students, students' motivation to raise their test scores, students' improvement, and the test impact on teaching. The results also showed little influence of the test on teaching. Very few teachers spent most of the class time on explicit test preparation. The amount of time devoted to explicit test instruction also varied depending on students' ability and the points in time during the semester. Only a quarter of teachers admitted that the TOEIC test influenced their teaching style in some way. Concerning learning, many of the teachers believed that the students were motivated to improve their test scores, and the majority of them thought that focusing on TOEIC could improve students' overall English ability. However, the researcher himself admitted that the study had many limitations, and that the results described above were rather impressionistic.

The most robust research on the impact of TOEIC that I know was carried out by Booth (2012). The researcher investigated the effects of the test on English learning in a large university in South Korea. Using Activity Theory as her theoretical framework, she not only recorded students' actions and beliefs during the test preparation process, but also put an emphasis on explaining the way different factors contributed to test impact. Thus, the

study took into consideration sociocultural, historical, political, and economic contexts which learners were situated in. Data was collected through focus group, semi-structured journals and multiple interviews and analysed both quantitatively and qualitatively. Journal data showed interesting differences between English-majors and other-majors in their ways of preparing for the TOEIC test. While other-majors focused directly on test preparation, English-majors prepared for the test indirectly. If English-majors directly focused on the test, they often used non-TOEIC materials. Other-majors often chose to go to language institutes for test preparation. When preparing for listening, other-majors paid attention to discrete language rules and vocabulary items, while English majors considered reading to be a more holistic activity. The study yielded further evidence of the complexity of test impact, which was illustrated by the model of Washback on The Processes and Products of Learner Test Activity. The model includes three main components: (1) the test-taker, community and test complex; (2) processes: learner action in context; and (3) products from student test activity. The first component consists of the test-taker, the extended community, and the test, which are interrelated to and interdependent on one another. The test-taker refers to individual learner's attributes such as individual differences, self/social/cultural identity, constructs of self as learner, attitudes, and beliefs. The extended community is composed of all groups and individuals related to the test and test-taker, for example, the test maker, the wider cultural, political, economic sector, preparation schools, textbook publishers, parents, siblings, friends, and peers. In accordance with sociocultural theory, the extended community was also defined as including not only the one that the test and the test-taker currently belong to but also the one that they belonged to in the past (previous learning experiences in different communities) and the one that they wish to belong to in the future. Some influences of the wider community on the test-taker's test preparation might include advice from peers, the culture of learning by memorization, and company requirements. The test includes factors such as test qualities (test construct, content, structure, format and tested skills) and other factors (scoring procedures, the status of the tested language, wider stakes, and degree to which the test is counter to current learning practices).

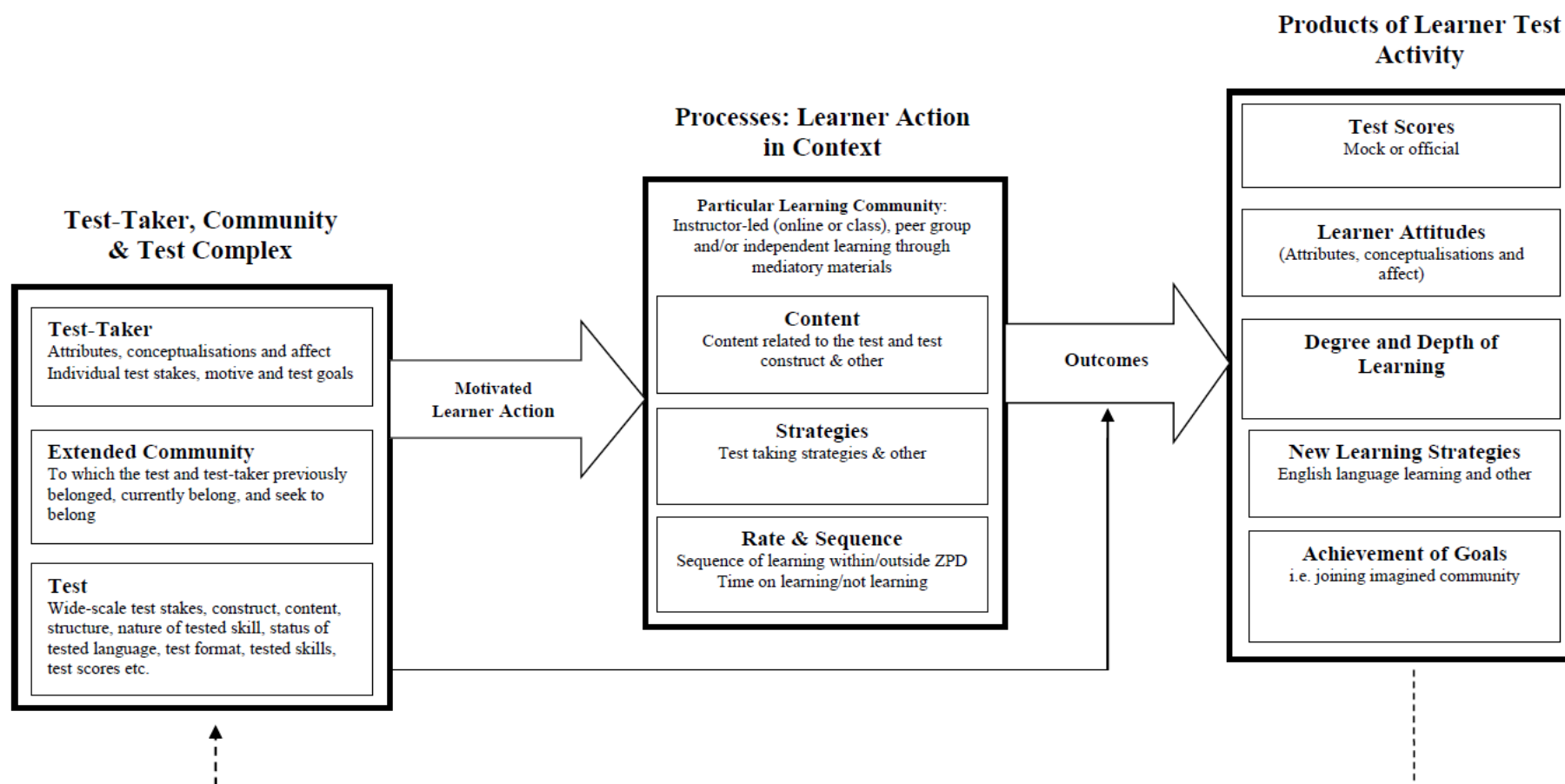


Figure 1 Washback on the Processes and Products of Learner Test Activity (Booth, 2012, p. 292)

Learner test activity is motivated by the test-taker, wider community, and test complex. Under the influence of this complex, the test-takers chooses to engage with others in certain learning communities, focus on certain content of learning (directly related to the test or indirectly related to the test), use certain learning strategies (direct test-taking strategies or other strategies), and learn with certain rate and sequence of learning (within or outside his/her Zone of Proximal Development).

The learning processes result in learning outcomes or products of learner test activity. These products might include the learner's test scores, his/her attitudes, the degree and depth of learning, the new learning strategies, and the extent by which the learner achieved their goals. This model of washback succeeds in capturing the complex and dynamic nature of the phenomenon. It shows that learner test preparation activity was influenced by a complex of interrelated and interdependent factors including the test-taker, the wider community, and the test.

2.3.5 Influencing factors and models of test impact on teaching and learning.

As can be seen from the previous sections, test impact on teaching and learning can vary a lot from context to context, and its formation involves a myriad of influencing factors. These can be stakes of the test and the status of the language it assesses (Shohamy et al., 1996), teachers' characteristics (Alderson & Hamp-Lyons, 1996; Watanabe, 1996), and school settings. From findings in the research literature, Watanabe made a list of five factors mediating washback: "test factors (e.g., test methods, test contents, skills tested, purpose of the test, and decisions that will be made on the basis of test results); prestige factors (e.g., stakes of the test, and status of the test within the entire educational system); person factors (e.g., teachers' educational backgrounds, and their beliefs about the best methods of teaching/learning); micro-context factors (e.g., the school setting in which the test preparation is being carried out); and macro-context factors, that is, the society where the test is used" (Cheng et al., 2004, p. 22). It would be, however, clearer and more systematic to classify the factors into three sets: test factors (including Watanabe's test factors and prestige factors), person factors, and context factors (micro-context factors and macro-context factors).

The variability of test impact depends on the specific involvement and interaction of these factors. Findings from a single study are not enough to explain the mechanism of the

process. In fact, there have been several different attempts to explain the complexity of the process of test impact generation. In 1993, when posing their question about the existence of washback, Alderson and Wall (1993) proposed fifteen washback hypotheses as suggestions for future washback research. These hypotheses were related to:

- the existence of washback
 - (1) A test will influence teaching.
 - (2) A test will influence learning.
- the aspects of teaching and learning which might be affected by a test
 - (3) A test will influence what teachers teach.
 - (4) A test will influence how teachers teach.
 - (5) A test will influence what learners learn.
 - (6) A test will influence how learners learn.
 - (7) A test will influence the rate and sequence of teaching.
 - (8) A test will influence the rate and sequence of learning.
 - (9) A test will influence the degree and depth of teaching.
 - (10) A test will influence the degree and depth of learning.
 - (11) A test will influence attitudes to the content, method, etc. of teaching and learning.
- variability of washback and factors causing this variability
 - (12) Tests that have important consequences will have washback.
 - (13) Tests that do not have important consequences will have no washback.
 - (14) Tests will have washback on all learners and teachers.
 - (15) Tests will have washback effects for some learners and some teachers, but not for others.

In the same year, Hughes (as cited in Bailey, 1996) proposed a mechanism of how washback operates. This mechanism includes three constituents: participants, process and product in teaching and learning. Participants refer to all the stakeholders “whose perceptions and attitudes towards their work students may be affected by a test” (p. 262) such as classroom teachers, administrators and publishers. Process includes “any actions taken by the participants which may contribute to the process of learning” (p. 262). Such processes can be material development, syllabus design, changes in teaching methodology, and the use of learning and/or test-taking strategies. Product means “what is learned (facts,

skills, etc.) and the quality of the learning (fluency, etc.)” (p. 262). Hughes (as cited in Bailey, 1996) explained how the mechanism works:

The trichotomy into participants, process and product allows us to construct a basic model of backwash. The nature of a test may first affect the perceptions and attitudes of the participants towards their teaching and learning tasks. These perceptions and attitudes in turn may affect what the participants do in carrying out their work (process), including practising the kind of items that are to be found in the test, which will affect the learning outcomes, the product of that work. (p. 262)

Combining Alderson’s and Wall’s washback hypotheses and Hughes’s ideas, Bailey (1996) presented a basic model of washback given in figure 1. She further distinguished between washback to the learner and washback to the programme arguing that the washback process would vary greatly across constituencies of participants. Washback to the learners refers to the direct effects of test-derived information on the test-takers, and washback to the programme – those effects on the teachers, administrators, curriculum developers, counsellors, and so on.

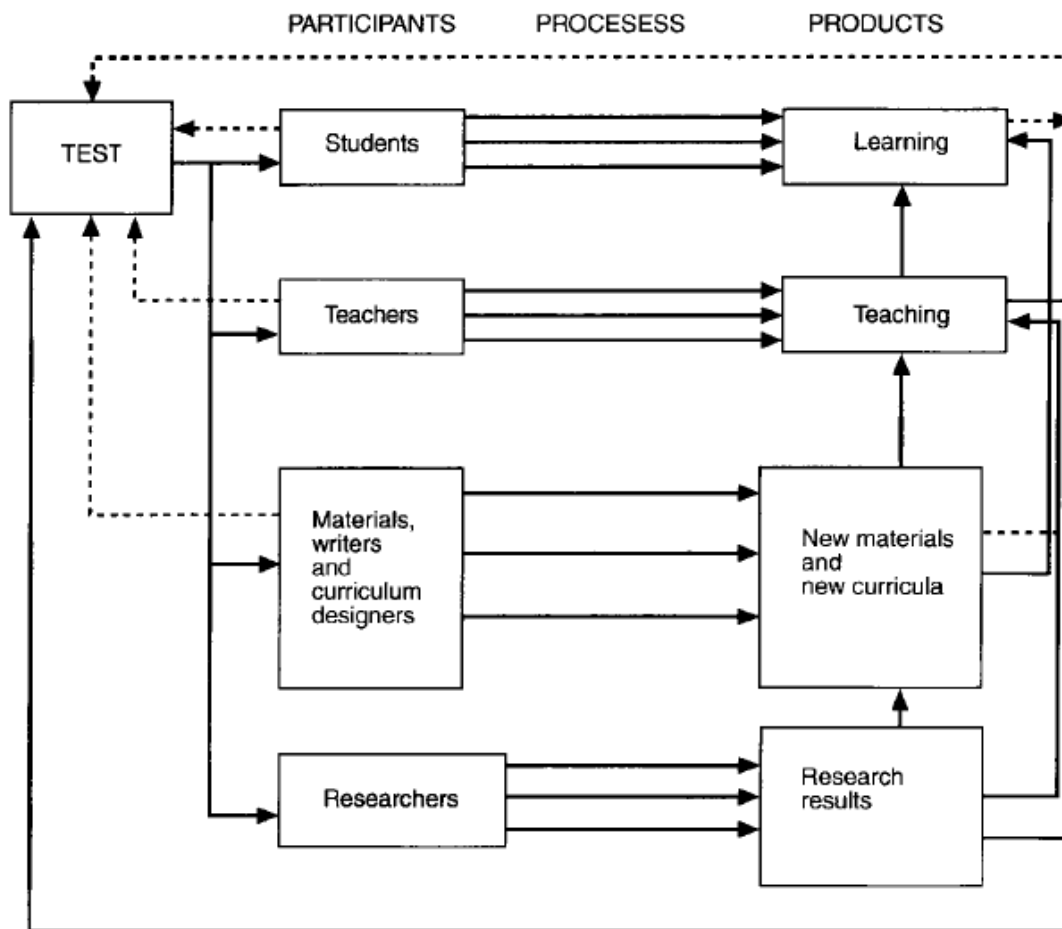


Figure 2 A basic Model of Washback (Bailey, 1996, p. 264)

However, Hughes' ideas and the model later elaborated by Bailey still seem simplistic compared to what we now know about washback. The washback process not only differs for different 'constituencies of participants', but also for participants of the same constituency. The model does not show and account for the variability of washback on the same group of participants.

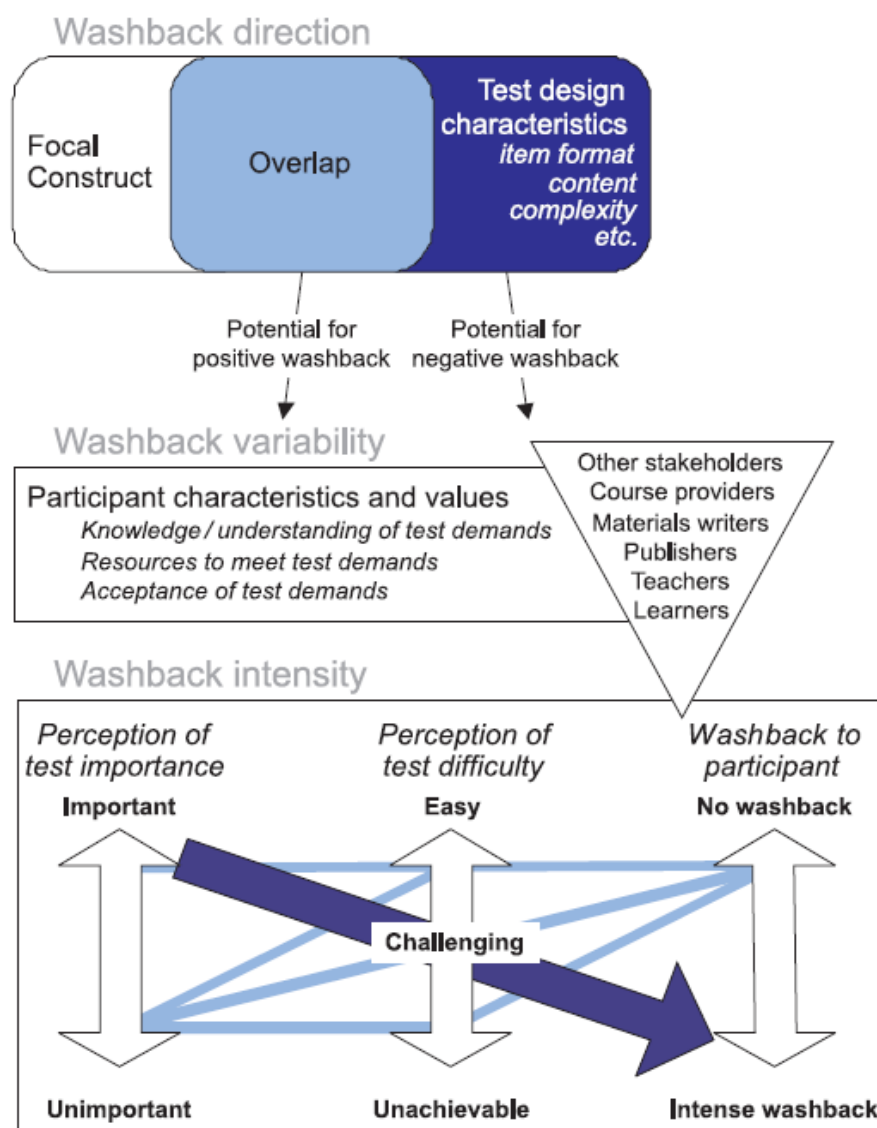


Figure 3 Green's Model of Washback Direction, Variability, and Intensity (Green, 2006, p. 340)

The later models attempted to fill this gap. Green's model of washback (Green, 2006, 2007a) explains the direction, variability and intensity of washback by showing the interaction of factors involved. Regarding the direction of washback, the focal construct (the target skills of a curriculum or a target domain as understood by course providers and learners) and the test characteristics are the responsible factors. The more they overlap, the more likely it is that positive washback occurs and vice versa. The variability of washback is explained by the differences in the participant characteristics and values which include their knowledge/understanding of test demands, resources to meet test demands that they have, and their acceptance of test demands. As for the intensity of washback, it is shaped by the interaction between the stake of the test and its difficulty. For example, a test which is perceived as important and challenging will have intense washback, while a test which is considered important, but easy will have no washback. This model, however, ignores

context as a factor responsible for washback variability. Some research findings (Diamond, 2007; Hamp-Lyons, 1998; McCarthy, 2008; Shih, 2007) have suggested that such factors as society, school settings, kinds of schools (high- or low-income), culture, and the influence of peers and other stakeholders can influence washback.

As reviewed in section 2.2.3, based on the results of his own and others' studies, Shih also proposed a model of washback on learning, which incorporates three groups of factors: extrinsic factors, intrinsic factors, and test factors (see Figure 4). Extrinsic factors include socio-economic factors; school and educational factors; family, friends, and colleagues; and personal factors. Intrinsic factors are composed of individual differences, personal characteristics, and personal perceptions of the test. Test factors are the stakes of the test, the immediate importance of the test, degree to which the test is counter to current learning practices, relative difficulty of the test in relation to students' proficiency, associated loopholes of the test, content of the test, structure of the test, nature of the tested skill, status of the tested language, format of the test, tested skills, and purpose of the test. This model takes into consideration the context, however, it does not show the aspects of test impact that each factor contributes to and direction of its influence like Green's (2007a) model does.

From a sociocultural theory perspective, Booth (2012) produced a model of test impact called Washback on the Processes and Products of Learner Test Activity (for a description of the model, see Section 2.3.4). Her model uses a different way of grouping the influencing factors from Shih's. It successfully illustrates the complexity of test impact emphasizing the interrelatedness and interdependence among the infinite number of influencing factors. However, like Shih's model, it does not show the aspects of test impact each factor contributes to as well as the significance of its contribution, either. It is clear that because test impact is too complicated, no single model can capture all of its aspects.

2.4 Measurement-Driven Instruction

The power of tests over educational systems in particular and society in general has long been documented. Tests are often used to control teaching and learning both in general education and language education. This use of tests is often referred to as measurement-driven instruction (MDI). According to Popham (1987), “measurement-driven instruction occurs when a high-stakes test of educational achievement, because of the important contingencies associated with the students’ performance, influences the instructional programme that prepares students for the test” (p. 680). However, there has been fierce debate about MDI. Proponents of MDI (Mehrens, 1984, 1998; Popham, 1987; Wiliam, 2010) argue for its benefit for improving student performance. Popham (1987) even considered MDI “the most cost-effective way of improving the quality of public education” (p. 679). The opponents of MDI (Kohn, 2000; Madaus, 1985; J. Smith & Kovacs, 2011; Thomas, 2004; Wright & Choi, 2006) point out the detrimental effects it has on teachers and students such as stress, pressure, the narrowing of curriculum, superficial teaching, and unfairness for ELL students. While the debate is still continuing, more and more high-stakes tests have been used to drive education. A well-known example of MDI in general education is the use of high-stakes tests in the USA to hold schools, districts, and states accountable for their teaching quality as required by the NCLB. Schools have to meet adequate yearly progress of their students; otherwise, they will face sanctions including loss of federal funding or closure (Bekis, 2008; McCarthy, 2008; Menken, 2009). Another example is the End-of-Key-Stage statutory national testing in England used to monitor individual progress over time to inform students and their parents and to monitor the performance of schools and local authorities (Collins et al., 2010).

In language education, the introduction of a new test is usually associated with curriculum innovations. Tests are often introduced or revised to motivate teaching and to be aligned with new teaching goals or philosophies. The revised HKCEE, for example, was introduced to change teaching and learning from “noninteractive, teacher dominated classrooms to more task-based teaching approaches” (Cheng, 2004, p. 149). Similarly, the new O-level English examination in Sri Lanka was adopted to reinforce the teaching of more practical English with an emphasis on everyday communication (Wall & Alderson, 1993).

However, research on the effectiveness of high-stake testing for instructional innovation has yielded mixed findings. Examining international experience in using high-stakes test to innovate education, D.W. Chapman and Snyder Jr (2000) concluded that “testing can be an effective mechanism for improving instructional practice, but its success is not ensured” (p. 457). What has been learned from the world experience is that in order to make educational improvement through high-stakes testing, a lot of issues should be taken into consideration. The following are several guidelines for successful use of high-stakes testing for educational change. First, there should be an alignment among teaching goals, teaching activities, and testing so that teaching and learning to the test will not be different from teaching and learning the curriculum, which will eventually help develop all the skills in the target domain (Andrews, 2004; Bailey & Masuhara, 2013; Green, 2007a; Messick, 1996). Weigle and Jensen (as cited in Bailey & Masuhara, 2013, p. 304) suggested that “when assessment procedures are aligned with instructional goals and activities, positive washback results.”

Second, according to Weigle and Jensen (as cited in Bailey & Masuhara, 2013, p. 304), “test tasks should require the same authentic, interactive language use promoted in the classroom.” The use of authentic assessment has become a new trend in language assessment (Cheng & Curtis, 2004). In some cases it has had beneficial effects on teaching and learning. For example, the teachers in Stecher et al’s (2004) study reported that “replacing or supplementing multiple-choice tests with more performance-based assessments has led to a dramatic increase in the amount of writing students do in school” (pp. 68-69).

Third, performance-based assessment cannot automatically ensure success. Researchers pointed out the necessity of understanding the sociocultural context, educational context, stakeholders’ concerns, and factors facilitating or inhibiting the success of innovation in assessment (Andrews, 2004; Cheng & Curtis, 2004; Wall, 2000). Wall (2000) suggested that test developers, who want to use assessment to innovate instruction, should apply the Hybrid Model of the Diffusion/Implementation Process proposed by Henrichsen (1989) to their innovation process. The model includes three components: Antecedents, Process, and Consequences. Test developers should examine the Antecedents component to understand the existing conditions in the educational context before the innovation is introduced. Then they should analyse the Process component to identify the factors in operation that

facilitate or hinder the innovation when it is being implemented. After analysing the Process, they should investigate the Consequences that result from the interaction between the factors in the Antecedents and the Process.

Fourth, D. W. Chapman and Snyder Jr (2000) have suggested that policymakers who want to use high-stakes testing to improve instruction should clarify and elaborate the linkages that connect testing to improved instruction because “It is not enough to call for more testing in the blithe hope that it will lead to the desired ends” (p. 467). They illustrated these linkages with a model of the influences of high-stakes testing on instructional practice, which showed the logical paths from the possible uses of high-stakes testing to improved instruction with one or more intermediate links. For example, one of the paths shows that when test scores, which are released to public, are used to motivate teachers to improve their teaching, parents and community will pressure teachers to improve their teaching. This pressure is linked with teachers’ understanding of what changes in instructional practice can raise student performance. This understanding will result in improved instructional practice and, consequently, in improved student performance.

Finally, Xie and Andrews (2013) suggested that for measurement-driven instruction to meet its target, “intended skills need to be endorsed by test-takers as necessary” and “Test-takers also need to consider the test valuable and doable” (p. 64). In brief, in order for educational reforms through examination change to be successful, examinations should be introduced with great care. “It should not be assumed that a ‘good’ test will automatically produce ‘good’ effects in the classroom, or that a ‘bad’ test will necessarily produce negative ones” (Wall, 2000, pp. 505–506).

2.5 Chapter Summary

This chapter reviewed research literature on test impact on teaching and learning in both language education and general education. It began with definitions of terms commonly used to refer to the phenomenon. Then it looked at five dimensions of test impact. Next, it reviewed research findings relating to areas of teaching and learning affected by testing and the mechanism of test impact. Finally, it concluded with the issue of using high-stakes testing to drive instructions.

What is now clear about test impact on teaching and learning is that it is far more complex and elusive than people used to think. It can exist in different form and intensity, and the

quality of the test alone cannot explain how it operates. To date, three main sets of factors have been pinpointed as the variables shaping test impact: test factors, person factors, and context factors. Nevertheless, because test impact is so context-specific, the list of factors in each set does not seem exhaustive. In each context, new factors emerge, a different combination of factors comes into play, and the factors involved interact with each other in a different way. For this reason, many language testers (Spratt, 2005; Tsagari, 2007) have called for more washback studies in new contexts. Also, research has shown that tests can be used to improve instruction, but great care needs to be taken to ensure their success.

Chapter 3: METHODOLOGY

3.1 Research Questions

The purpose of the study was to investigate the impact of the TOEIC as a university exit test on English teaching and learning in the Vietnamese context. To achieve that purpose, the study aimed at answering the following research questions:

1. What are policymakers' perspectives on the TOEIC-as-exit-test policy?
 - 1.1. What are policymakers' perceptions of the purposes of language teaching at university?
 - 1.2. What are policymakers' rationales for choosing TOEIC as a university language exit test?
 - 1.3. What is the intended impact of the TOEIC-as-exit-test?
2. What is the impact of the TOEIC-as-exit-test on English teaching and learning in the Vietnamese context?
 - 2.1. What are teachers' perceptions of the impact of the TOEIC test on university English teaching?
 - 2.2. What are students' perceptions of the impact of the TOEIC test on university English learning?
 - 2.3. Are there differences between large city universities and provincial universities in teachers' and students' perceptions of the impact of the TOEIC-as-exit-test?
3. What brings about the impact of the TOEIC-as-exit-test on university English teaching and learning?
 - 3.1. What test factors, context factors, and person factors play major roles in shaping the test impact?
 - 3.2. How do the factors involved interact to produce the impact?

3.2 Overall Research Design

To answer the research questions, a mixed-methods research design was utilized. According to Fraenkel and Wallen (2009), mixed-methods research has three advantages. First, it can “help to clarify and explain relationships found to exist between variables” (p. 558). Second, “it allows us to explore relationships between variables in depth” (p. 558).

Finally, it “helps confirm or cross-validate relationships discovered between variables” (p. 558).

The use of multiple methods in this project is a combination of two types of design: explanatory design and triangulation design. In the explanatory design, a researcher employs a quantitative method followed by a qualitative one to “refine the quantitative findings” (Fraenkel & Wallen, 2009, p. 560). Thus, the current project was conducted in two phases utilising two different methods. In Phase 1, questionnaires were used with teachers and students to identify the general patterns of the test impact. In Phase 2, a teacher and student case study was conducted to explain the patterns found in Phase 1. According to Flyvbjerg (2011), these two methods could complement each other well:

...if you want to understand a phenomenon in any degree of thoroughness ..., you need to do case studies. If you want to understand how widespread the phenomenon is, how it correlates with other phenomena and varies across different population, and at what level of statistical significance, then you have to do statistical studies. If you want to understand both, which is advisable if you would like to speak with weight about the phenomenon at hand, then you need to do both case studies and statistical analyses. The complementarity of the two methods is that simple, and that beautiful (p. 314).

In the triangulation design, “researchers make use of multiple and different sources, methods, investigators, and theories to provide corroborating evidence” (Creswell, 2013, p. 251). Triangulation is often used to seek convergent evidence, but it can also be used to “ascertain multiple forms of interpretation” (Duff, 2008, p. 30) or to reveal inconsistencies and contradictions which may be useful for reshaping our framework and findings (Anzul, Ely, Freidman, Garner, & McCormack-Steinmetz, 1991). The current study took advantage of data and methodological triangulation to understand the complexity of the test impact phenomenon and the related issues. Thus, in Phase 2 of the study, the evidence of the impact of the TOEIC test on teaching was collected through semi-structured interviews and classroom observations, and the signs of its influence on learning – through semi-structured interviews and weekly journals.

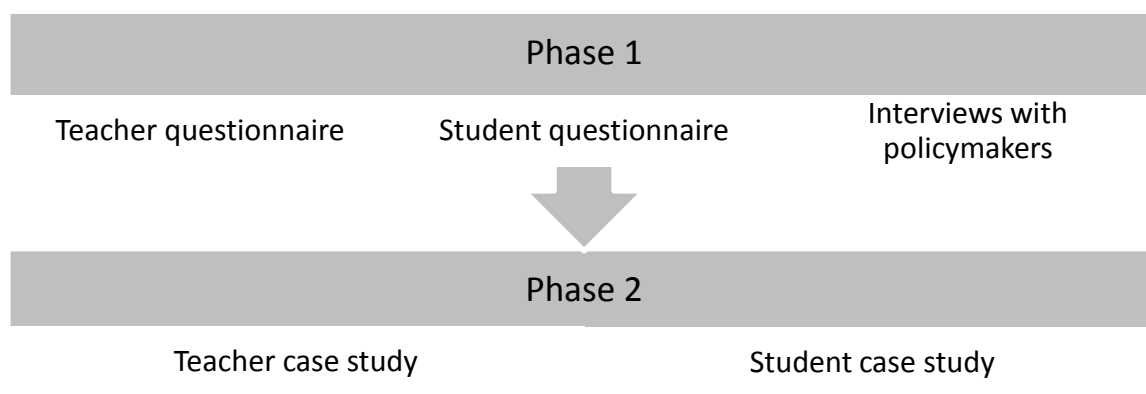


Figure 5. Overall Research Design

3.3 Participants and Research Sites

The data collection was carried out in various places in Vietnam in 2012. The following is the description of the participants and research sites.

3.3.1 Policymakers.

Table 3.1 Policymaker Participants

	Types of participants	No of participants	Organizations
1	MOET officials (MOs)	2	MOET based in Hanoi
2	Heads of Department (HODs)	12	Universities around Vietnam

Policymakers were interviewed face to face or on the telephone. They consisted of two MOET officials, who were based in Hanoi, eleven Heads of English Departments and the like, and one Head of Training Department, who worked at different universities around Vietnam.

3.3.2 Teachers and students.

3.3.2.1 Teachers and students participating in questionnaire piloting.

A total of 102 students and 15 teachers from Department of Tourism and Department of Information Technology at a private university in Ho Chi Minh City were invited to participate in piloting the questionnaires (see Table 3.2). At this university, the TOEIC benchmarks were introduced in 2010, and the first cohort of students under the language exit test policy will have to take the test in 2014. In the university's English programme,

TOEIC preparation is taught as one component of some English modules. At the time the piloting was carried out, only the students at the Department of Tourism had been taught some TOEIC preparation. The teachers who participated in the piloting stage either had had experience teaching TOEIC at this university or at other institutions.

The piloting process went through three steps. Relevant responses of the teachers participating in the first and second steps of the teacher questionnaire were also used for the third step data analysis as the third version was only slightly modified compared to version one and version two.

Table 3.2 Participants in Questionnaire Piloting

	Department of Tourism		Department of Information Technology	
	Students	Teachers	Students	Teachers
First piloting	2	1		
Second piloting	3	1		
Third piloting	75	3	22	10
Total	80	5	22	10

3.3.2.2 Teachers and students participating in main study.

3.3.2.2.1 Research sites.

Phase 1 survey and Phase 2 case study were conducted at the same three universities in South Vietnam. These research sites were chosen based on four criteria. First, they should have the TOEIC-as-exit-test policy. Second, they should be large in size (Each of them recruited approximately 4,000 new students in 2011). Third, they should be located in either a large city or a province because one of the main purposes of the study was to explore the role of school settings in test impact generation. Last but not least, they should be accessible to the researcher.

The researcher's initial intention was to carry out research at two large city universities and two provincial ones. However, she was unable to find a second provincial university which was accessible to her as well as met her selection criteria. Thus, University A and University B were located in a large city, and University C was situated in a province near the large city. While University A and University B provided courses in only one field,

business and technology, respectively, University C is a comprehensive one, which had courses in both natural and social sciences (see Table 3.3). Universities A and B had a reputation as top universities in Vietnam. In 2014, University A's admission score was 21, and University B's admission score was from 20 to 21 depending on majors. In the same year, University C admitted students having a university entrance exam score in the range from 12 to 14.5.

The three universities differed greatly in their English programmes. University A had four English modules (BE1, BE2, BE3, BE4), which taught Business English with a small proportion of TOEIC preparation. There were two TOEIC benchmarks: 550 points for a few majors including Tourism, Foreign Trade, and International Business and 450 points for the other majors. Students had to take the international TOEIC test. They could also submit other international English certificates such as IELTS and TOEFL instead of TOEIC for graduation.

At University B, students studied three modules of General English and one module of TOEIC Preparation at the university's English Language Centre. They might also take ESP courses, which were usually optional, through their own departments. An international TOEIC certificate with a score of 450 was a prerequisite for graduation. Like students at University A, students here could also submit other international English certificates; however, according to the Head of the Training Department, the majority of students studied for and took the TOEIC test.

University C provided TOEIC Preparation courses to all students at their Basic English Department. There were five levels of TOEIC preparation: TOEIC 1, TOEIC 2, TOEIC 3, TOEIC 4, and TOEIC 5. Apart from TOEIC preparation courses, students had to take ESP courses through their own departments. Most ESP courses varied between 45 and 60 periods. However, due to their specific characteristics, some departments reserved more time for ESP courses. Students had to either pass a TOEIC-format test called internal TOEIC, which was written and administered by the university, or submit an international TOEIC certificate. The pass score was 400. Other international certificates such as TOEFL and IELTS were also accepted.

Table 3.3 Research Sites for Main Study

	University A	University B	University C
Location	Large city	Large city	Province
2011 recruitment	4,000	Approximately 3,800	Approximately 3,911
Private/ state	State	State	Private
Training programmes	Business	Technology	Comprehensive
English courses	Business English with a TOEIC Preparation component	<ul style="list-style-type: none"> - General English; - ESP (optional); - TOEIC Preparation 	<ul style="list-style-type: none"> - TOEIC Preparation; - ESP
Textbooks	<ul style="list-style-type: none"> - Market Leader Elementary - Market Leader Pre-intermediate - Further Practice Book 	<ul style="list-style-type: none"> - Cutting Edge Pre-intermediate - Longman Preparation For The New TOEIC Test: Introductory Course 	<ul style="list-style-type: none"> - TOEIC Preparation Course - Materials selected or compiled by teachers teaching ESP courses
TOEIC Benchmarks	<ul style="list-style-type: none"> - TOEIC 550 (for Foreign Trade, International Business and Tourism majors) - TOEIC 450 (for other majors) 	<ul style="list-style-type: none"> - TOEIC 450 	<ul style="list-style-type: none"> - TOEIC 400

At all three universities, students had to pass the course assessment of all subjects including English and achieve the TOEIC benchmark in order to graduate. They had to submit the required TOEIC score before the graduation exams or graduation review. At university B, apart from the benchmark of 450 that students had to achieve before the graduation review, they were required to submit a TOEIC score of 400 in the middle of year 4 in order to register for their thesis. Without the TOEIC score of 400, they would be suspended from their studies for one semester to have time to prepare for the test. The consequences for not achieving the TOEIC benchmark were the same at these universities:

Those students who failed to meet the benchmark could not get their bachelor's degrees and had to prepare for the test themselves to submit the required score within two years of the end of their programs. If they did not get it within this time period, their bachelor's degrees would be cancelled. At University C, failing students were provided with temporary graduation certificates, which might be accepted for job application at some businesses and institutions. However, the consequences for not passing the course assessment were slightly different at the three universities. At University A, students always had to retake the course that they failed. At University B, generally, students also had to retake the course, but some programs allowed them to retake the exam once before retaking the course. At, university C, all failing students had an opportunity to resit the exam once. If they still failed, they had to retake the course. Both the course assessment and the TOEIC benchmark were important for students' graduation. However, the interviews with the case study students showed that they cared about the TOEIC test more than the course assessment including the final exams because the former was perceived to be more difficult.

3.3.2.2 Teachers and students answering questionnaires.

Teachers

All the teachers who taught TOEIC preparation courses or were involved in preparing students for the TOEIC test at the three universities described above were selected for the survey. Thus, at University A and University C, the teacher questionnaires were handed out to all the English teachers who taught non-English major students, but at University B, they were only distributed to those who were teaching or had taught TOEIC preparation courses before. A total of 81 teachers returned the questionnaire with the response rate being 83.65%.

Table 3.4 Number of Teachers Participating in the Main Study – Phase 1

	University A	University B	University C
Number of participating teachers	23	32	26

Students

Student participants were selected by class. At University A and University C, students enrolled in classes based on their year at university and major, so all the students in a class

were of the same year and had the same major or at least came from the same department. However, University B used a credit-based training system, so students could enrol in any class that they preferred as long as they met the course pre-requisites. Therefore, each class might be composed of students of different majors and years. As a result, the principles of student selections were different for the three universities. At University A and University C, the researcher tried to include different groups of majors into the sample and to ensure that the sample relatively reflected proportions of the majors in the population. However, this was impossible for University C; therefore, at this university only convenient classes were chosen.

At each university, the highest level classes at the time the research was being carried out were selected for the study because these classes had had some experience with the TOEIC test, and the test impact on them was believed to be the strongest. Therefore, students who were studying Business English 4 at University A, English 4 (TOEIC preparation) at University B, and TOEIC 4 at University C were selected (At the time the survey was conducted, no TOEIC 5 classes were available at University C). A total of 1501 students returned the questionnaire with the response rate being 97.91%.

Table 3.5 Number of Students Participating in the Main Study – Phase 1

	University A	University B	University C
Course	Business English 4	English 4	TOEIC 4
Year	Year 3	Different years, but mainly year 2	Year 2
Number of students doing the course at the time of research	Approximately 4,700	2193	3911
Number of participating students	521	464	516

3.3.2.2.3 Case study participants.

Teachers

Initially, two teachers at each of the three universities were invited to participate in the case study. At University A and University B, they were either recommended by the HOD or selected by the researcher herself. The recommendations were made based on the researcher's criteria for participant selection. At University C, the participants were chosen by the researcher herself based on the information given to her by the HOD. The

information given included the teachers' names, qualifications, email address, phone numbers and teaching schedules. The first selection criteria were the teachers' involvement in TOEIC preparation and the stage of their class in the English programme. The researcher wanted to observe the highest level classes at the time the research was being conducted because research has shown that test impact is the most intense and noticeable when the test date is close (Shohamy et al., 1996; Wall & Alderson, 1993; Zhan & Wan, 2013). However, as the three universities differed greatly in terms of their English programme, a different principle was used at each university with regard to these criteria. At University A, there was no TOEIC preparation course; TOEIC preparation was only a small component of the Business English courses. Therefore, teachers who were teaching Business English 3 were considered. At University B, the first three modules of the English programme were normal general English, and the last one was TOEIC preparation, so only teachers who were teaching this last module were contacted. At University C, all of the modules were TOEIC preparation courses, so like in University B, only teachers teaching the module nearest to the end of the programme were aimed at.

Besides, the researcher planned to select participants on the basis of their age and qualification. The first reason for this intention was that the Phase 1 data showed that there were small to medium correlations between the teachers' age and many effects of TOEIC. In-depth information from the youngest and the oldest teachers could be very useful for explaining the differences of the test impact on them. The second reason was the consideration of the teachers' representativeness. Obviously, qualification is a very important factor that influences teachers' beliefs and teaching. Demographic information from the Phase 1 data showed that at University A and University B, the majority of the teachers held an MA degree (91.3% and 78.1%, respectively), while at University C, just over half of the teachers (53.8%) had an MA. Therefore, the researcher aimed at recruiting two teachers with an MA at University A and University B, and one teacher with an MA and one teacher with a BA at University C. However, due to the timetable clash and some teachers' unwillingness to participate, both of the teachers selected at University C were young. Later, it turned out that young teachers did not know about student learning before the TOEIC-as-exit-test policy, so an older teacher was asked to participate as soon as the researcher could arrange time to observe her class. However, at the time the research was conducted, she was not teaching TOEIC 5 as the other participants were and, as the course was about to end, the researcher could only observe her twice.

Table 3.6 Teacher Case Study Participants (N=7)

		University A	University B	University C	Total
Number of participants		2	2	3	7
Gender	Male	1		1	2
	Female	1	2	2	5
Highest qualification	MA	2	2	1	5
	BA			2	2
Age	20–30	1		2	3
	31–40		1		1
	Above 40	1	1	1	3
English teaching experience (In years)	1–10	1		2	3
	11–20	1	2		3
	21–30			1	1
TOEIC teaching experience (In years)	1–5	1	1	2	4
	6–10	1		1	2
	No Information		1		

Students

Two students from each observed class were selected for the case study. The selection of students from the teacher participants' classes allowed the researcher to observe both the teachers and the students at the same time. Also, as teachers usually know their students well, the teachers could recommend suitable students for the study. The researcher intentionally chose one top and one bottom student in each class because she believed students' English ability could have a strong effect on their preparation for tests. In fact, Green (2006) shows in his predictive model of test washback that participants' perception of test difficulty is one of the factors moderating the intensity of test washback. However, one of the teachers, Lan, refused to recommend any specific students explaining that she did not know her students' ability. Therefore, the researcher had to ask the students in her

class to volunteer. Two students agreed to participate in the study. The total number of students participating in the case study came up to 12.

Table 3.7 Student Case Study Participants (N=12)

		University A	University B	University C	Total
Number of participants		4	4	4	12
Genders	<i>Male</i>	2	3	2	7
	<i>Female</i>	2	1	2	5
Language ability as perceived by teachers	<i>Top</i>	2	1	2	5
	<i>Bottom</i>	2	1	2	5
	<i>No information</i>		2		2

3.4 Reliability and Validity

Reliability and validity are two essential qualities of both quantitative and qualitative research. However, these criteria are interpreted differently in the two types of research, and, accordingly, the strategies that are used to ensure them are also different. As my study involved both qualitative and quantitative methods, both interpretations and types of strategies are discussed below. Descriptions of strategies used to ensure reliability and validity of the current study are also included.

3.4.1 Reliability and validity in quantitative research.

3.4.1.1 Reliability.

In quantitative research, *reliability* refers to the consistency of research results over time, over instruments, and over groups of respondents (L. Cohen, Manion, & Morrison, 2007). Many authors discuss three types of reliability: reliability as stability, equivalence and internal consistency (L. Cohen et al., 2007; Lodico, Spaulding, & Voegtler, 2006). Reliability as stability is “a measure of consistency over time and over similar samples” (L. Cohen et al., 2007, p. 146). Reliability as equivalence is achieved when equivalent forms of an instrument yield similar results. Reliability as internal consistency is the consistency within the instrument, e.g., the consistency between the two halves of an instrument or

among the items of a multi-item scale. However, Lodico et al. (2006) suggested that it was not necessary to conduct all of the types of reliability on a single instrument.

In the current study, internal consistency among the items of the scales was established for the teacher and student questionnaire using the Cronbach alpha. Details of the assessment of the questionnaire reliability are described in Section 3.6.1.4.

3.4.1.2 Validity.

When discussing *validity*, many authors (Lodico et al., 2006) refer to the accuracy of the research instrument. According to them, validity is ensured when an instrument measures what it claims to measure. The following is the description of the types of validity related to survey research, which are most often discussed in research literature: content validity, criterion-related validity, and construct validity.

According to Fraenkel and Wallen (2009), an instrument with content validity is one that contains “an adequate sample of the domain of content it is supposed to represent” and has a clear and appropriate format (p. 150). Content validity for a survey can be assessed by having participants examine it to ensure the comprehensiveness of the coverage and appropriateness of the items. In some cases, a panel of experts is needed to determine whether the content is appropriate or not (Lodico et al., 2006). Criterion-related validity “reflects the degree to which two scores on two different measures are correlated” (Lodico et al., 2006, p. 93). To establish criterion-related validity, researchers usually calculate the correlations of two scores on two different measures. Construct validity shows “evidence that an instrument is accurately measuring an abstract trait or ability” (Lodico et al., 2006, pp. 94–95). Establishing construct validity is a lengthy process, which involves several procedures. Some of them are clearly defining the meaning of the construct based on research or well-developed theories related to the construct and performing factor analysis (Lodico et al., 2006).

In the current study, the content and construct validity of the questionnaire were established through careful and thorough processes of designing and piloting. These processes are described in Section 3.6.1. To assess the content validity of the questionnaires, these were given to different groups of target participants for their comments on different aspects including the content coverage. To obtain their construct validity, the meaning of the constructs to be measured was defined on the basis of a

comprehensive review of literature on theory of language teaching and test impact. Factor analysis was also performed on the collected data. (See Section 3.6.1.)

3.4.2 Reliability and validity in qualitative research.

According to some authors, quality of qualitative research must be assessed by criteria different from those used in quantitative research. Lodico et al. (2006) said that these should “focus on how well the researchers have provided evidence that their descriptions and analysis represent the reality of the situations and persons studied” (p. 273). According to LeCompte and Preissle (L. Cohen et al., 2007), the reason for this difference is that quantitative methods are characterized by the possibility of replication, while the strength of qualitative methods lies in maintaining the uniqueness and idiosyncrasy of situations, and so it is impossible to replicate a qualitative study. Accordingly, instead of the terms reliability and validity, authors suggested terms such as *credibility*, *neutrality*, *confirmability*, *dependability*, *consistency*, *applicability*, *trustworthiness* and *transferability*. The most important terms are discussed below.

3.4.2.1 Dependability.

Many authors such as Lodico et al. (2006) and Lincoln and Guba (2007) considered *dependability* a criterion that parallels *reliability* in quantitative research. “Dependability refers to whether one can track the procedures and processes used to collect and interpret the data. Good qualitative studies will provide detailed explanations of how the data are collected and analysed” (Lodico et al., 2006, p. 275). In this thesis, the data collection methods and data analysis are described in detail in Sections 3.7.2 and 3.7.3, respectively.

3.4.2.2 Credibility.

In qualitative research, *credibility* is often used instead of *validity* (Lodico et al., 2006). According to Creswell and Miller (2000), validity in qualitative research refers to “how accurately the account represents participants’ realities of the social phenomena and is credible to them” (pp. 124–125). They also suggested a two-dimensional framework for selecting validity procedures. The first dimension is the lens or the view used by the researcher. This can be the lens of the researcher, the lens of the participants or the lens of individuals external to the study. The second dimension is the researcher’s paradigm assumptions or worldviews, which can be postpositivist, constructivist, and critical. Depending on the lens and the paradigm used, the researcher can choose from nine

procedures for establishing validity: triangulation; member checking; the audit trail; disconfirming evidence; prolonged engagement in the field; thick, rich description; researcher reflexivity; collaboration; and peer debriefing. Many of those strategies were used to obtain the credibility of this study. These included triangulation, member checking, and thick, rich description, which are described in Sections 3.6 and 3.7.

3.4.2.3 Transferability.

Transferability is the term often used to replace *external validity* or *generalizability* in quantitative research (L. Cohen et al., 2007; Lodico et al., 2006). However, qualitative researchers are “less definitive, less certain about the conclusions they draw from their research” and some are “concerned less with the question of whether their findings are generalizable, but rather with the question of to which other settings and subjects they are generalizable” (Fraenkel & Wallen, 2009, p. 433).

Unlike quantitative research, where the degree of generalizability of findings is decided by the researcher, in qualitative research, this quality is judged by the reader. Therefore, it is demanding for qualitative researchers to provide clear, richly detailed or thick description of the settings, context and background, participants, and interactions so that readers can determine whether the findings are generalizable to their own situation (L. Cohen et al., 2007; Fraenkel & Wallen, 2009; 2006).

Furthermore, according to Flyvbjerg (2011), “generalizability of case studies can be increased by the strategic selection of cases” (p. 306). The author argued that when the object is to get the maximum amount of information on a problem or phenomenon, it might not be the most appropriate to use a representative or random sample. Instead, extreme/deviant, maximum variation, critical or paradigmatic cases would produce more insight.

In the current study, transferability was ensured by detailed description of the settings and participants as well as the purposeful selection of maximum variation cases. The chosen teacher and student cases were different on certain dimensions, e.g., teachers’ qualifications and age/experience and students’ language ability.

3.5 Ethical Considerations

The project received ethics approval from Victoria University of Wellington's Human Ethics Committee on 28 February, 2012. In accordance with the approval, information sheets and consent forms were presented to MOET and the four participating Universities to obtain permission for conducting research at their institutions and to get access to participants. When being recruited for the research, the potential participants were also given information sheets and consent forms to read and sign if they agreed to participate. As for the survey respondents, only information sheets were given, and their voluntary participation in filling out the questionnaires implied their informed consent. In addition, the participants had the opportunity to ask any questions concerning the research and their participation. All of the documents were translated into Vietnamese to ensure full understanding. For a copy of the ethics approval, see Appendix I.

3.6 Phase 1 Methodology

In Phase 1, a survey was carried out to discover the major patterns of the test impact on teaching and learning and the factors that produced the impact and their relationships. In addition, semi-structured interviews with policymakers were also used to find out MOET's and universities' perspectives on the TOEIC-as-exit-test policy.

A survey is a study in which "investigators ask questions about peoples' beliefs, opinions, characteristics, and behaviour" (Ary, Jacobs, & Razavieh, 2009, p. 372). It was suitable for the research purpose for several reasons. First, it can be used to observe patterns of response in and make generalizations about the target population. Second, it provides different types of information including descriptive, inferential, and explanatory. Third, it can ascertain correlations. Last but not least, it is economical and efficient as researchers can collect data on a one-shot basis. All of these reasons were mentioned in L. Cohen et al. (2007, p. 206).

The survey data were collected through questionnaires. Questionnaires were a preferred data collecting technique because it is time and cost effective and more efficient for "identification of underlying patterns than more qualitative and ethnographic methods" thanks to "the size of the respondent group as significant for the quantification of data produced through the questionnaire" (Dunne, Pryor, & Yates, 2005, p. 43). The questionnaires were directly administered to respondents. According to methodologists

(Ary et al., 2009; Fraenkel & Wallen, 2009) self-administered questionnaires ensure high response rate, low cost, and opportunity to explain the research and answer possible questions from respondents.

On the other hand, interviews were more appropriate to elicit policymakers' opinions relating to the research topic because it was impossible for the researcher to get access to a number of respondents which was sufficient for statistical data analysis.

3.6.1 Teacher and student questionnaires.

3.6.1.1 Questionnaire development.

Two questionnaires were developed by the researchers to collect data from teachers and students for Phase 1 of the study. The common purpose of both questionnaires was to identify the patterns of test impact and measure the factors that might shape the impact. The only difference between the questionnaires was that the student questionnaire aimed at eliciting information regarding student learning while the teacher questionnaire focused on teacher teaching. The processes of developing the questionnaires were also similar, starting with question writing, followed by informal piloting with non-target population, questionnaire translating and translation checking, then informal and formal piloting with the target population, and finally internal consistency checking and questionnaire validating. Following is the description of the stages involved in the questionnaire development process.

Stage 1: Question writing.

The first stage of the questionnaire designing process was question writing. At this stage the researcher tried to generate as many questions as possible. First of all, a list of variables that must be measured in order to answer the research questions was produced. Then operational definitions of these variables were formulated. As the purpose of the questionnaires was to identify the impact of the TOEIC test on teaching and learning and the factors that shaped this impact, it was first necessary to formulate operational definitions of variables such as teaching content, learning content, teaching methods (the grammar-translation method, audio-lingual method, communicative approach), learning strategies, student achievement, perception of tests, teachers' and learners' beliefs about language teaching and learning, and contextual factors. This was done by searching literature on theories of language teaching and learning (Brown, 2007; Nunan, 1999;

Richards, 2006; Richards & Rodgers, 2001) to see how these terms were understood in the language teaching field. Other test impact studies and their reviews (Tsagari, 2007) were also used to find out how these variables had been defined and how the test impact on them had been measured.

Questionnaire items were then generated based on the operational definitions of the variables and other researchers' questionnaires (Cheng et al., 2004; Chen, 2002a, 2002c; Griffiths, n.d.; J. Smith & Kovacs, 2011; Stoneman, 2006). The items taken from other questionnaires were slightly modified to suit the context of the present study. The following are some examples.

- | | |
|-------------|---|
| 1) Original | <u>The BCT has</u> little impact on what I teach. |
| Modified | <u>The TOEIC test has had</u> little impact on what I teach. |
| 2) Original | <u>I skip</u> over certain sections in the textbooks because they are less likely to be tested on the BCT. |
| Modified | <u>I have been skipping</u> over certain sections in the textbooks because they are less likely to be tested on the TOEIC test. |
| 3) Original | <u>I have my students do the Ministry of Education mock tests</u> to familiarize students with the <u>BCT</u> . |
| Modified | <u>I have been using mock tests</u> to familiarize students with <u>the TOEIC test</u> . |
| 4) Original | <u>For me, a common university exit language test</u> helps me to set goals for my English learning activities. |
| Modified | <u>The TOEIC test</u> helps me set goals for my English learning activities. |
| 5) Original | <u>For me, a common university exit language test</u> makes me become more aware of the importance of English. |
| Modified | <u>The TOEIC benchmark</u> makes me more aware of the importance of English learning. |

The first drafts of the questionnaires were sent to the researcher's supervisors for their comments and were discussed at face-to-face meetings with them. Many useful comments and suggestions were made about the content of the questionnaires, the scales used, and the wording of the questions. The questionnaires were revised based on the supervisors' comments and suggestions and then sent to them again for new comments. Following the

supervisors' suggestions, the original five-point scales were also changed to seven-point scales for finer distinctions of responses. Some questions were reworded, for examples:

- 1) Original I teach students how to do multiple-choice test effectively. (ambiguous)
Reworded I teach students more multiple-choice test-taking strategies than before.
- 2) Original If there were not the TOEIC test, students would have no motivation to learn the tested skills.
Reworded Without the TOEIC test, students would not be motivated to learn the tested skills.
- 3) Original It is mandating that I get the required TOEIC score before I graduate.
Reworded I must get the required TOEIC score before I graduate.
- 4) Original TOEIC benchmarks have been ensuring graduates' ability to do jobs that require effective use of English
Reworded TOEIC benchmarks have ensured that graduates have the ability to effectively use English in their jobs.

At the end of this stage, both of the questionnaires included 2 parts: Personal Information and Questions. The second part was divided into 2 sections:

Section A: Impact of the TOEIC Test on Teaching/ Learning

Section B: Factors Associated with Teaching/ Learning

There was a total of 10 and 11 big questions in the teacher and student questionnaires, respectively. One of them was an open-ended question and placed at the end of section A. The other questions were closed-ended with a number of seven-point scale items following each question.

Stage 2: Informal questionnaire piloting with non-target population.

The questionnaires were first tried out with several PhD students majoring in Linguistics or Applied Linguistics at Victoria University of Wellington and other universities and an English lecturer in Vietnam. PhD students were asked for help because they had experience in questionnaire designing, and so they could make a lot of contribution. Some Vietnamese speakers of English were deliberately invited to participate in the informal piloting because of their knowledge and understanding of the English teaching and

learning practices in Vietnam, which would enable them to point out problem items for the Vietnamese context and to suggest important ideas to include.

Hard copies of the questionnaires were provided when possible, so that the piloting was as close to reality as possible. The participants first answered the questions as if they were students or teachers at universities that used the TOEIC test for graduation purposes and timed how long it took them to answer the questions. Then they could look at another version of the questionnaires with notes about the variables each group of questions tried to measure and tell the researcher if the questions were clear, valid or necessary and if any other questions were needed. It often took the respondents from 12 to 30 minutes to complete the questionnaires. One participant, however, spent 50 minutes on the student questionnaire. Several types of useful feedback were made:

- 1) There were suggestions to improve the format of the questionnaires, for example, adding a space between subgroups of items to give respondents a rest when they answer the questionnaires, adding three dots to the beginning of each item in question 2 to make it clear that it was part of a statement, using bold letters or capital letters to highlight contrasted points in some places, and numbering all the items continuously for easy reference regardless which questions they belonged to.
- 2) The participants also pointed out possible confusions for the target respondents, for example, the instruction to teachers about how to choose a course to keep in mind when answering the questionnaires and the use of technical terms such as metalanguage, sound instruction, social expectations, and achievement test.
- 3) Some questions needed to be reworded because they were found to be difficult to answer, negative or ambiguous. An example is the question regarding whether teachers had changed their instructional time devoted to skills or teaching and learning activities. Participants argued that the answers to these questions varied depending on the timing and the focus of the lessons. The questions were, therefore, rewritten as follows.

Original: To what extent has your instructional time devoted to teaching and learning activities changed as a result of the introduction of the TOEIC test?

Revised: To what extent has your total instructional time devoted to teaching and learning activities changed as a result of the introduction of the TOEIC test during the whole semester?

Another example is item 2.f

Original: ...I have still been teaching whatever I think is important no matter whether this can improve my students' TOEIC scores or not. (This item sounded negative because it might imply "I don't care about my students' test scores").

Revised: ...I have still been teaching whatever I think is important regardless of the TOEIC test content.

In the next example, the original question had to be rewritten because of its ambiguity:

Original: I have been using different teaching methods.

Revised: I have been using teaching methods that I didn't use before.

- 4) The participants also suggested combining similar items together (for example, "watching TV in English" and "watching movies/ videos in English") and adding more items (for example, adding teaching language functions as a teaching activity). Some of these ideas were used in the later version of the questionnaires.

The respondents' answers also helped the researcher realize wording problems in the questionnaires. For instance, a respondent chose similar values of 2 and 1 for items 10.n and 10.o respectively, while these items were meant to be logically contradictory.

10.n The number of students in my class prevents me from organizing certain teaching and learning activities.

10.o The number of students in my class doesn't affect my teaching plan.

Later clarification with the respondent showed that he had thought item 10.o was asking about the process of lesson planning. He then suggested changing it to "The number of students in my class doesn't affect my lesson plans/ what I plan to do in class."

Stage 3: Questionnaire translating and translation checking.

The questionnaires were translated into Vietnamese by the researcher herself. The translations together with the English versions were sent to a Vietnamese PhD student in Australia and a very experienced English lecturer in Vietnam for checking. After the translations were revised according to the comments and suggestions received, they were shown to three of the researcher's family members, who were native speakers of Vietnamese and were not experts in the field of language teaching, to check if the questions sounded natural, and clear to them and if the technical terms were easy to understand. These family members were also asked to make suggestions for improvement. As a result, the title of the questionnaire was reduced to “Tác động của bài thi TOEIC” (“The Impact of the TOEIC Test”) because the Vietnamese equivalent of “The Impact of the TOEIC-as-exit-test” was long and sounded clumsy; the introduction paragraph was also shortened and made more general to make it more user-friendly; more precise or more comprehensible Vietnamese equivalents were adopted for some jargons such as mock tests, cramming, learning strategies; and more natural ways of translating some expressions were chosen.

Stage 4: Informal questionnaire piloting with target population.

In March and April, 2012 the questionnaires were piloted on the target population. The informal piloting included two steps. Because it was impossible for the researcher to get many teachers to try out the questionnaire, only one teacher was involved at each of these steps. In the first step, the researcher tried out the questionnaires with individual participants with each session lasting for approximately one hour and 15 minutes. The researcher gave each respondent a questionnaire and asked them to answer it in her presence item by item. After each item, she asked them what their answer was and why they chose that answer. She also asked them how they understood the question and what their suggestions for revising it were if it was confusing or unclear to them. Notes were taken by the researcher during the sessions.

Most questions were clear and easy to understand to the participants. The students only asked for clarification for some items. One example is the question about their experience learning English where they were not sure whether this was about the time since they had started learning English at school or since they had started learning English at home

(Nowadays many children begin learning English before the starting of the foreign language program at school). The teacher also needed clarification for only some terms such as information-gap activities, information gathering activities, and test format. Later, an example of each of the teaching activities mentioned was included in the questionnaires. The term “test format” was changed to “test tasks” in both the teacher and students questionnaires. Besides, the teacher participant also suggested words and expressions that were more natural or more culturally acceptable, such as “mong muốn” instead of “yêu cầu” for “expect” in item 82, “không có tác dụng” instead of “không thể giúp tôi” for “not able to prepare” in item 19, and “Xin Thầy/Cô vui lòng” instead of “Xin vui lòng” for “please” used when addressing teachers. Her answers also helped spot some misplaced definitions of scale values in the translation.

In the second step, the revised versions of the questionnaires were tried out with some other respondents. This time, the researcher asked a group of three students to complete the whole questionnaire at the same time before her. They were also asked to mark the questions that they found difficult to understand or to answer and ask the researcher for clarification on the spot or after they had finished the questionnaire. The time for the questionnaire completion was also measured. Only a couple of more minor clarifications were needed. As for the teacher questionnaire, after several meeting cancellations by the teacher, the researcher had to send her the questionnaire together with instructions for completing it so that she could work on it on her own. The researcher argument was that the small number of teachers involved Stage 4 did not affect the task quality because, in fact, no further revision was needed after the second try-out.

Stage 5: Formal questionnaire piloting with target population.

The questionnaires were formally piloted on other teachers and students at the same university as the one in Stage 4. The researcher administered the student questionnaire to the students in class and had them complete it in her presence so that she could give them clarification if necessary. Only two students asked for minor clarifications.

As for the teacher questionnaire, the researcher distributed them and collected them during the teachers’ break time. Some teachers completed the questionnaire in her presence, but no questions were asked. As the researcher had decided to survey only teachers who had

taught TOEIC or were teaching TOEIC, the instruction for respondents to choose a type of course to keep in mind when answering the questionnaires were no longer needed.

(See Table 3.8 for information about the participants involved in the questionnaire piloting and translation checking process.)

Table 3.8 Participants Involved in Questionnaire Piloting and Translation Checking

Stages in questionnaire development		Student questionnaire	Teacher questionnaire
Stage 2: Informal piloting with non-target population		<ul style="list-style-type: none"> - 5 PhD students including 4 Vietnamese students - 1 English lecturer in Vietnam 	<ul style="list-style-type: none"> - 7 PhD students including 2 Vietnamese students - 1 English lecturer in Vietnam
Stage 3: Questionnaire translation checking		<ul style="list-style-type: none"> - 1 Vietnamese PhD student - 1 experienced English lecturer in Vietnam - 3 non-expert native speakers of Vietnamese 	<ul style="list-style-type: none"> - 1 Vietnamese PhD student - 3 non-expert native speakers of Vietnamese
Stage 4: Informal piloting with target population	Step 1	<ul style="list-style-type: none"> - 2 students 	<ul style="list-style-type: none"> - 1 teacher
	Step 2	<ul style="list-style-type: none"> - 3 students 	<ul style="list-style-type: none"> - 1 teacher
Stage 5: Formal piloting with target population		<ul style="list-style-type: none"> - 97 students 	<ul style="list-style-type: none"> - 13 teachers

Stage 6: Internal reliability checking and questionnaire validating

The purpose of this stage is to produce shorter versions of the questionnaires by removing items that did not load on any factor or that caused low internal reliability of scales. Responses from 97 students participating in stage 5 were used for data analysis at this stage. Concerning teachers' data, relevant responses from stage 4 were also included for analysis in this stage because the teacher questionnaire was only slightly revised after step 1, stage 4: one question was removed, one question was added, and some expressions were reworded. The total number of teacher answers used for stage 6 was, therefore, 15.

Factor analysis and internal consistency analysis were done for the student questionnaire, but only internal consistency analysis was conducted for the teacher questionnaire due to the small number of cases (15). Because of the thesis word limit, these analyses are not described in detail here. However, they were done in the same way as those performed on the questionnaire data collected in the main study (see Section 3.6.1.4).

Table 3.9 Description of Teacher Questionnaire

	Types of questions/ items	Focus of questions	Items
PART 1	<ul style="list-style-type: none"> - Boxes to tick - Spaces to complete 	Personal information (gender, age, highest qualification, experience teaching English, Experience teaching TOEIC)	1–5
PART 2			
SECTION A			
Question 1	<ul style="list-style-type: none"> - Closed-ended - Seven-point scales 	Test impact on feelings and attitudes towards teaching TOEIC	1–6
Question 2		Test impact on content of teaching	7–15
Question 3		Test impact on teaching methods	16–18
Question 4		Changes in instructional time devoted to skills	19–25
Question 5		Changes in instructional time devoted to teaching and learning activities	26–34
Question 6	Open-ended	Other impact on teaching	
SECTION B			
Question 7	<ul style="list-style-type: none"> - Closed-ended - Seven-point scales 	Perception of the TOEIC test	35–55
Question 8		Beliefs about language teaching and learning	56–66
Question 9		Test pressure	67–72
Question 10		Context factors	73–85

Table 3.10 Description of Student Questionnaire

	Types of questions/ items	Focus of questions	Items
PART 1	<ul style="list-style-type: none"> - Boxes to tick - Spaces to complete 	Personal information (gender, experience learning English, experience preparing for TOEIC, year at university, major, city/town where students went to high school)	1–6
PART 2			
SECTION A			
Question 1	<ul style="list-style-type: none"> - Closed-ended - Seven-point scales 	Test impact on learning and attitudes towards learning English/ TOEIC	1–7
Question 2		Test impact on content of learning	8–16
Question 3		Test impact on learning activities	17–22
Question 4		Changes in amount of time devoted to skills	23–29
Question 5		Changes in amount of time devoted to learning activities	30–38
Question 6		Test impact on achievement	39–44
Question 7	<ul style="list-style-type: none"> - Open-ended 	Other test impact on learning	
SECTION B			
Question 8	<ul style="list-style-type: none"> - Closed-ended - Seven-point scales 	Perception of the TOEIC test	45–59a
Question 9		Beliefs about learning languages	60–66
Question 10		Test pressure	67–72
Question 11		Context factors	73–80

The final versions of the student and teacher questionnaires had the same structure (see Table 3.9 and Table 3.10). They included two parts. Part 1 asked respondents for their personal information. Part 2 was the main part of the questionnaires which was composed of two sections: Section A – impact of the TOEIC test on respondents’ learning (student questionnaire)/teaching (teacher questionnaire) and Section B – Factors associated with

English learning (student questionnaire)/English teaching (teacher questionnaire). (See Appendices A and B for the complete questionnaires.)

3.6.1.2 Data collection procedures.

Before administering her questionnaires to students and teachers, the researcher had to obtain permission from the universities. First, she contacted the people in authority at the universities by phone, briefly explained the research project to them, and asked them for a meeting. Then, she visited the universities, met the universities' representatives, presented them with the information sheets, consent forms and her research plan, and requested their permission for data collection. The questionnaire administration was carried out after the researcher got the permission from the universities.

3.6.1.2.1 Student questionnaires.

The student questionnaires were distributed to students in class by the researcher, her assistant, or the teachers in charge of the classes. A total of 40 classes were selected at the three universities. First of all, the purposes of the project and the kind of help needed were briefly explained to the students. Then they were invited to participate in the project. Each student was given an information sheet about the project and a copy of the questionnaire to complete on the spot. The students were offered opportunities to ask questions about the project or the questionnaires if they wanted to. In case the questionnaire was not administered by the researcher herself, a detailed instruction sheet was prepared to inform the teachers/researcher's assistant about the steps to follow when administering the questionnaires and points to clarify to students.

3.6.1.2.2 Teacher questionnaires.

At Universities B and C, the teacher questionnaires were distributed to the teachers in class. Before the break time, the researcher and her assistant came to classes, briefly explained the project to the teachers, and invited them to participate in the project. None of the teachers refused participation. Each teacher was then given an envelope containing an information sheet, a copy of the questionnaire, instruction on steps to follow, and an empty self-adhesive envelope for them to put their answer in. After the break time, the researcher and her assistant came back to collect the answers. If the researcher administered the questionnaires to both the teacher and students of a class, the teacher completed their answers at the same time as their students. At University A, the questionnaires were

distributed to the teachers by the Head of the Foreign Language Department during the Department's meeting. Fifteen teachers returned their answers on the day and eight other teachers returned them to the Department later.

3.6.1.3 Screening and cleaning data.

It is vital to screen and clean data before doing statistical analysis to avoid possible trouble later (Pallant, 2011). The researcher, therefore, examined the collected questionnaire data very carefully to make sure that the information used for statistical analysis was valid and truthful, and that this information was recorded correctly. First of all, to ensure the validity of the data, all problematic responses were discarded. Several types of problems were found:

- 1) Respondents only answer questions about personal information on the first page of the questionnaire.
- 2) Respondents used only one or two values to answer all the items.
- 3) Respondents selected their answers using repeated patterns of values in all or several parts of the questionnaires. For example, they selected 1 for item 1, 2 for item 2, 3 for item 3, 4 for item 4 for the first four items and then repeated this pattern for the next four items and so on.
- 4) Two or three respondents who were sitting close to each other gave the same responses. The responses were identical for either the whole questionnaire or for some parts of the questionnaire. Those questionnaires which were determined to have 40% or more consecutive identical responses were discarded.
- 5) Respondents gave an answer that was impossible to be true. For instance, some students indicated that they had been studying TOEIC for more than 11 years, while the TOEIC test was only introduced to Vietnam less than 11 years prior and had only been popular for about 6 years. In addition, considering students' age and English learning experience, they could not have been studying TOEIC for such a long time.

When problems 1, 2, 3, and 4 were found, the whole questionnaires were discarded because the respondents obviously did not answer the questions seriously and their responses, if used, would damage the validity of the data. In total, 56 questionnaires (3.7%) were excluded from the 1,501 student questionnaires collected. All the teacher

questionnaires were useable. Problem 5 was treated as missing information, and the other parts of the questionnaires were still used.

Second, the data files were checked for errors in entering data. After the information was entered into SPSS, descriptive statistics such as frequencies, maximum and minimum values, and means were produced to spot out-of-range values, for example a value of 3 for gender, when “male” was specified as 1 and “female” was specified as 2, or a value of 9 in a seven-point scale. Next, the input data from every three cases were checked for accuracy with the help of the researcher’s family members. These people randomly chose chunks of responses from the questionnaires and read them to the researcher, who checked these values on the computer. Very few errors were found.

3.6.1.4 Questionnaire validation and internal consistency check.

The purpose of the questionnaire validation is to ensure the construct validity of the variables measured in the questionnaires. Internal consistency was calculated to check their reliability. These processes were the same for both the teacher and student questionnaires. Due to a large number of items and variables in both of them, factor analysis had to be performed many times. To avoid repetition, the common procedure, techniques, and principles that were used to validate the questionnaire are described first, and the statistics for each variable is reported later. The following are the steps involved in the validation process.

(1) Assessment of the suitability of the data for factor analysis

SPSS 18 was used to perform principal component analysis (PCA) on all the items in each question or group of questions to discover the latent factors or variables that the items measured. Before the items were subjected to PCA, the suitability of data for factor analysis was evaluated by examining the correlation coefficients, the Kaiser-Meyer-Olkin value and the statistical significance level that Bartlett’s Test of sphericity reached. According to Pallant (2011) PCA can be performed only when there are many coefficients of .3 or above, the Kaiser-Meyer-Olkin value is equal or greater than .6, and Bartlett’s Test of sphericity is significant ($p < 0.5$). The student questionnaire analysis showed that there were many correlation coefficients of .3 or above, Bartlett’s Test of sphericity was always significant with $p = .000$, and the Kaiser-Meyer-Olkin values were in the range from .633 to .838. Similarly, the teacher questionnaire analysis revealed that there were many

correlation coefficients of .3 or above, Bartlett's Test of sphericity was always significant with $p=.000$, and most of the Kaiser-Meyer-Olkin values ranged from .627 to .823. Only two values were .552, which would be .6 if rounded. This suggested that the data from both questionnaires was suitable for factor analysis.

(2) Factor extraction

Principal component analysis (PCA) was used for factor extraction. The number of factors was determined based on three criteria. The first one is Kaiser's criterion: Only factors having an eigenvalue of 1.0 or more were retained for further investigation. The second one is Scree test. As eigenvalues often suggest too many factors, a Scree test is usually used on top of them. As a rule, only factors above the elbow were retained because they "contribute the most to the explanation of the variance in the data set" (Pallant, 2011, p. 184). The final criterion is parallel analysis: Only those eigenvalues that were above the corresponding values from a randomly generated data set of the same size as my sample were retained.

(3) Internal reliability check

After the number of factors had been determined, internal consistency analysis was conducted for each factor to identify items that caused low Cronbach's alpha. These items were removed one by one, starting with the one most negatively affecting Cronbach's alpha, until a satisfactory alpha was reached. The satisfactory alpha values were determined by taking into consideration several principles. First of all, the alpha must be as high as possible. Generally, an alpha of .7 or more is acceptable (Pallant, 2011). Second, the variables must be made up of three or more items. Third, the items clumping together as a variable must make up a theoretically meaningful unit. However, there were some rare exceptions to these principles, which are mentioned later in the questionnaire validation results sections.

Above is a description of the general validation process used for the study. However, the actual process was not always so straightforward. Usually, different solutions had to be experimented with to find the most satisfactory one, which met the requirements for meaningful constructs, the number of items making up the variables, and the internal consistency of the items. Sometimes, an item was removed from a factor because it did not conceptually belong even though its removal lowered the Cronbach's alpha. This was the

case with the direct test preparation variable of the teacher questionnaire with item 17 excluded.

Student questionnaire validation results

All the 81 Likert-like items in the student questionnaire were subjected to PCA. The results yielded 18 composite variables, three of which were composed of only two items. All of the variables achieved alpha values above .7, except for three items – time devoted to tested skills (.677); test purpose-tested skills (.683); and test difficulty (.612). The alpha values for the time devoted to tested skills and test purpose-tested skills were only slightly lower than .7. The test difficulty variable had lower alpha probably due to participants' ignorance of the reverse wording of the items measuring the variable. Those items that did not cluster with other items were either discarded (six items) or retained as single item variables (six items). Table 3.11 shows the results of factor analysis of the student questionnaire.

Teacher questionnaire validation results

All the 85 Likert-like items in the teacher questionnaire were subjected to PCA. The results yielded 18 composite variables, three of which were composed of only two items. All of the variables achieved alpha values above .7, except for positive feelings (alpha=.635); time devoted to traditional activities (alpha=.587); test difficulty (alpha=.678); belief in communicative approach (alpha=.655); belief in necessity of direct test preparation (alpha=.678); and context factors-opportunities (alpha=.656). Thirteen items did not load with other items. They were either discarded (nine items) or retained as single item variables (four items). Table 3.12 shows the results of factor analysis of the teacher questionnaire.

Table 3.11 Results of Factor Analysis of Student Questionnaire

Questions	Original items	Factors found	Items making up the factor	Cronbach's alpha
Question 1: Impact on feelings and attitudes	1, 2, 3, 4, 5, 6, 7	Motivation	3, 4, 6, 7	.784
		Stress	2	
Questions 2 and 3: Impact on learning content and learning activities	8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22	Direct test preparation activities	10, 11, 13, 14, 15, 21, 22	.819
		Reduction of content	8, 9	.772
		Overall management of learning	17, 18, 19	.715
Question 4: Impact on time devoted to skills	23, 24, 25, 26, 27, 28, 29	Time devoted to non-tested skills	24, 26, 29	.713
		Time devoted to tested skills	23, 25, 27, 28	.677
Question 5: Impact on time devoted to learning activities	30, 31, 32, 33, 34, 35, 36, 37, 38	Time devoted to language use activities	30, 31, 32, 33, 34	.796
		Time devoted to form-focused activities	35, 36, 37, 38	.792
Question 6: Impact on achievement	39, 40, 41, 42, 43, 44	Achievement	39, 40, 41, 42, 43, 44	.791
Question 8: Perceptions of TOEIC	45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 59a	Test value	51, 56, 57, 58, 59	.809
		Test difficulty	52, 53, 54	.612
		Test purpose-non-tested skills	45, 48, 50	.819

Questions	Original items	Factors found	Items making up the factor	Cronbach's alpha
		Test purpose-tested skills	46, 47, 49	.683
		Test format	59a	
Question 9: Beliefs about language teaching and learning	60, 61, 62, 63, 64, 65, 66	Belief in traditional methods	61, 62, 63	.759
		Belief that language learning is for communication	60	
		Belief in necessity of learning all four skills	64	
		Belief that overall improvement results in improved test scores	65	
		Belief that test-taking strategies are necessary to improve test scores	66	
Question 10: Test pressure	67, 68, 69, 70, 71, 72	Test pressure	67, 68, 69, 70, 71, 72	.665
Question 11: Context factors	73, 74, 75, 76, 77, 78, 79, 80	Availability of resources	73, 74, 75, 76	.761
		Teacher's influence	77, 78	.847
		Time constraints	79, 80	.855

Table 3.12 Results of Factor Analysis of Teacher Questionnaire

Questions	Original items	Factors found	Items making up the factor	Cronbach's alpha
Question 1: Impact on feelings and attitudes	1, 2, 3, 4, 5, 6	Positive feelings	1, 4, 5	.635
		Anxiety	2	
Questions 2 and 3: Impact on teaching content and teaching activities	7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18	Direct test-preparation	9, 10, 13, 14	.847
		Extent of impact	15, 18	.7
Question 4: Impact on time devoted to skills	19, 20, 21, 22, 23, 24, 25	Time devoted to tested skills	19, 21, 23, 24	.818
		Time devoted to non-tested skills	20, 22, 25	.723
Question 5: Impact on time devoted to learning and teaching activities	26, 27, 28, 29, 30, 31, 32, 33, 34	Time devoted to communicative activities	30, 31, 32, 33, 34	.907
		Time devoted to traditional activities	26, 27, 28	.587
		Use of Vietnamese	29	
Question 7: Perceptions of TOEIC	35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55	Test value	41, 46, 48, 49, 50, 51, 55	.903
		Test purpose – tested skills	36, 37, 39	.727
		Test purpose – non-tested skills	35, 38, 40	.812
		Test difficulty	43, 44, 45	.678
		Test format	42	

Questions	Original items	Factors found	Items making up the factor	Cronbach's alpha
Question 8: Beliefs about language teaching and learning	56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66	Belief in communicative approach	56, 60, 61, 63	.655
		Belief in traditional methods	57, 58, 59	.741
		Belief in the necessity of direct test preparation	64, 65	.678
		Belief in overall improvement	66	
Question 9: Test pressure + items 52, 53, 54	67, 68, 69, 70, 71, 72, 52, 53, 54	Consequences for teachers	71, 72, 53, 54	.808
		External expectations	67, 68, 69, 70	.707
Question 10: Context factors	73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85	Constraints	82, 83, 78, 79	.743
		Opportunities	73, 75, 76, 80, 81	.656

3.6.1.5 Data analysis

Three statistical techniques were used to analyse the questionnaire data. First of all, descriptive statistics, including the means and the standard deviations, on the variables were obtained. These statistics showed the strength of the perceived impact of TOEIC on different aspects of teaching and learning and the factors potentially influencing the test impact on teaching and learning.

Second, bivariate correlation analysis was performed to discover the relationships between the factors potentially influencing the test impact on teaching and learning and the test impacts. The results of the analysis showed which of the potentially influencing factors were actually involved in shaping the test impact and how much they were involved.

Third, independent-sample t-test was conducted to compare the impact of the TOEIC test and the factors potentially influencing the test impact in the large city universities with those in the provincial university.

3.6.2 Interviews with policymakers.

3.6.2.1 Instrument.

Policymakers including heads of English departments and MOET officials were interviewed. The interviews had the purpose of finding out their perspectives with regard to the TOEIC-as-exit-test policy. The interviews with both groups were semi-structured and had two parts (see Appendices C and D). The first part asked interviewees about their involvement in universities' foreign language teaching policy and their views on the goals of language teaching. The second part of the interview asked them about their rationales for choosing TOEIC to be a language exit test or supporting this use of TOEIC and the intended impact of the policy. As these were semi-structured interviews, the actual number of questions, the wording of the questions, and their order varied slightly depending on the interviewees' answers.

3.6.2.2 Data collection procedures.

The researcher contacted the interviewees by email or telephone to ask for an interview and arrange a time for it. A brief description of the project was given on the phone or in the email. An information sheet on the project and a consent form were also sent to the interviewees before the interviews in most cases when their email addresses were available.

In addition, the researcher obtained permission from the Ministry of Education and Training for interviewing MOET officials. She sent the Head of the Higher Education Department a package of documents including the information sheet for MOET, consent form, introduction paper from the researcher's institution in Vietnam, and the researcher's letter of request for conducting interviews with MOET officials.

Almost all of the interviews were conducted face-to-face; only one of them was carried out over the telephone. At the interviews, the interviewees were presented with hard copies of the information sheet and consent form. The interviewees also had an opportunity to discuss the research topic and ask questions for further information or clarification. Most of the interviews were audio-recorded; however, for two of them, only notes were taken because the interviewees did not permit recording.

3.6.2.3 Transformation of data and member checking.

The 14 interviews resulted in only 12 recordings because two of the participants did not allow the researchers to record them. All the recorded interviews were transcribed verbatim and summarized after Phase 1 data collection had been completed. The transcripts and summaries, together with some questions for clarification if needed, were emailed to participants for member checking. The two unrecorded interviews, however, were immediately transcribed (based on the researcher's notes and memory), and the transcripts were sent without a summary to the interviewees for feedback once they had been done to avoid memory decay issues.

Twelve of the participants replied with their feedback either right after the researcher sent them the transcripts and summaries or after her reminder. Four of them confirmed their complete agreement; the others made some changes to the verbatim transcripts and/or the summaries. These changes were categorised into nine groups (see Table 3.13). There were five changes for which the researchers could not work out the reasons because the interview was not audio-recorded. It could be that the researcher did not manage to take notes of all the details that the interviewee gave, or the details were added later by the interviewee herself. This group was, therefore, labelled "unknown reasons."

Table 3.13 Results of Member Checking with Policymaker Participants

Types of changes	No of changes	Examples	
		Original transcripts/ summaries	Changed transcripts/ summaries
• Participants corrected typos and spelling errors	5		
• Participants rephrased themselves to tone down their expressions.	3	- The first cohort of students... was very highly evaluated by employers.	- The first cohort of students... was highly evaluated by employers.
• Participants rephrased themselves to express their ideas more accurately, clearly, and grammatically correctly.	9	- of <u>foreigners</u> - *students socialised	- of <u>English speaking countries</u> - socialising learning
• Participants corrected facts and details they originally stated.	3	- <u>Phòng Đào tạo</u> đã tổ chức (The <u>Training Department</u> have organized) - University <u>X</u>	- <u>Trường</u> đã tổ chức (The <u>University</u> has organized) - University <u>D</u>
• Participants added more details	13	- ...the member universities teach foreign language [subjects] called General English, called Basic English. - the TOEIC benchmark	- ...the member universities teach foreign language [subjects] called General English, called Basic English, and <u>ESP (English for Specific Purposes)</u> . - The TOEIC benchmark or equivalent
• Participants corrected the researcher's misunderstanding.	1	- <u>Acting Head</u> of the Language Practice Department	- <u>Head</u> of Language Practice Department
• Participants removed information they had originally stated.	1		
• Participants added details from the verbatim transcripts into the researcher's summaries.	4	- ...students of previous cohorts who <u>returned to retake</u> a course found...	- ...students of previous cohorts who <u>had to retake the course together with the current ones</u> found...
• Unknown reasons	5	- ... <u>they can communicate fluently</u> . - ... Concerning linguistic aspect, TOEIC vocabulary is business related...	- ... <u>they can speak and listen at a normal level</u> . - ... Concerning linguistic aspect, TOEIC vocabulary is business related, <u>suitable for students of Economics majors</u> .
Total	44		

Two participants did not respond to the researcher's request for member checking even after they were sent a reminder. This reminder, however, also informed them that the researcher would assume that they agreed with the transcripts and the summaries if she did not hear from them after a certain date.

3.6.2.4 Data analysis.

A quantitative approach to the analysis of the interviews with policymakers was adopted. Thus, the interview transcripts were coded into categories, using NVIVO 10. After that, their frequency was counted and reported.

Approaches to coding are “determined by the nature of the research” (Gu, 2014, p. 79). In the current study, both the top-down and bottom-up coding approaches were used due to the semi-structuredness of the interviews. In the interviews with policymakers, the questions mostly aimed at eliciting information of pre-determined categories, e.g., the purpose of language teaching and the rationales for using TOEIC score requirements. Therefore, a list of broad categories or codes was developed as a starting point. During the coding process, subcategories or subcodes were added under these categories as they were spotted. For instance, several subcategories such as communication in the workplace, academic purposes, and reading specialized materials were added under a broad predetermined category called purpose of language teaching. However, the researcher was always open to new categories if they arose. An example of a broad category that emerged as a result of the bottom-up approach is test impact on university's policies.

The coding process was an iterative one. After the first three interviews were coded, the codes were checked and refined to make sure that the code labels truly reflected the coded content, the codes were organized into a logical hierarchical structure, and there were no overlapping categories. This revising process was repeated several times until all the interviews were coded. Finally, every category was checked one more time to make sure no response from participants was missed out.

3.7 Phase 2 Methodology

3.7.1 Research design.

The purpose of Phase 2 was to seek in-depth information about the patterns of test impact found in Phase 1 and to explain the way these patterns were generated. A case study was

conducted to meet this purpose. A case study is defined as a qualitative approach that focuses on a single unit in a specific setting, and information on the unit is gathered through multiple sources and perspectives (Ary et al., 2009; Creswell, Hanson, Plano, & Morales, 2007; Lodico et al., 2006). This approach was chosen for the current investigation because “case studies may be carried out to follow up and put flesh on the bones of a survey” (Bell, 2005, p. 10), and it is “ideal when a ‘how’ or ‘why’ question is being asked about a contemporary set of events over which the researcher has no control” (Gray, 2004, p. 124). Multiple cases were selected because they often yield more compelling results, which are “more likely to lend themselves to valid generalization” (Fraenkel & Wallen, 2009, p. 431).

3.7.2 Data collection methods.

Multiple data collection methods were used to triangulate data. Interviews and classroom observations were employed to collect information on all cases. As for students, weekly journals were an additional method. The interviews enabled the researcher to capture the participants’ perspectives, but the observations and journals allowed her to see the fact, i.e., what they actually did. The following is the description of the instruments, their designing process, and data collection procedures used for each method.

3.7.2.1 Interviews.

Semi-structured interviews were carried out with the teacher and student cases. They were chosen for the case study because “An interview has the advantage of supplying large volumes of in-depth data rather quickly” (Ary et al., 2009, p. 439). Also, they can ensure that all important topics were discussed with all of the participants, and, at the same time, there were opportunities for participants to have some control over what they said and for the researcher “to ‘probe’ for more detailed responses” (Gray, 2004, p. 214). Interview schedules were prepared in advance. The questions for the interviews were all open-ended and generated on the basis of Phase 1 results (see Appendices E and F). The interview schedules were sent to the researcher’s supervisors for comments. Some more questions were added, for example, questions asking students to rate their own English level on a scale from 1 to 10 and questions about student family background, personality, and job aspiration. Some ambiguous questions were reworded, for example:

Original question: Have you been spending more time on learning the tested skills

Reworded question: Have you been spending more time on learning the tested skills than in high school? and Have you been spending more time on learning the tested skills than on the non-tested skills?

Both the teacher and student interview schedules included five main parts. Part 1 aimed at eliciting personal information such as participant's age, family background, education, teaching/learning experience, personality, job aspiration/career goals. Part 2 asked participants about their beliefs about language teaching and learning. Part 3 included questions about the perception of the TOEIC test. Part 4 tried to get information about the influence of context factors on their teaching and learning. Part 5 obtained information about the test impact on participants' teaching/learning and impact on student learning in the case of the teacher interview.

As for teachers, apart from the questions described above, there were also post-observation interview questions. These questions aimed at finding out whether the TOEIC test influenced teachers' specific instructional decisions observed in class, and whether any other factors were involved in those decisions. For each point of interest, the following set of possible questions was asked:

1. Why did you do this [thing the teacher did in the observed class]?
2. Have you always been doing this?
3. How long have you been doing this?
4. What did you do in the past?
5. Why did you change?
6. Did anyone or anything contribute to this change?

As the interviews were semi-structured, during the interview process, the actual order of the questions and wording were adapted to suit the participants' answers. The interviews with the teachers were usually carried out right after the classroom observations. The interviews with the students were conducted at a time convenient to both the students and the researcher. All of them were audio-recorded.

3.7.2.2 Observations.

Observations were used as a data collection method to triangulate the data from the interviews because "what people do may differ from what they say they do" (Fraenkel & Wallen, 2009, p. 396). The researcher assumed that in class, Vietnamese students only did

what their teachers told them to do, and so it was impossible to observe the impact of the test on the students' learning, so the classroom observations could only help reveal the students' attitude towards learning, the level of their participation, and their English proficiency. The classroom observations were, therefore, mainly used to discover the impact of the TOEIC test on teachers' teaching.

As the objects of the observations had been clearly identified based on a considerable number of previous test impact studies, the structured observation type was chosen. An observation scheme was, therefore, prepared for classroom observation (see Appendix G). The first part of the scheme was used to record information about the class being observed, e.g., the university, the teacher's name, and the date of the observation.

Previous studies have showed that a test can have impact on types of teaching and learning activities, types of participant organization, skills taught, and materials used, so these points became the main focus of the observation scheme. Several columns were used to record the time when an activity/episode started and ended, type of activity, type of participant organization, skills being taught, materials being used, teacher's activities, students' activities, other information, and questions to be asked after the observations. After the scheme was designed, it was sent to the researchers' supervisors for comments. Next, it was tried out with an English Proficiency Class at the English Language Institute at Victoria University of Wellington. The researcher observed the class for 45 minutes, using the scheme to record what happened in the classroom. After the piloting, the position of some columns was changed and horizontal lines were added to make the recording process easier.

3.7.2.3 *Weekly journals.*

A journal sheet was to be completed by students every week. The sheet was designed by the researcher herself. It aimed at finding out what and how students learned, and whether the TOEIC had any influence on their learning and their feelings about their learning. The sheet was also sent to the researcher's supervisors for comments. Before the journal sheet was used, it was tried out with three students at University A. The students said the questions were clear to them, but it was hard for them to know how much time they had spent on each item every week. Therefore, a table was added to the end of the sheet, so that students could record their learning time every time they studied English and then transfer the total time to their answers at the end of the week.

The journal sheets included 7 questions (see Appendix H). All the questions were open-ended, except for question 1, which provided students with a list of skills, so that they could check the one that were relevant to them. A sample answer was provided for question 3 to help students understand more clearly what types of information were required.

The researcher also carried out a short training session with the participants right after they were recruited. The students had no difficulty understanding how to complete the journal sheet. In fact, all of them did their job well. At the end of each week, they returned their sheet to the researcher and got another hard copy of the sheet.

Table 3.14 Description of Journal Questions

Questions	Information elicited
1	<ul style="list-style-type: none"> - Skill(s) students have been learning - Total time students spent on each skill - Reasons for learning the skill(s)
2	<ul style="list-style-type: none"> - Material(s) students have been using - Reasons for using the material(s)
3	<ul style="list-style-type: none"> - Activity(ies) students have been doing to learn English - Time students have been spending on the activity(ies) - Reasons for doing the activity(ies)
4	<ul style="list-style-type: none"> - Important things students have learned
5	<ul style="list-style-type: none"> - Students' satisfaction/dissatisfaction with their achievement(s)/underachievement(s)
6	<ul style="list-style-type: none"> - Difficulty(ies) students have had to deal with
7	<ul style="list-style-type: none"> - Students' feeling(s) about their learning - Reasons for their feeling(s)

3.7.3 Transformation of data and member checking.

The interviews with teachers and students resulted in 882 and 1,259 minutes of recording, respectively. All the interviews were transcribed and then summarized. The detailed summaries of the transcripts were sent to the participants for member checking together with additional questions or questions for clarifications if there were any. Almost all of the respondents replied and answered the additional questions or questions for clarification

except for one teacher who confirmed only one of the four summaries sent to her, one student who did not reply at all, and one student who only confirmed two of the three summaries sent to him. Those participants who did not reply to the member checking requests were sent a reminder and informed that the researcher would assume that they agreed with the summaries if she did not hear from them after a certain date. Very few changes to the summaries were made by the participants. The participants gave additional information themselves (five times); corrected themselves (three times); corrected the wordings of the summaries (four times); and corrected the place names in one summary (two times).

A total of 22 observations were made by the researcher at the three universities (see Table 3.15). Each teacher at University A was observed four times, and each teacher at Universities B and C, except for Kim, was observed three times. Each of the teachers was observed for a period of time that covered at least one completed textbook unit. The observation schemes were used for analysis.

Table 3.15 Information on Classroom Observations.

Teachers	Universities	Types of courses observed	Levels observed	No of sessions observed	No of minutes observed
Phuong	University A	BE	BE3	4	635
Thoai	University A	BE	BE3	4	499
Lan	University B	TOEIC preparation	English 4	3	483
Hoa	University B	TOEIC preparation	English 4	3	393
Quyen	University C	TOEIC preparation	TOEIC 5	3	368
Phong	University C	TOEIC preparation	TOEIC 5	3	454
Kim	University C	TOEIC preparation	TOEIC 3	2	191
Total				22	3,023

3.7.4 Data analysis.

3.7.4.1 Interviews.

Interviews with teachers and students were analysed both quantitatively and qualitatively because the researcher was not only interested in identifying the patterns of test impact across cases, but also in explaining the complexity of the phenomenon. Thus, the interviews were reduced to codes and their frequencies were counted. The number and frequency counts gave the researcher the confidence in her conclusions about the patterns of test effects. Apart from coding and frequency counts, each individual case was analysed as a whole. This gave the researcher a holistic picture of the case and a deeper insight into the complex interaction of factors which determined the test impact on the case. An extensive description of representative cases with the key issues was presented to illuminate their complexity.

Like the coding of the interviews with policymakers, the coding of the interviews with teachers and students was done using both the bottom-up and top-down approaches for the same reasons. In this explanatory research design, the case study in Phase 2 was conducted to explain the patterns found in Phase 1; therefore, many categories had already been informed not only by previous studies on the topic, but also by the findings from Phase 1.

3.7.4.2 Observations.

The observations were analysed quantitatively and qualitatively. First, the observation schemes were coded using a top-down approach. The coding focused on four aspects of teaching: participant organization, teaching and learning activities, teaching materials, TOEIC content vs. non-TOEIC content. A list of predetermined categories was used for coding each aspect of language teaching. For example, pair work, group work, whole class, and individual work were used for coding the participant organization.

Second, qualitative analysis of teachers' teaching was also performed. The analysis focused on the ways they taught the language skills and their use of English and Vietnamese. Verbatim transcriptions of teaching episodes were presented to illustrate the points analysed.

3.7.4.3 Weekly journals.

Students' journal sheets were analysed quantitatively. Students' responses to each question were coded, frequency counts were done, and the results were reported in numbers. Only the results from four out of the seven questions were reported in this thesis. The following is a description of the ways the responses to those questions were coded.

Question 1 asked students to tick the skills they had been working on in the past week, indicate the time they had spent on each skill, and give the reasons they had been working on those skills. Therefore, students' responses were coded into categories such as listening, speaking, and writing. These categories included six subcategories about the reasons why the students had been working on the skills. These were TOEIC, TOEIC and other, general English, special cases, other, and unclear. The subcategories resulted from the coding of the journal sheets as well as students' clarifications of those categories at the subsequent interviews with them. The total number of times students ticked the skills; the total amount of time they spent on them; and the reasons they worked on them during the four weeks of keeping weekly journals were reported.

Question 2 asked students to report what materials they used to learn English and the reasons for using them. The materials were coded into six categories: TOEIC, Non-TOEIC materials for TOEIC, non-TOEIC materials for other, non-TOEIC materials for TOEIC and other, non-TOEIC materials for general English, and unclear.

Question 3 asked students to report the activities they did to learn English, the time they spent on them, and the reasons for doing them. The responses were coded into direct test preparation, indirect test preparation, general English, other, and unclear. The number of time they did the activities was reported.

Question 4 asked students to report their achievement if any for each of the language skills. Their responses were coded into achievement, no achievement, and I can't say or no information. Under achievement there were subcategories labelled in accordance with the listed language skills.

3.8 Chapter Summary

This chapter gave an overview of the research methodology to achieve the purpose of the study. The chapter started with the statement of purpose and the research questions. Next,

it presented the overall research design together with the rationales for it. Then it described the research sites and participants in each phase of the study. After that, it discussed the validity and reliability in both quantitative and qualitative research as well as the specific strategies used in the current study to ensure these qualities. Following the discussion of research quality was a description of measures to ensure human research ethics. The final two sections described the methodology of the two phases of the study in detail. In the next three chapters, findings from Phase 1 are reported with some discussion included.

PHASE 1 FINDINGS

The three chapters in this part report the findings from Phase 1. Chapter 4 presents the findings from the interviews with policymakers. Chapter 5 reports the results of the teacher questionnaire analysis. Chapter 6 deals with the student questionnaire findings.

Chapter 4: FINDINGS FROM INTERVIEWS WITH POLICYMAKERS

This chapter describes the results of the policymakers' interview analysis. The interviews were used to seek their perspectives regarding the TOEIC-as-exit-test policy with a focus on the following questions:

1. What are policymakers' perceptions of the purposes of language teaching at university?
2. What are policymakers' rationales for choosing TOEIC as a university language exit test?
3. What is the intended impact of the TOEIC-as-exit-test?

The report of the findings is organized in accordance with these questions.

4.1 Demographic Information

The policymaker participants were made up of two groups. The first group included 12 Heads of Department from different universities. Eleven of them were Heads of the English or Foreign Language Department, and one of them was Head of the Training Department. Although not particularly in charge of language teaching, the Head of the Training Department was also selected because he was deeply involved in the TOEIC-as-exit-test policy at his university. Generally, the Heads of the English Department were mainly responsible for academic affairs. They made decisions on the English syllabus, teaching materials, and course assessment. However, when making decisions on syllabi and assessment, they were somehow restricted by MOET's framework programmes. Concerning the TOEIC-as-exit-test policy, the HODs had various roles. Some of them confirmed that they themselves initiated the policy, while others said that the policy was a top-down decision, and they were only consulted about its suitability. Other participants had a mixed role, saying the university leadership suggested using TOEIC, and they accepted it because they thought it was suitable.

The second group consisted of two MOET officials (MOs) in charge of higher education. One of them had retired a short time ago, but she was working when the TOEIC policy started. Talking about MOET's role in the TOEIC policy, both officials asserted that

MOET only required universities to publicise their learning outcomes without forcing them to use any language benchmarks as indications of these outcomes.

4.2 Purposes of Language Teaching at University

Table 4.1 Policymakers' Opinions Regarding the Purpose of Language Teaching at University (N=14)

Purposes	No of participants	No of comments
For communication in the workplace	11	13
For academic purposes	3	3
For students to have basic English to read and/or translate specialized materials later	2	2
For an international certificate	1	1
To train language skills	1	1
Total	17	20

Although five purposes of English teaching at university were mentioned, the overwhelming one was communication in the workplace (see Table 4.1). It was cited by 11 out of 14 participants, and 13 comments were made. Some participants simply stated that the purpose of English teaching was to make students to be able to communicate in their working environment. Some others further explained the reasons for this purpose. The most common reason was that English is required by companies for recruitment. Following is a typical comment:

[The purpose is that], with a minimum instructional time plus an additional English programme [...], after graduation, students will have minimum communicative ability because nowadays, apart from specialized knowledge, foreign language is an extremely important factor for getting employment at good companies. Having realized this matter, the university has aims at improving students' foreign language ability, that is, when students graduate and have job interviews, they must meet the minimum standard required by the companies (HOD 2).

This reason was also associated with the trend of integration and globalization that is happening in Vietnam and around the world.

So that students can get better jobs in this time of integration. People think that nowadays English is very necessary. English has already been an international means, a means of international communication, so if students know English, this is favourable to them in many ways (HOD 12).

The other purposes of English teaching were mentioned by very few participants. Teaching for academic purposes, which was the second most often specified one, received only three comments. According to participants, universities taught English so that their students could use English as a tool to learn their specialized subjects for their current studies or to meet the English requirements for further education abroad. This was one of their comments:

The purpose that the university as well as the Foreign Language Department have set for English teaching is to provide students with some knowledge and skills so that they can access their learning materials. You know, there are a lot of economic materials in English (HOD 1).

A lower expectation of the learning outcomes was students' having basic English. Two participants were pessimistic about expecting graduates to be able to communicate in English. Therefore, they only aimed at a basic English level so that students can read and/or translate their specialized materials for their work later.

From my personal point of view as well as the department's point of view, English teaching at university has to provide students with a certain amount of basic knowledge, and then... Actually, this is also our hope that the training process will help students form their very first language use skills because after graduation... to be honest, after so many years I've been working here, every year, graduates' ability to use English has been evaluated as very bad. Students at University H are, mainly, only pretty good at reading. They're very good at reading books to do research or to perform their jobs later, but their communicative ability is, generally, unsatisfactory (HOD 10).

Only one participant said that his university wanted their students to be able to get an international English certificate before graduation. Another argued that the purpose of English teaching is to train skills, not to teach knowledge about the language.

4.3 Rationales for Choosing TOEIC as a University Exit Test

Policymakers cited various reasons for choosing TOEIC as a university exit test (see Table 4.2). The most common reasons were its suitability, prestige, and good quality and the poor testing and assessment practice in Vietnam.

Table 4.2 Rationales for Choosing TOEIC as a University Exit Test (N=14)

Rationales	No of Participants	No of Comments
Suitability of TOEIC	13	23
Prestige of TOEIC	8	18
Good quality of TOEIC	7	14
Usefulness of TOEIC	5	14
Poor testing and assessment practice in Vietnam	11	18
TOEIC is an independent assessment	3	4
TOEIC is students' preference	3	4
Other	6	6

4.3.1 First rationale – TOEIC is suitable.

Policymakers seemed to feel most strongly about the suitability of TOEIC for graduation purposes because 13 out of the 14 participants addressed it with 23 comments. The suitability of the test was often assessed in relation to other major international tests (see Table 4.3). Half of the participants said that TOEIC was chosen to be the language exit test because it assessed language in the workplace. All of them saw this feature of the test in relation to TOEFL or IELTS, which they believed were academic in nature: “The University considers TOEIC suitable. TOEFL and IELTS are academic, but TOEIC assesses language in the workplace” (HOD 4). This seems to be logical because the overwhelming number of participants said that the purpose of English teaching was for employment.

The second feature that makes TOEIC suitable was that it was more appropriate for students' language proficiency than other international tests: “The university chose TOEIC to be the language exit requirement because, although TOEFL or IELTS came before TOEIC, they are too high [difficult]” (HOD 2). The next most frequently mentioned reason was that TOEIC had the most suitable test-taking fees. This does not necessarily mean that the majority of students can afford the test-taking fee, but it is much cheaper compared to TOEFL and IELTS.

Table 4.3 Reasons Why TOEIC is Suitable (N=14)

Reasons	No of Participants
TOEIC assesses language in the workplace	7
TOEIC is more suitable for students' proficiency	4
TOEIC has most suitable test-taking fees	4
TOEIC is suitable for my students' majors	2
TOEIC is suitable for limited instructional time allowed	1
TOEIC is not ideal, but most suitable for our current situation	2

It is interesting that one participant considered TOEIC suitable for graduation purposes because it suits the limited instructional time allowed by universities or MOET.

First, from the academic perspective, it has only two skills. Considering our instructional time, it is pretty suitable. For example, we'll try to focus on those two skills. It's easier to focus on those two skills than teaching for other certificates which require all four skills (HOD 1).

Two participants said that the test content was suitable for their students' majors, namely banking, economics, and IT.

Nevertheless, being aware of weaknesses of the test, two policymakers said that TOEIC is not ideal, but most suitable for their situation. One of them also added that it is unnecessary to set the goal of assessing students' communicative ability at the time because language teaching at university had not been communicative. From her personal observations, English classes were usually large; therefore, teachers only let students read a passage and then corrected their pronunciation, and there were absolutely no communicative activities.

4.3.2 Second rationale – TOEIC is prestigious.

Prestige of TOEIC is the third most common reason cited for using or supporting TOEIC with a total of 18 comments made by 8 participants. Participants "trust the prestige of international tests." Most comments emphasized the fact that TOEIC is an international test, and so it is prestigious and recognised:

(...) [our] point of view is to choose an independent testing centre. And it is even better if the centre belongs to English-speaking countries, and has a brand name, an

internationally prestigious brand name. TOEIC meets the criterion of being an independent centre. It has been built on international standards and it has a brand name, so the university has chosen it to be the language exit requirement (HOD 2).

Some others added that as an international certificate, TOEIC could give students an advantage when they applied for a job:

So the University has used an international certificate for a benchmark to give [students] a competitive advantage in the labour market. For example, when you apply for a job, if you show them [employers] an international certificate, then they... you are at an advantage, and it's easier for you to get the job (HOD 1).

One of the participants even trusted the test through the prestige of other universities. He chose the TOEIC because “famous universities such as the University of Technology, the Foreign Trade University, and the Economic University have used it” (HOD 6).

4.3.3 Third rationale – TOEIC has good quality.

Participants not only trusted the prestige of TOEIC, but many of them did believe that the test was of good quality. They commented on different aspects of the test validity and reliability. Nine out of the eleven participants who commented on the validity of TOEIC said that TOEIC scores accurately reflected learners' language ability, one of them said it did when the scores were high, and only one of them said it did not. Seven of them cited the test validity as a reason for choosing it to be their language exit test.

Participants gave different reasons for why they believed in the accuracy of the test results. Two of them mentioned the similarity of students' TOEIC test results with their ability observed in class or their course assessment results, which is called concurrent validity in language testing. Almost all of them attributed this accuracy to the strictness, objectiveness, and standardization of the test. They mentioned the strict organizational procedures and the absence of subjective elements, leniency, cheating, and the influence of any other factors on the test results except learners' ability. For example, when talking about the reasons for supporting TOEIC, MO 1 said:

All the tests that foreign countries produce, such as TOEFL, IELTS, or GMAT, are strict. They are strict in the way that people can't copy from somebody else; they can't do anything; no relationship, nothing can affect [the test results]. This strictness forces people to truly learn. That's what we want.

This suggests that when giving rationales for choosing TOEIC, the policymakers generally referred to test validity as the ability of the test to measure accurately what it actually

measured rather than the overall language ability. This was clearly evident in HOD 2's comment:

Although the test has only two skills, it accurately reflects learners' ability because its administration is very strict, very strict like other international tests such as TOEFL. There are never issues called leniency. They are totally absent. When it's strictly administered, it certainly reflects students' ability accurately.

Two participants commented on the reliability of the test. HOD 4 said, "Reality shows that some students took the test many times, but the results were always the same." HOD 5 talked about the consistency of the parallel forms of the test when the scores from different forms indicated the same ability.

4.3.4 Fourth rationale – Poor testing and assessment practice in Vietnam.

The judgement of the prestige and good quality of TOEIC was often made in relation to the perceptions of the poor language testing and assessment practice in Vietnam. In fact, seven out of eight participants who cited the prestige of TOEIC as a reason for their choice also mentioned the poor testing and assessment practice in Vietnam. Similarly, among the nine participants who commented on the good quality of TOEIC, eight also mentioned the poor practice in Vietnam. This poor practice was seen in two aspects. The first one was the distrust in the Vietnam National English Certificates. According to the participants, these certificates were not prestigious or recognised by the society anymore. The reasons were that they could be bought, a lot of people used forged certificates, there was cheating or leniency in marking or administering the tests, and the difficulty levels of the test varied from one administering institution to another. Here are some comments:

MOET required universities to build their learning outcomes, but what is available now? The old traditional A, B, C [certificates] are not recognized by anyone any more. No one recognizes them! Then the University thought of an international test, which is prestigious and accurately assesses [learners' ability]. With the old A, B, C, each place is different. They [students] can get a certificate somewhere very easily. They can even buy one easily (HOD 1).

Secondly, we don't recognise A, B, C certificates. Lots of forged ones. You know, other modes of learning such as on-the-job training (...) used to allow students to submit A, B, C certificates to be exempt from the English subjects. But too many forged certificates were found, so students are required to take the exit test. (...). They can spend only 500,000 dong and get a certificate (HOD 1).

All these problems led to the invalidity of the National Certificates. Participants thought the National Certificates “can’t accurately assess students’ language ability very well” (HOD 3). HOD 6 added more details to illustrate the invalidity:

...you know, in our country, there are people who hold even a level C certificate but can't speak English... Obviously, I've met people with level B and level C certificates who can't speak. They are not good, but they still pass the test. If they take the TOEIC test, how can they pass? At least the test [TOEIC] is difficult. The structure of the test is already difficult (HOD 6).

Participant MO 2 commented on the meaning of test results: “Due to our implementation, the borders between levels A, B, and C have become blurred. You are saying that someone has achieved level C, but your test might be easier than other institutions’ tests” (MO 2).

The second type of problem that Vietnam is facing is associated with in-house test development. According to participants, it was impossible to develop a reliable and valid test due to poor measurement and assessment knowledge:

Honestly, during the 2004-2005 period and later as well, we were confused about language measurement and assessment and, more broadly speaking, our assessment in general. You studied education, so you know that our measurement and assessment are bad; [teachers] are almost not trained in assessment. I mean our test questions are funny. They mainly measure rote learning or application. They are completely subjective. We didn't have an assessment tool, and people had different opinions about the A, B, C, D [certificates], so we chose something reasonable. At that time, many people supported TOEIC, of course, specialists supported it, so the Ministry supported the policy (MO 1).

Later participant MO 1 added,

At present, there may be different benchmarks, and they are recognised. We think that Vietnam is behind, so we should use an available measurement tool. It's troublesome to develop our own tool. Secondly, as I said, later if we want to build an assessment system based on the European Common Framework, to be honest, we still have to rely on foreign countries; we can't think of a [system] ourselves. Our teachers can't make tests themselves.

Other participants also mentioned other difficulties in developing in-house tests such as the time and cost needed or the lack of a department specializing in test development: “If Vietnamese teachers develop a test themselves, then it's necessary to go through a very long procedure to validate the questions, which would cost a lot of money. Therefore, in our condition, the university chose whichever was fast and effective” (HOD 9).

The last issue that prevents universities from using their in-house tests was teachers’ leniency. Teachers may write easy questions so that a certain percentage of students can

pass, or they can pass their students out of sympathy when students beg them for it. This is what HOD 8 said:

International certificates evaluate students' real ability better than our own tests because there are negative issues related to our University's assessment. For example, some students ask their teachers for more marks. They sometimes say, "We've failed only one subject, so please pass us!" and teachers may increase their scores a little bit, so that they can pass (HOD 8).

4.3.5 Fifth rationale – TOEIC is useful.

A few participants pointed out the usefulness of TOEIC as a rationale for using it as an exit test. The overwhelming number of comments (9 out of 14) claimed that TOEIC scores were used by employers for recruitment: "...a lot of businesses say outright what TOEIC scores are required when they recruit new employees" (MO 1). Two participants said that TOEIC can be useful for further education. They explained that TOEIC could help students improve their ability, and so they could take the opportunities to study abroad, or TOEIC scores could be converted into TOEFL or IELTS, so students could use them to apply for postgraduate study programmes.

But, of course, when we were working with TOEIC [provider], we knew that TOEIC scores can be converted into TOEFL or IELTS scores, so that when you go on your master's or PhD degree, you can use them. In fact, many countries where English is not the mother tongue do use TOEIC for admission to master's programmes. Japan is an example. When I was working with TOEIC in Japan, in the office in Japan, they introduced to me a lot of universities that require certain TOEIC scores for admission to their master's programme, for instance (MO 2).

Only one participant said that TOEIC could give students feedback on their learning and so helped them set their own learning goals. In her opinion, the continuous TOEIC score scale let students themselves know clearly how much progress they had made each term, and so they had a goal to strive for. Another participant argued that because TOEIC was serious and strict, it made students truly learn.

4.3.6 Sixth rationale – TOEIC is an independent assessment.

Three participants mentioned that they wanted an independent organization that was not involved in the teaching process to assess their students. HOD 3 said, "Actually, the university is concerned about students' English learning quality, so an independent testing organization is needed to assess them accurately." Later he explained that they wanted an assessment by an independent organization because the assessment could be objective: "Our university wants to rely on an objective organization to assess [our students]. We

don't want to teach and then administer the A, B, C tests ourselves.” One of the participants gave another reason for an independent testing organization: Her university was outsourcing people from another institution to teach English to their students, so they need a third party to make the assessment.

4.3.7 Seventh rationale – TOEIC is students' preference.

Three participants said one of the reasons they chose TOEIC was that their students liked it the most. When they were given a choice among different international tests, most of them submitted the TOEIC certificates. At one university, when the new programme started, students were given a choice between TOEFL and TOEIC certificates to study for; they chose the TOEIC test because their purpose was to get a job after graduation, not to go on to further education abroad.

4.3.8 Other.

There were six other rationales for the use of TOEIC, each of which was mentioned only once. Those were the popularity of TOEIC compared to other international tests, the university's board of directors' determination to make a prominent achievement, convenience of learning for and taking the TOEIC test; successful promotional campaign by TOEIC Vietnam, and so on. One participant, acknowledging the weakness of TOEIC, which assessed only the two receptive skills, still supported the use of the test. According to him, teaching and testing were different: Although the test evaluated students on only two skills, universities still taught their students all four skills in their language programmes, and their purpose was not to teach students for TOEIC or TOEFL.

4.4 Intended Impact of the TOEIC-as-exit-test Policy

Table 4.4 Intended Impact of TOEIC

Intended impact	No of participants	No of comments
To impact on learning	14	31
To impact on teaching	8	10
To impact on language programmes	8	11
Other	3	3

4.4.1 Intended impact on learning.

Almost all of the participants said they wanted the test to impact on students' learning (see Table 4.4). The most common type of intended impact was impact on students' motivation, which received 31 comments from 14 participants. The participants hoped that the test would make students aware of the importance of the English subject, change their attitude towards learning it, and make them exert themselves to the utmost. HOD 5 said:

When we change the assessment, learners' attitudes towards the subject will always change. The fact that students are required to take an English test as soon as they enter the university, then take four English modules, and meet the English exit requirement when they graduate will make them feel that this is important, and this will become a motive for them to change (HOD 5).

According to participants, the TOEIC test could motivate students in two ways. On the one hand, it served as a barrier which obliged students to make an effort in order to be able to jump over it.

The University's policy is to use TOEIC as a prerequisite for students to graduate. This is a way to standardise them. This means that, with the language exit requirement, students have to be responsible for their own learning. If we only let students learn normally, they won't hurry. But now that we have set a certain score, students will study (HOD 6).

On the other hand, TOEIC was used as a scaffolding strategy to encourage students to go further.

But I think this [a TOEIC score of 450] will be an advance, which will help students get used to [learning English]. If they have achieved that [the TOEIC benchmark], they will feel that English is not difficult, and that they can further improve it themselves (HOD 6).

However, more participants mentioned the first way than the latter. The second type of intended impact on students' learning mentioned by the majority of participants (10) was to make them meet the requirements of the labour market.

Four other comments indicated the specific directions of impact that participants wanted to have on students' learning. Two participants said that the test would push students towards communication: "The university hopes that students [...] will be able to communicate in English, unlike what people usually think about learning languages at universities with students being like dumb and deaf people" (HOD 8).

One participant said the test would have impact on the content of learning, for instance, students would listen more. Another participant said he wanted the test to make students truly learn.

4.4.2 Intended impact on teaching.

There were fewer comments on teaching than on learning. Six of the ten comments on teaching were concerned with teaching methods. Two of them were rather vague. Participants only said that teachers should look for the best teaching methods or improve their teaching methods, or teaching should be more specific.

Two participants said teachers should change their traditional teaching methods into the communicative approach. HOD 8 said: "Teachers have to improve their pedagogical skills, their teaching methods towards communication. Language used to be taught academically, i.e., teaching translation of texts, a deep knowledge of grammar, reading and writing" (HOD 8).

Having the same opinion, HOD 1 pointed out the influence of the new textbook on teachers' teaching methods:

Teachers... obviously, with the previous ESP they still used, I don't know all, but possibly a lot of teachers used the old teaching methods [...], which was reading-translation, i.e., reading, translation, and vocabulary explanation. Now the method is different. They have to... and the textbook, they have to follow the textbook. Market Leader is very much communicative (HOD 1).

One participant said that as teachers now had to teach test-taking skills to students, they would have to change their teaching methods and teach in such a way that would attract their students.

Two participants hoped the new policy would push teachers to improve themselves, i.e., they would have to be more dynamic or to try hard to meet the university's standards of teacher quality. An example was that teachers would have to obtain high TOEIC scores themselves.

... teachers have to be more dynamic. This university paid for teachers to attend IIG Vietnam's workshop. When doing the workshop, teachers also had to take a TOEIC mock test and then learned [TOEIC] for a while. After the workshop, teachers had to improve themselves because the learning time was short, only a few days. So teachers have to try harder, and actually, whenever teachers teach high levels, then they'll have to improve and perfect themselves (HOD 11).

One participant made a vague statement that teaching should be partly oriented towards the assessment.

4.4.3 Intended impact on language programmes.

There were 11 comments about the intended impact of the test on language programmes, i.e., on syllabi, teaching materials, and course assessment. In fact, most of the universities where the HODs interviewed were working had made changes to their English syllabus after the TOEIC test was introduced. Some of them turned their whole programme into TOEIC preparation with a small speaking component. Others removed the ESP modules and either replaced them with General English/Business English with a small proportion of TOEIC preparation in each module or replaced them with one General English module and one or a few TOEIC preparation modules. Together with the changes in the structure of the syllabus, many universities also changed the instruction time allocated to their English modules. However, when asked about the intended impact of the TOEIC test, only four participants mentioned that they wanted to see change in the syllabus. They talked about the changes in the structure of the syllabus or instructional time allocation to make the programmes more suitable to the language exit requirements: "We don't teach ESP anymore. Instead, we teach Business English to make it [the syllabus] more in line with our language exit requirements. That is [the impact] on the programme" (HOD 1). Another one said:

When the language exit requirement was A, B certificates, there were only 3 terms [English modules], but when the university decided to use the TOEIC benchmark, it increased the number of terms to five. Each term is 60 periods. The purpose was to give students more opportunities to learn English, to have good English ability, so that they can meet their job requirements later (HOD 3).

One participant, however, just gave a vague answer: “Yes, of course. Because when you introduce such a programme, all the functional departments have to get involved: building the syllabus, preparing the human resources, and then preparing the material base” (HOD 2).

Similarly, although most universities had changed their textbooks after the introduction of TOEIC, only two participants said that their intention was to impact on teaching materials and their answers were also vague: “Of course, there should be innovations, conferences to put forward the best teaching methods and select the best textbooks. There should be changes!” (HOD 2) or “Students have materials to practice” (HOD 9).

It seems that, although participants mentioned the changes as their intended impacts or even sometimes stressed that they intended to do this when they planned to use the TOEIC test, these should be better considered as measures universities took to implement the TOEIC policy rather than the end-purpose of the policy.

The largest number of comments (5 out of 11) was about the impact of the exit test on assessment. First, the introduction of the TOEIC was intended to cause change in course assessment. According to a participant, this impact was mediated by the changes in the syllabus:

With the previous ESP, we only focused on reading comprehension. There were no listening and speaking exams at all. Now testing also needs to be changed. There should be exams for all four skills. That is the impact on the programme (HOD 1).

At this university, after the TOEIC was introduced, ESP was replaced with Business English, and ESP reading materials were replaced with Market Leader, which was built on the communicative approach with attention paid on all four skills.

Second, the introduction of the TOEIC also aimed at influencing the difficulty levels of the tests that teachers write:

If we didn’t use the TOEIC benchmark, if we used our normal tests, teachers would design tests based on students’ ability level. We would design such tests that a certain percentage of students could pass. But when we have a benchmark, this is clearly pressure for both students and teachers. Now we have to write tests according to the benchmark, although we can’t reach the level which is internationally recognised, but we have to emulate it, relatively, but can’t write a too much easier test (HOD 11).

One of the participants did not specify any intended impact on teaching, learning, and the language programme. She said, “Actually, this is something new, so generally everyone is

expecting change or results which are different from before” (HOD 10). Her statement showed that her university had no clear objectives when they adopted the TOEIC policy.

4.5 Chapter Summary and Discussion

This chapter reported findings from the interviews with policymakers. The report focused on three big issues related to policymakers’ perspectives on the TOEIC-as-exit-test policy: The purpose of language teaching at university, the rationales for choosing the TOEIC test to be the university language exit test, and the intended impact of the TOEIC-as-exit-test policy.

The overwhelming number of policymakers considered the purpose of language teaching as improving students’ ability to communicate in the workplace. The main reason for this purpose was employers’ demands. Another reason was the current trend of integration and globalization happening around the world. A small number of participants mentioned some other purposes related to current and future learning activities and the ability to read and translate specialist materials at work.

Participants gave a lot of different rationales for choosing TOEIC to be their language exit test. The most common ones were the suitability of TOEIC, its prestige, its good quality, its usefulness, and the poor testing and assessment practice in Vietnam. Most of these rationales essentially involved the employment factor more than other factors: TOEIC assessed language in the workplace; it gave students job advantages.

The policymakers also intended to use the TOEIC-as-exit-test policy to produce certain impact on teaching, learning, and language programme. The most important intended effects of the test on learning were to motivate students to learn and to make them meet the requirement of the market by establishing a language barrier for students to overcome. A small number of participants wanted to push students towards communication. Affecting teachers did not appear to be a main concern of the TOEIC-as-exit-test policy. Not everyone mentioned intended impact on teaching. Some of the comments were very vague without clearly stating the directions of change and the mechanism in which the policy would cause the change. The intentions also varied across participants. In other words, there did not seem to be a clear philosophy under the intended impact on teaching. The other intentions mentioned were to change the language programme (syllabus, textbooks, and course assessment). Initially motivated by the MOET’s requirement for publicizing

learning outcomes, the TOEIC policy was perhaps a situational response to the requirement rather than a well-thought out plan for curriculum innovation. As shown in the analysis, one participant said that his university adopted the policy because famous universities had done it; another participant did not specify any clear intended impact.

From the interviews with the policymakers, communicative ability for employment clearly emerged as the most important theme. The purpose of language teaching was for students to have the communicative ability to work. Many of the rationales for the TOEIC policy were job-related. Major intended impact of the policy was also for students to meet the labour market demands. This probably reflected the role of economic factors in language education policy as remarked by Ross (2008): “Language education policy in Asian nations is particularly driven by economic goals” (p. 789).

Chapter 5: TEACHER QUESTIONNAIRE FINDINGS

This chapter describes the results of the teacher questionnaire analysis. The teacher questionnaire was used to seek the answers to the following questions:

1. What are teachers' perceptions of the impact of the TOEIC test on university English teaching?
2. What are teachers' perceptions of the factors potentially involved in forming the impact of the TOEIC test on teaching?
3. To what extent are test factors, context factors, and person factors involved in shaping the test impact?
4. Are there differences between large city universities and provincial universities in teachers' perceptions of the impact of the TOEIC test on teaching?
5. Are there differences between large city universities and provincial universities in teachers' perceptions of factors potentially involved in forming the impact of the TOEIC test on teaching?

The report of the findings is organized around these research questions. Participants' responses to the open-ended question in the questionnaire are not presented because although they were very interesting, they mostly repeated the information already obtained from the closed-ended questions. The interpretation of the mean scores in this and the subsequent chapters is based on the response values of the Likert scales used in the questionnaires with 1 being *strongly disagree*, 2 – *disagree*, 3 – *somewhat disagree*, 4 – *neither agree nor disagree*, 5 – *somewhat agree*, 6 – *agree*, 7 – *strongly agree*. (For information about the items making up the composite variables, refer to Table 3.12.)

5.1 Demographic Information

Responses from 81 teachers were used for the data analysis. Table 5.1 and Table 5.2 show their demographic information. The participants had an average of approximately 37 years of age ($SD=10.45$); 11 years of experience teaching English ($SD=7.84$), but just over 3 years of experience teaching TOEIC ($SD=1.93$). The majority of them held an MA degree (75%); most of the remainder had a BA degree (23.8%); only one participant had a doctoral qualification. The number of females was double the number of males.

The teachers from University A had much higher age and teaching experience means than those from Universities B and C. Almost all of them had an MA degree.

Table 5.1 Demographic Information of Teacher Participants – Part 1 (N=81)

Variables		University A	University B	University C	Total
No of participants		23 (28.4%)	32 (39.5%)	26 (32.1%)	81 (100%)
Gender	Male	3	9	14	26 (32.10%)
	Female	20	23	12	55 (67.90%)
Qualifications	BA	1	7	11	19 (23.46%)
	MA	21	25	14	60 (70.07%)
	PhD	1			1 (1.23%)
	Missing			1	1 (1.23%)

Table 5.2 Demographic Information of Teacher Participants – Part 2 (N=81)

Variables		University A	University B	University C	All three universities
Age (in years)	Mean	42.17	33.84	35.12	36.62
	SD	8.48	9.40	11.42	10.45
Experience teaching English (in years)	Mean	15.63	9.43	9.83	11.34
	SD	6.54	6.80	8.79	7.84
Experience teaching TOEIC (in years)	Mean	3.20	3.90	3.62	3.62
	SD	2.63	1.59	1.63	1.93

5.2 Perceived Impact of TOEIC on Teaching

Seven composite variables and two single item variables with regard to test impact on teaching were analysed (see Table 5.3). The “positive feelings and attitudes” variable had a mean of 4.04 with a standard deviation of 1.23. This means that most teachers could not decide whether they felt optimistic about their students’ test results and had positive attitudes towards preparing students for the TOEIC test or not. On the contrary, they

slightly agreed that they felt anxious about improving their students' test scores with the mean for "anxiety" being 5.15 ($SD=1.25$).

Table 5.3 Descriptive Statistics on Perceived Impact of TOEIC on Teacher Teaching (N=81)

Test impact	N	Mean	SD
Positive feelings	79	4.04	1.13
Anxiety	80	5.15	1.25
Direct test preparation	78	5.72	1.04
Extent of impact	81	5.12	1.36
Time devoted to tested skills	80	5.85	0.89
Time devoted to non-tested skills	80	3.31	1.21
Time devoted to communicative activities	80	3.24	1.41
Time devoted to traditional activities	78	4.09	1.18
Teacher's use of Vietnamese	81	4.14	1.40

The majority of teachers agreed that they had been doing some direct test preparation activities since the introduction of the TOEIC test, that is, they had been teaching their students test-taking strategies, using mock tests or test preparation materials. The "direct test preparation" variable had a pretty high mean score of 5.72 with a standard deviation of 1.04. The time teachers used Vietnamese seemed to remain almost the same with the mean being only 4.14, but teachers differed in this respect with the standard deviation being 1.40.

Teachers indicated that their instructional time devoted to the tested skills (listening, reading, grammar, and vocabulary) had increased moderately ($M=5.85$ and $SD=0.89$). On the contrary, time devoted to non-tested skills (speaking, writing, and pronunciation) had decreased slightly ($M=3.31$), but teachers were less unanimous regarding this variable than the previous one with the standard deviation being 1.21 compared to only 0.89.

"Time devoted to traditional activities", including grammatical analyses of texts, mechanical drills, and use of metalanguage, had remained almost the same with the mean for the variable being 4.09 ($SD=1.18$). However, "time devoted to communicative

activities” slightly decreased ($M=3.24$). It is necessary to note that teachers’ reports on this variable varied with the standard deviation being 1.41.

Although the statistics revealed some perceived influence of the test on teaching, this influence was probably not too strong because teachers only slightly agreed with the questionnaire items on “extent of test impact” ($M=5.12$, $SD=1.36$).

In brief, the teachers’ responses to the questionnaire showed that the introduction of the TOEIC test clearly created distinct patterns of impact on teaching. Due to the introduction of the test, teachers carried out direct test preparation activities and spent more time on the tested skills, while they reduced time devoted to the non-tested skills and communicative activities (only time spent on traditional activities remained the same). They also slightly felt anxious about improving students’ test scores. However, they did not report a large extent of impact.

5.3 Factors Potentially Involved in Forming the Impact of TOEIC on Teaching

Research literature suggests that a variety of factors may contribute to the test impact on teaching and learning. These factors can include stake-holders’ perceptions of the test, their personal beliefs, and the context factors. This section of the thesis reports on teachers’ perceptions of these factors.

5.3.1 Teachers’ beliefs about language teaching and learning.

The results showed that nowadays most teachers are well aware of, and believed in, the communicative approach. The “belief in communicative approach” variable had a high mean of 6.34 and a low standard deviation of 0.59. On the contrary, the belief in traditional methods, which considered the most important part of language teaching to be teaching vocabulary words, grammar rules, and translation, received a lower mean score of 3.50. However, teachers’ opinions with regards to this respect seemed to vary with the standard deviation being 1.17.

Regarding the ways of improving test scores, teachers believed in both direct test preparation activities and the overall improvement in language learning, but the latter had a higher mean of 5.81 compared to 5.48 ($SD=1.07$ and 1.10, respectively).

Table 5.4 Descriptive Statistics on Teachers' Beliefs about Language Teaching and Learning (N=81)

Variables	N	Mean	SD
Belief in communicative approach	81	6.34	0.59
Belief in traditional methods	80	3.50	1.17
Belief in the necessity of direct test preparation	80	5.48	1.10
Belief in overall improvement	81	5.81	1.07

5.3.2 Teachers' perceptions of TOEIC.

Table 5.5 Descriptive Statistics on Teachers' Perceptions of TOEIC (N=81)

Variables	N	Mean	SD
Test value	78	4.62	1.22
Test purpose-tested skills	77	5.51	1.03
Test purpose-non-tested skills	77	2.84	1.37
Test difficulty	79	5.21	1.04
Test format	80	6.20	1.08
Consequences for teachers	79	3.20	1.20
External expectations	81	5.00	1.07

Factor analysis of the teacher questionnaire revealed the same variables of test perceptions as that of the student questionnaire. Teachers less than slightly agreed that the TOEIC test was valuable ($M=4.62$), that is, it measured the most important knowledge and skills that university students should have; it accurately reflected students' English proficiency; it provided useful information as feedback to their teaching and their students' learning; it helped students set goals for their learning; it was useful for their students' careers; and it was suitable for evaluating English learning outcomes of university students.

Regarding the purpose of the test, teachers generally agreed that the test assessed the actually tested skills (grammatical knowledge, reading skills, and listening skills) with the mean score being 5.51. They somewhat disagreed that the test measured the actually non-

tested skills ($M=2.84$). It is noticeable that the former opinion was more unanimous than the latter ($SD=1.03$ vs. $SD=1.37$).

Most of the teachers somewhat agreed that the TOEIC test was difficult for their students ($M=5.21$, $SD=1.04$). They also agreed that it had a multiple-choice format ($M=6.20$, $SD=1.08$).

Test pressure on teachers was measured by two variables: consequences for teachers and external expectations. While teachers mildly denied consequences of the test for them ($M=3.1$, $SD=1.20$), they slightly felt the external expectations ($M=5.00$, $SD=1.07$). Thus, teachers got a little pressure from the other stake-holders' expectation for them to prepare their students for the test rather than from the real consequences to them.

5.3.3 Teachers' perceptions of context factors.

The teachers' responses showed that teachers both faced constraints and had opportunities. Both of these variables had means of over 5.00 (see Table 5.6), but the responses varied more with regard to constraints ($SD=1.20$). The constraints included class sizes and time limitation. The opportunities variable was made up of freedom to use textbooks selectively and access to different kinds of resources.

Table 5.6 Descriptive Statistics on Teachers' Perceptions of Context Factors (N=81)

Context factors	N	Mean	SD
Constraints	81	5.13	1.20
Opportunities	80	5.40	0.83

5.4 Involvement of Test Factors, Person Factors, and Context Factors in Forming the Impact of TOEIC on Teaching

The involvement of test factors, person factors, and context factors in shaping the impact of the TOEIC test on teaching was measured by Pearson correlation coefficients between teachers' perceptions of these factors and their perceptions of the test impact. The strength of the correlations reported in this thesis was interpreted according to the guidelines given by J. Cohen in his book *Statistical Power Analysis for the Behavioral Sciences* (1969).

Small $r=.10$ to $.29$

Medium	$r=.30$ to $.49$
Large	$r=.50$ to 1.0

5.4.1 Person factors and perceived impact of TOEIC on teaching.

“Teachers’ age” had both positive and negative significant correlations with the perceived test impacts, but most of them were negative (see Table 5.7). This variable had a significant positive relationship with time devoted to non-tested skills and time devoted to communicative activities ($r=.27$ and $.35$, respectively).

On the other hand, it had significant negative correlations with “anxiety,” “direct test preparation,” “extent of test impact,” “time devoted to tested skills,” and “teacher’s use of Vietnamese.” Three of these associations were small with the coefficients ranging from $-.224$ to $-.261$. However, the relationship between age and “time devoted to tested skills” and “teacher’s use of Vietnamese” reached a medium strength ($r=-.49$ and $-.33$, respectively). The fact that all of these correlations were negative suggested that the younger the teachers, the more they taught to the test.

“Experience teaching English” had significant positive correlations with time devoted to non-tested skills and time devoted to communicative activities. The former had a small strength ($r=.22$), and the latter – medium strength ($r=.36$). On the other hand, this variable had negative relationships with anxiety, direct test preparation, extent of test impact, using past exam papers, time devoted to tested skills, teacher’s use of Vietnamese. Among these correlations, the correlation with time devoted to tested skills was large ($r=-.56$); the correlations with direct test preparation and teacher’s use of Vietnamese were medium ($r=-.3$ and $-.36$, respectively); all the other ones were small (r ranging from $-.23$ to $-.27$).

Thus, experience teaching English had significant relationships with almost the same test impacts as the teachers’ age did. Moreover, the significant correlations between experience teaching English with the test impacts also had the same directions as those between teachers’ age and test impacts. This could be explained by the fact that age is normally associated with experience.

“Experience teaching TOEIC” only significantly correlated with direct test preparation. The relationship was positive and small in strength with the coefficient being $.23$.

The belief in the communicative approach had no statistically significant correlations with any types of test impact. Similarly, the belief in the traditional methods only positively correlated with time devoted to traditional activities ($r=.26$). This is probably because teachers were unanimous with regard to beliefs about teaching and learning methods, especially the communicative approach.

The belief in direct test preparation positively significantly correlated with only direct test preparation, time devoted to tested skills, and teacher's use of Vietnamese. The strength of the relationship was medium for direct test preparation ($r=.35$) and small for the others. It is strange, however, that the belief in overall improvement in language learning was also associated with direct test preparation ($r=.26$), and this is the only statistically significant correlation found.

In brief, among the person factors, teachers' age and experience teaching English were associated the most with test impact, but these associations were negative. The older and the more experienced the teachers, the more they were distant from teaching to the test. It is surprising that teachers' beliefs about teaching methods hardly had a role in shaping the test impact. Naturally, the belief in the necessity of direct test preparation for score improvement and experience teaching TOEIC tended to correlate with direct test preparation activities and the tendency of teaching to the test.

Table 5.7 Statistics on Correlations between Person Factors and Perceived Impacts of TOEIC on Teaching (N=81)

		Age	Experience teaching English	Experience teaching TOEIC	Belief in communicative approach	Belief in traditional methods	Belief in direct test preparation	Belief in overall improvement for test score improvement
Positive feelings	Pearson Correlation	-.040	-.073	.182	.126	-.061	-.013	.161
	N	79	78	77	79	78	78	79
Anxiety	Pearson Correlation	-.224*	-.225*	.073	.122	.073	.128	.060
	N	80	79	78	80	79	79	80
Direct test preparation	Pearson Correlation	-.261*	-.305**	.226*	.100	.093	.349**	.263*
	N	78	77	76	78	77	77	78
Extent of test impact	Pearson Correlation	-.243*	-.273*	.147	.031	-.012	.052	-.045
	N	81	80	78	81	80	80	81
Time devoted to tested skills	Pearson Correlation	-.485**	-.558**	.046	.014	.128	.217	.010
	N	80	79	77	80	79	79	80
Time devoted to non-tested skills	Pearson Correlation	.269*	.219	.060	.058	.102	-.117	.119
	N	80	79	77	80	79	79	80
Time devoted to communicative activities	Pearson Correlation	.352**	.364**	-.021	.024	.200	-.103	.081
	N	80	79	77	80	80	79	80
Time devoted to traditional activities	Pearson Correlation	-.145	-.201	-.018	.134	.258*	.064	.079
	N	78	77	75	78	78	77	78
Teacher's use of Vietnamese	Pearson Correlation	-.331**	-.355**	.014	-.084	-.061	.282	.033
	N	81	80	78	81	80	80	81

*. Correlation is significant at the 0.05 level (2-tailed).

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

5.4.2 Perceptions of TOEIC and its perceived impact on teaching.

Test value significantly correlated with nearly half of the types of test impact (see Table 5.8). All the relationships were positive with a medium strength. They included correlations with positive feelings ($r=.43$), anxiety ($r=.32$), direct test preparation ($r=.44$), and time devoted to tested skills ($r=.30$).

The perception that the TOEIC test assesses the actually tested skills significantly correlated with five types of test impact, all of them being positive and three of them having a medium strength. Small correlations were found with anxiety ($r=.26$) and time devoted to traditional activities ($r=.27$). Medium correlations were found with direct test preparation ($r=.46$), time devoted to tested skills ($r=.40$), and teacher's use of Vietnamese ($r=.32$).

The perception that the TOEIC test assesses the actually non-tested skills only significantly negatively correlated with teachers' use of Vietnamese ($r=-.39$). It had significant positive correlations with three other types of test impact, one of them having a small coefficient and the other two having large coefficients. The small correlation was with positive feelings ($r=.28$). The large correlations were with time devoted to non-tested skills ($r=.59$) and time devoted to communicative activities ($r=.52$).

Test difficulty was significantly correlated with two kinds of test impact, that is, positive feelings ($r=-.27$) and time devoted to tested skills ($r=.32$). The more difficult teachers believed the test was for their students, the less positive they felt about students' future test results and the more time they devoted to the tested skills.

Table 5.8 Statistics on Correlations between Perceptions of TOEIC and Its Perceived Impacts on Teaching (N=81)

		Test value	Test purpose-tested skills	Test purpose-non-tested skills	Test difficulty	Test format	Consequences for teachers	External expectations
Positive feelings	Pearson Correlation	.426**	.047	.280*	-.274*	.036	.248*	.168
	N	76	75	75	77	78	77	79
Anxiety	Pearson Correlation	.324**	.257*	-.084	.184	.259*	.153	.280*
	N	77	76	76	78	79	78	80
Direct test preparation	Pearson Correlation	.435**	.461**	-.067	.077	.255*	.291*	.584**
	N	75	74	74	76	77	76	78
Extent of impact	Pearson Correlation	.039	.154	-.196	.017	.145	.077	.279*
	N	78	77	77	79	80	79	81
Time devoted to tested skills	Pearson Correlation	.299**	.400**	.019	.322**	.144	.077	.361**
	N	77	76	76	78	79	78	80
Time devoted to non-tested skills	Pearson Correlation	.214	-.131	.590**	-.091	-.170	.195	-.049
	N	77	76	76	78	79	78	80
Time devoted to communicative activities	Pearson Correlation	.178	-.083	.519**	-.173	-.076	.133	-.166
	N	77	76	76	78	79	78	80
Time devoted to traditional activities	Pearson Correlation	.178	.272*	-.021	-.106	.095	.160	.375**
	N	75	74	74	76	77	76	78
Teachers' use of Vietnamese	Pearson Correlation	.019	.316**	-.393**	.144	.056	-.041	.335**
	N	78	77	77	79	80	79	81

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

* . Correlation is significant at the 0.05 level (2-tailed).

Test pressure consisted of two variables – consequences for teachers and external expectations. The consequences for teachers variable only significantly positively correlated with positive feelings and direct test preparation (see Table 5.8). Both of them had a small strength ($r=.25$ and $.29$, respectively).

The external expectations were more associated with test impact than the test consequences were in terms of the number of significant correlations and their strength. All of the significant correlations found were positive with the strength being in the small to large range. The small correlations were with anxiety ($r=.28$) and extent of test impact ($r=.28$). The medium correlations were found with time devoted to tested skills ($r=.36$), time devoted to traditional activities ($r=.38$), and teacher's use of Vietnamese ($r=.34$). The correlation with direct test preparation was the only one that was large ($r=.58$). It was likely that external expectations had more influence on the shaping of the test impact than the test consequences to teachers did. This could be explained by the fact that the consequences to teachers were very small, while the external expectations were much higher.

In brief, among the test factors, test value, test purposes, and external expectations were associated the most with the test impact. They not only correlated with the most types of test impact, but also had the largest correlations. It is also interesting that teachers' reported behaviours tended to be consistent with their perceptions of the test design. The more they believed that the test assessed the actually tested skills, the more they were involved in teaching to the test. In contrast, the more they believed that the test assessed the actually non-tested skills, the greater their time increase for the non-tested skills and communicative activities.

5.4.3 Perceptions of context factors and perceived impact of TOEIC on teaching.

The constraints variable significantly negatively correlated with positive feelings with a medium strength ($r=-.33$). It significantly positively correlated with teacher's use of Vietnamese with a small strength ($r=.26$). The opportunities variable had more significant correlations with test impacts than the constraints variable did. The correlations between opportunities and test impacts were usually larger in strength than those between constraints and test impacts. All of the statistically significant correlations found were

positive with their strengths ranging from small to medium. Small correlations were discovered with positive feelings ($r=.23$), extent of impact ($r=.22$), and time devoted to tested skills ($r=.23$). Medium correlations were found with anxiety ($r=.37$), direct test preparation ($r=.35$), and time devoted to traditional activities ($r=.34$). The correlations suggested that the more opportunities teachers had, the more they taught to the test. For the statistics on these relationships, see Table 5.9.

Table 5.9 Statistics on Correlations between Perceptions of Context Factors and Perceived Impacts of TOEIC on Teaching (N=81)

		Constraints	Opportunities
Positive feelings	Pearson Correlation	-.328**	.226*
	N	79	78
Anxiety	Pearson Correlation	.076	.369**
	N	80	79
Direct test preparation	Pearson Correlation	-.010	.354**
	N	78	77
Extent of test impact	Pearson Correlation	.035	.224*
	N	81	80
Time devoted to tested skills	Pearson Correlation	-.037	.232*
	N	80	79
Time devoted to non-tested skills	Pearson Correlation	-.124	.053
	N	80	79
Time devoted to communicative activities	Pearson Correlation	-.129	-.049
	N	80	80
Time devoted to traditional activities	Pearson Correlation	-.011	.335**
	N	78	78
Teacher's use of Vietnamese	Pearson Correlation	.258*	.137
	N	81	80

*. Correlation is significant at the 0.05 level (2-tailed).

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

5.5 Differences between the Large City Universities and the Provincial University in the Perceived Impact of TOEIC on Teaching

Independent sample t-test analysis revealed that the large city universities and the provincial university were very similar in the perceived impact of the TOEIC test on teaching (see Table 5.10). There was only significant difference in direct test preparation activities. The large city universities had a mean score of only 5.49 with a standard deviation of 1.11, while the provincial university had a mean score of 6.11 with a standard

deviation of .67; $t = -3.03$, $p = .003$, two-tailed. The magnitude of the differences in the means was large with Cohen's d being -0.79.

Table 5.10 Statistics on Differences in Perceived Impact of TOEIC on Teaching between Large City Universities and Provincial University

Types of test impact	Location	N	Mean scores	SD	t	p (2-tailed)	Cohen's d
Positive feelings	LC	53	4.06	1.03	0.15	.884	0.04
	P	26	4.01	1.34			
Anxiety	LC	54	5.11	1.30	-0.40	.692	-0.10
	P	26	5.23	1.18			
Direct test preparation	LC	53	5.49	1.11	-3.03	.003	-0.79
	P	25	6.22	0.67			
Extent of impact	LC	55	5.02	1.42	-0.95	.344	-0.23
	P	26	5.33	1.22			
Time devoted to tested skills	LC	54	5.78	0.94	-1.02	.309	-0.25
	P	26	6.00	0.78			
Time devoted to non-tested skills	LC	54	3.27	1.19	-0.50	.620	-0.12
	P	26	3.41	1.28			
Time devoted to communicative activities	LC	55	3.19	1.46	-0.49	.627	-0.12
	P	25	3.35	1.30			
Time devoted to traditional activities	LC	54	4.14	1.16	0.47	.641	0.11
	P	24	4.00	1.23			
Teacher's use of Vietnamese	LC	55	4.04	1.41	-0.93	.357	-0.22
	P	26	4.35	1.38			

Note. LC = large city universities; P = provincial university.

5.6 Differences between the Large City Universities and the Provincial University in Factors Potentially Involved in Forming the Impact of TOEIC on Teaching

5.6.1 Differences between the large city universities and the provincial university in teachers' perceptions of TOEIC.

Table 5.11 Statistics on Differences in Teachers' Perceptions of TOEIC between Large City Universities and Provincial University

Test factors	Location	N	Mean scores	SD	t	p (2-tailed)	Cohen's d
Test value	LC	54	4.38	1.25	-2.75	.008	-0.71
	P	24	5.17	0.95			
Test purpose-tested skills	LC	51	5.49	1.02	-0.19	.847	-0.05
	P	26	5.54	1.07			
Test purpose-non-tested skills	LC	51	2.68	1.34	-1.41	.162	-0.34
	P	26	3.14	1.40			
Test difficulty	LC	53	5.14	1.03	-0.88	.381	-0.21
	P	26	5.36	1.07			
Test format	LC	54	6.26	1.01	0.70	.484	0.16
	P	26	6.08	1.23			
Consequences of the test for teachers	LC	54	3.02	1.20	-2.00	.049	-0.49
	P	25	3.59	1.12			
External expectations	LC	55	4.80	1.14	-2.63	.010	-0.67
	P	26	5.44	0.76			

Note. LC = large city universities; P = provincial university.

There were only three statistically significant differences in teachers' perceptions of various aspects of the TOEIC test between the large city universities and the provincial university: test value, consequences of the test for teachers, and external expectations (see

Table 5.11). Large city teachers considered the TOEIC test less valuable than provincial teachers. The former had a mean score of 4.38 with a standard deviation of 1.25; while the latter had a mean score of 5.17 with a standard deviation of 0.95; $t=-2.75$, $p=.008$. This difference was nearly large ($d=-0.71$).

Large city teachers also experienced less test pressure than provincial teachers. They perceived the test consequences for themselves slightly less serious than provincial teachers did. The first group had a mean score of 3.02 with a standard deviation of 1.20; while the second group had a mean score of 3.59 with a standard deviation of 1.12; $t=-2.00$, $p=.049$, two-tailed. This difference was nearly medium with Cohen's d being -0.49.

The external expectations were also smaller for the large city universities ($M=4.80$, $SD=1.14$) than the provincial university ($M=5.44$, $SD=0.76$; $t=-2.63$, $p=.010$, two-tailed). This difference was medium with Cohen's d being -0.67.

5.6.2 Differences between the large city universities and the provincial university in teachers' beliefs about language learning and teaching.

There were no significant differences between the large city universities and the provincial university in teachers' beliefs about language teaching and learning (see Table 5.12).

Table 5.12 Statistics on Differences in Teachers' Beliefs about Language Teaching and Learning between Large City Universities and Provincial University

Teachers' beliefs	Location	N	Mean scores	SD	t	P (2-tailed)	Cohen's d
Belief in communicative approach	LC	55	6.30	0.66	-1.05	.297	-0.27
	P	26	6.44	0.38			
Belief in traditional methods	LC	55	3.39	1.21	-1.20	.233	-0.30
	P	25	3.73	1.09			
Belief in the necessity of direct test preparation in improving test scores	LC	55	5.46	1.10	-0.14	.892	-0.03
	P	25	5.50	1.15			
Belief in overall improvement in proving test scores	LC	55	5.82	1.11	0.04	.968	0.01
	P	26	5.81	1.02			

Note. LC = large city universities; P = provincial university.

5.6.3 Differences between the large city universities and the provincial university in teachers' perceptions of context factors.

Table 5.13 Statistics on Differences in Teachers' Perceptions of Context Factors between the Large City Universities and Provincial University

Context factors	Location	N	Mean scores	SD	t	p (2-tailed)	Cohen's d
Constraints	LC	55	5.06	1.20	-0.72	.476	-0.17
	P	26	5.27	1.22			
Opportunities	LC	55	5.43	0.85	0.46	.644	0.11
	P	25	5.34	0.78			

Note. LC = large city universities; P = provincial university.

There were no significant differences in context factors between large city universities and provincial university (see Table 5.13).

In brief, large city and provincial teachers' perceptions of the factors potentially involved in test impact differed very little. There were absolutely no significant differences in their beliefs about language teaching and learning and the context factors. There were only significant differences in their perceptions of three test factors, namely test value, consequences for teachers, and external expectations with the mean scores for provincial teachers being higher.

5.7 Chapter Summary and Discussion

This chapter reported findings from the teacher questionnaire. It first presented the descriptive statistics on the perceived impact of the TOEIC test on teaching and teachers' perceptions of factors potentially involved in the generation of test impact. Next, it reported the results of the bivariate correlation analysis between the factors potentially involved in the generation of the test impact and the types of test impact. Finally, it provided readers with the findings of independent-sample t-test, showing the differences in the means between the large city universities and the provincial one.

The findings showed an obvious tendency of teaching to the test under the influence of TOEIC. This was clearly evident in teachers' direct test preparation activities and the changes in the time allotment to language skills and teaching and learning activities.

Contrary to popular belief that Vietnamese teachers inclined towards traditional teaching methods, the results showed that the teachers at the three universities favoured the communicative approach over them. This is probably because the communicative teaching approach is now the main focus of any language teaching methodology courses in Vietnam. Communicative textbooks are also widely available.

As for their perceptions of the test factors, teachers appeared to clearly understand the test design. They did not perceive the test as too difficult for their students. They less than “slightly agreed” that it was valuable. Also, they only felt a little test pressure from external expectations such as students’, administrators’ and employers’ expectations, but the test did not seem to have direct consequences for them.

Regarding the context factors, the findings suggested that teachers had some opportunities such as availability of teaching resources and freedom to use them, but they were a little constrained by limited instructional time and class sizes.

The extent to which the three groups of factors – person factors, test factors, and context factors – were involved in the generation of the test impact on teaching was shown by the bivariate correlation analysis results. The most important groups of factors in shaping the test impact on teaching were likely test factors and person factors. Only one of the context factors – the opportunities – had a considerable number of relationships with test impact.

The findings interestingly suggested that the older and more experienced the teachers, the less they taught to the test, and the more communicative their teaching approach was. Generally, teachers’ reported behaviours were logically consistent with their perceptions of the nature of the test, which supported Hughes’ (as cited in Bailey, 1996) ideas about the mechanism of test impact. The more they perceived that the test assessed the actually tested-skills, the more they taught to the test and the more traditional they were (i.e., being engaged in direct test preparation, teaching the tested skills, doing traditional teaching activities, and using Vietnamese). On the contrary, the more they perceived that the test assessed the actually non-tested skills, the more they taught the non-tested skills and the more communicative they were (i.e., using communicative teaching activities and using less Vietnamese).

The teachers in the large city and the province did not seem to differ much. Absolutely no statistically significant differences in their beliefs about language teaching and learning

and context factors were found. Teachers at the two locations were also very similar in their perceptions of the test, except for test value and test pressure. Provincial teachers more strongly believed in the value of the test and felt greater test pressure than the large city teachers did. Concerning the impact of TOEIC on teaching, the teachers at the two locations only significantly differed in direct test preparation with the provincial teachers being more engaged in it.

The teacher questionnaire data confirmed that the relationship between a test and stakeholders' behaviours was not straightforward because it involved a number of factors. It also again proved the important role of the test factors and certain person factors, namely age and teaching experience (Alderson & Hamp-Lyons, 1996; Cheng et al., 2004; Shohamy et al., 1996). However, while some studies have highlighted the importance of teachers' beliefs as a person factor in test impact generation (Cheng et al., 2004; Watanabe, 1996, 2000), in the current study, teachers' beliefs had very little connection with test impact. It is, therefore, necessary to explain this difference in findings. Was this because teachers in the present study were in consensus about their beliefs in the communicative approach, the questionnaire did not capture all aspects of teachers' beliefs, or the effects of teachers' beliefs were overridden by another factor?

It is also noticeable that the dichotomy between large city and province had little effect on teachers' teaching. This might be explained by the fact that teachers differed very little in their beliefs about language teaching and learning, their perceptions of the test, and their perceptions of context factors.

In brief, the teacher questionnaire findings showed the general patterns of the impact of TOEIC on teaching as well as revealed which factors were involved in shaping this impact, and which factors were most responsible for it. However, these results were only based on self-report data, so they might not completely reflect the real impact of the test. There were also several questions that still needed to be answered.

Chapter 6: STUDENT QUESTIONNAIRE FINDINGS

This chapter describes the results of the student questionnaire analysis. The student questionnaire was used to seek the answers to the following questions:

1. What are students' perceptions of the impact of the TOEIC test on university English learning?
2. What are students' perceptions of factors potentially involved in forming the impact of the TOEIC test on student learning?
3. To what extent are test factors, context factors, and person factors involved in shaping the test impact?
4. Are there differences between large city universities and provincial universities in students' perceptions of the impact of the TOEIC test on learning?
5. Are there differences between large city universities and provincial universities in factors potentially involved in forming the impact of the TOEIC test on student learning?

The report of the findings is organized around these research questions. Participants' responses to the open-ended question are not presented because although they were very interesting, they mostly repeated the information already obtained from the closed-ended questions. For information about the items making up the composite variables, refer to Table 3.11.

6.1 Demographic Information

After the questionnaire data had been screened and cleaned, responses from 1,445 students remained for data analysis. The participating students from each university almost equally made up one third of the sample. The majority of the students were in their second year and third year at university (52.2% and 42.1%, respectively). Male students exceeded females in number by approximately 8% (54.2% vs. 45.8%). The students had an average of 118.52 months of experience learning English (approximately 10 years), but very little experience preparing for TOEIC (16.37 months, i.e., 1.4 years).

Table 6.1 Demographic Information of Student Participants (N=1445)

Variables		Participants	Percentage
Universities	U 1	500	34.6
	U 2	462	32.0
	U 3	483	33.4
Gender	Male	781	54.0
	Female	661	45.7
	Missing	3	0.2
Year at university	First year	1	.1
	Second year	754	52.2
	Third year	608	42.1
	Fourth year	82	5.7

Table 6.2 Descriptive Statistics on Students' Learning Experience (N=1445)

Variables	N	Missing	Mean	Standard deviation
Experience learning English (in months)	1424	21	118.52	21.69
Experience preparing for TOEIC (in months)	1408	37	16.37	11.79

6.2 Perceived Impact of TOEIC on Student Learning

Seven composite variables and one single item variable with regard to test impact on learning were analysed. Descriptive statistics of these variables clearly showed that students did perceive the influence of the TOEIC test on most aspects of their learning although not to a very high degree (see Table 6.3). Generally, their perceptions seemed unanimous with regard to most aspects, except for “stress”, because the standard deviations for most of them varied from 0.83 to 1.09 only.

The test could have had a positive influence on students' motivation with the mean score for this variable being 5.3. However, students might also have felt a little stressed ($M=5.2$) although their responses on this aspect differed ($SD=1.6$).

Table 6.3 Descriptive Statistics on Perceived Impact of TOEIC on Student Learning (N=1445)

Test impact	N	Mean	SD
Motivation	1428	5.33	1.10
Stress	1439	5.17	1.58
Direct test preparation activities	1406	5.36	0.93
Time devoted to non-tested skills	1428	4.76	1.03
Time devoted to tested skills	1423	5.40	0.83
Time devoted to language use activities	1432	4.56	0.89
Time devoted to form-focused activities	1432	5.18	0.91
Achievement	1420	5.10	0.83

In addition, students' responses suggested that the TOEIC test impelled them to spend more time on learning the tested skills (listening, reading, grammar, and vocabulary) as well as the non-tested skills (speaking, writing, and pronunciation) than before. Similarly, they showed that students spent more time on both the form-focused activities and the language use activities than before (see Table 6.3). However, time devoted to the tested skills and the form-focused activities ($M=5.4$ and 5.2 , respectively) clearly increased more than time devoted to the non-tested skills and the language use activities did ($M=4.8$ and 4.6 , respectively).

Besides, the introduction of the TOEIC test may have driven students towards a variety of direct test preparation activities. These included learning test-taking skills, doing TOEIC past exam papers and/or mock tests, using TOEIC test preparation materials, finding information about the test, and seeking help from English teachers about the test ($M=5.4$, $SD=0.9$).

A positive influence of the TOEIC test on the improvement of students' language ability was also reported. The "achievement" variable had a mean score of 5.1 , which means a slight improvement.

In brief, the findings suggested that the introduction of the TOEIC test positively impacted on student learning. It reportedly enhanced students' motivation, made them learn both the

tested content and non-tested content more than before, and consequently improved their achievement. However, it is important to note that, according to students' responses, time devoted to what students perceived as being tested increased more than what they perceived as non-tested. Also, time spent on language forms had a bigger perceived increase than time spent on real language use. Additionally, the influence was probably not strong.

6.3 Factors Potentially Involved in Forming the Impact of TOEIC on Student Learning

6.3.1 Students' perceptions of TOEIC.

Table 6.4 Descriptive Statistics on Students' Perceptions of TOEIC (N=1445)

Test factors	N	Mean	SD
Test value	1398	5.24	1.01
Test difficulty	1378	4.97	1.03
Test purpose-non-tested skills	1423	4.19	1.42
Test purpose-tested skills	1426	5.38	0.95
Test format	959	5.78	1.22
Test pressure	1414	5.38	0.82

The following are the findings on students' perceptions of different aspects of the TOEIC test, namely the test value, test difficulty, test purpose, and test format. Students slightly agreed that the TOEIC test was valuable ($M=5.24$ and $SD=1.01$). The test value was composed of its accurate reflection of students' English proficiency, its usefulness for their future career, its useful feedback to their learning, its usefulness in helping them set goals for English learning, and its ability to measure the most important English knowledge and skills that university students should have.

Students also slightly agreed that the TOEIC test was difficult ($M=4.97$, $SD=1.03$). However, the listening section was perceived as more difficult than the reading section ($M= 5.13$ and 4.85 , respectively; $SD=1.39$ and 1.30 , respectively).

Students agreed that TOEIC measured what we called the actually tested skills, that is grammatical knowledge, listening skills, and reading skills ($M= 5.4$, $SD=1.0$). However, they could not decide whether the test assessed the actually non-tested skills (communicative ability, speaking skills, and writing skills) or not.

As shown in Table 6.4, the number of students answering the item about the test format was unusually low. This was due to a printing error. The item was removed from the questionnaire by mistake before printing. When the error was spotted, a number of copies had already been administered to students. The item was later added to the other copies manually. However, those students who responded to the item seemed to be more certain about the format of the test than its other aspects. They agreed that it had a multiple-choice format ($M=5.78$ and $SD=1.22$).

As for the test pressure, students generally agreed that the TOEIC test put some pressure on them, but the degree of pressure was not too high. The mean score of this variable was only 5.38. This opinion was probably shared by the majority of students with the standard deviation being only 0.82.

6.3.2 Students' beliefs about language learning.

It is interesting that students appeared to have practical beliefs about language learning. The results showed that they generally believed that language learning was for communication, and it was necessary to learn all four language skills – listening, speaking, reading and writing. The mean scores for these two items were the highest, both being nearly 6.1 ($SD=1.00$ and 1.13 , respectively). The belief in traditional methods, which considered the most important parts of language learning to be vocabulary words, grammar rules, and translation, received a lower mean score of 5.10 ($SD=1.02$). Thus, students were likely to have stronger beliefs in learning a language for practical use than for knowledge about the language. However, it is natural that they also somewhat agreed that vocabulary, grammar, and translation are important because these are necessary for communication. In addition, students agreed that both overall improvement in language learning and learning test-taking strategies resulted in improved test scores ($M=5.65$ and 5.61 , respectively).

Table 6.5 Descriptive Statistics on Students' Beliefs about Language Learning (N=1445)

Beliefs	N	Mean	SD
Belief in traditional methods	1422	5.10	1.02
Belief that language learning is for communication	1436	6.09	1.00
Belief in necessity of learning all four skills	1429	6.06	1.13
Belief that overall improvement results in improved test scores	1430	5.65	1.19
Belief that test-taking strategies are necessary to improve test scores	1435	5.61	1.19

6.3.3 Students' perceptions of context factors.

The context factors were composed of “availability of resources,” “teacher’s influence,” and “time constraints.” Students somewhat agreed that they had enough resources for their learning ($M=5.15$, $SD=1.00$). Their teachers had some influence on their decisions about learning but not to a large extent ($M=5.17$, $SD=1.12$). Time constraints did not seem to be a serious problem ($M=4.85$, $SD=1.42$). Unlike availability of resources and teacher’s influence, time constraints received less unanimous responses with the standard deviation being the highest (see Table 6.6).

Table 6.6 Descriptive Statistics on Students' Perceptions of Context Factors (N=1445)

Context factors	N	Mean	SD
Availability of resources	1425	5.15	1.00
Teacher’s influence	1429	5.17	1.19
Time constraints	1433	4.85	1.42

6.4 Involvement of Test Factors, Person Factors, and Context Factors in Forming the Impact of TOEIC on Student Learning

The Pearson product-moment correlation coefficient was used to determine whether there were any relationships between the three sets of factors including person factors, test factors, context factors and the impact of the TOEIC test on student learning. Below are the results of the correlation analysis for each set of factors.

6.4.1 Person factors and perceived impact of TOEIC on student learning.

The set of person factors was made up of students' experience learning English, their experience preparing for TOEIC, and their beliefs about language learning. The results of the correlation analysis revealed a considerable number of significant associations between these factors and the perceived impact of the test on student learning. However, most of these correlations were small, the largest coefficient being only .25 (see Table 6.7).

There were almost no significant correlations between experience learning English and experience preparing for TOEIC and the test impacts. This could be explained by the fact that the participating students were very similar in their experience learning English. The descriptive statistics showed that the average number of years learning English was approximately 10 years ($M=118.5$ months) with the standard deviation being approximately 2 years (21.7 months). As for experience preparing for TOEIC, it was still short for most students, except for the provincial students. The mean score for this item being only 16.37 months ($SD=11.79$).

A larger number of significant correlations were found between the belief in the traditional learning methods and the test impacts, but they were also small. This person factor had the strongest correlations with students' motivation, time devoted to form-focused activities and achievement, $r = .25, .23$, and $.24$, respectively; $n = 1405, 1410$, and 1399 , respectively; $p = .000$ for all the correlations.

The belief that the purpose of language learning is communication had significant correlations with most of the test impacts, except for motivation. However, all of these correlations were smaller than those between the belief in the traditional learning methods and the test impacts, r ranging from $.12$ to $.22$. This belief had the strongest correlation with the time students devoted to language use activities, $r = .22$, $n=1427$, $p=.000$. The second strongest correlation was with the time devoted to the tested skills, $r=.21$, $n=1416$, $p=.000$.

The belief that it was necessary for university students to learn all four skills also had small significant correlations with all the test impacts. The coefficients varied from $.11$ to $.21$, all being significant. This belief had the strongest correlations with time-devoted to non-tested skills and time devoted to form-focused activities.

The belief that overall improvement results in improved test scores had low correlations with all test impacts. The largest correlation was found between the belief and achievement, $r=.22$, $n=1407$, $p=.000$.

The belief in learning test-taking strategies for improving test scores also had small correlations with all test impacts. The largest correlations were with motivation and achievement ($r=.25$ and $.24$, respectively; $n=1418$ and 1411 , respectively; $p=.000$ for both the correlations). The smallest correlations were with time devoted to non-tested skills and time devoted to language use activities. It is interesting that believing in the role of test-taking strategies had a larger correlation with achievement than believing in overall improvement in improving test scores ($r=.24$ and $.22$, respectively).

In brief, the person factors being investigated did not appear to contribute a lot to the perceived test impact. Among these factors, the belief in traditional methods seemed to play the most prominent role.

Table 6.7 Statistics on Correlations between Person Factors and Perceived Impacts of TOEIC on Student Learning (N=1445)

		Motiva- tion	Direct test preparation activities	Time devoted to non-tested skills	Time devoted to tested skills	Time devoted to language use activities	Time devoted to form- focused activities	Achieve- ment
Experience learning English	Pearson Correlation	-.084**	.013	-.028	-.007	.049	-.073**	-.044
	N	1407	1385	1407	1402	1412	1411	1400
Experience preparing for TOEIC	Pearson Correlation	.048	.092**	.078**	.040	.057*	.040	.097**
	N	1391	1371	1391	1386	1396	1395	1384
Belief in traditional methods	Pearson Correlation	.254**	.163**	.194**	.105**	.069**	.235**	.241**
	N	1405	1386	1405	1402	1413	1410	1399
Belief that language learning is for communication	Pearson Correlation	.073**	.132**	.136**	.209**	.218**	.157**	.140**
	N	1419	1397	1419	1416	1427	1424	1412
Belief in necessity of learning all four skills	Pearson Correlation	.158**	.162**	.205**	.153**	.116**	.205**	.195**
	N	1412	1391	1412	1408	1420	1418	1406
Belief that overall improvement results in improved test scores.	Pearson Correlation	.178**	.154**	.186**	.136**	.131**	.127**	.217**
	N	1413	1392	1413	1409	1420	1418	1407
Belief that test-taking strategies are needed to improve test scores.	Pearson Correlation	.246**	.219**	.176**	.170**	.118**	.167**	.241**
	N	1418	1396	1418	1414	1425	1423	1411

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

* . Correlation is significant at the 0.05 level (2-tailed).

6.4.2 Perceptions of TOEIC and its perceived impact of on student learning.

The test factors probably had clearer relationships with the test impact than the person factors did (see Table 6.8). The “test value” variable had small to medium positive correlations with all the test impacts. This means that, generally, the more students believed that the test was valuable, the larger impact the test had on their learning. Medium correlations were found with motivation, direct test preparation activities, time devoted to non-tested skills, time devoted to form-focused activities and achievement with the coefficients being .44, .30, .33, .31, and .49, respectively (all of them were significant, $p=.000$). However, it is worth noticing that test value had lower correlation with time devoted to language use activities than with time devoted to form-focused activities, $r=.17$ and .30, respectively.

Test difficulty only had a positive medium correlation with stress ($r=.37$, $n=1,373$, and $p=.000$). This suggested that the more difficult students perceived the test to be, the more stress they suffered from, and the more they skipped the non-tested content. All the other significant correlations found were too small.

The perception that the purpose of the TOEIC test is assessing the actually non-tested skills had small positive correlations with many of the test impacts including motivation, direct test preparation activities, time devoted to language use activities, time devoted to form-focused activities (see Table 6.8). It had medium correlations with time devoted to non-tested skills and achievement, $r=.44$ and .42, respectively; $n=1,407$ and 1,400, respectively; $p=.000$ for all the coefficients). However, it did not correlate with stress and time devoted to tested skills.

The perception that the purpose of the TOEIC test is assessing the actually tested skills, generally, had larger correlations with the test impacts than the perception of the test as assessing the actually non-tested skills did. Those larger correlations were found between the perception and motivation, direct test preparation activities, time devoted to tested skills, time devoted to language use activities, time devoted to form-focused activities, and achievement (r ranging from 2.1 to 3.5). The correlations with time devoted to tested-skills and form-focused activities seemed obvious, but it is interesting that the correlations were also higher for time devoted to language use activities.

The perception of the test as assessing the actually tested skills also had a small correlation with time devoted to non-tested skills, $r=.21$, $n=1409$, $p=.000$. This suggested that, although students believed that the test evaluated them on listening, reading, vocabulary and grammar, they still spent time on the actually non-tested skills such as speaking, writing, and pronunciation. However, this correlation was smaller than the correlation between the perception of the test as assessing the actually non-tested skills and time devoted to the actually non-tested skills ($r=.44$). This difference clearly demonstrated that the phenomenon of “learning to the test” existed. It is strange that this perception also had a lower correlation with achievement than the other perception did with $r=.35$ compared to $r=.42$. Besides, it did not correlate with stress.

The perception that the TOEIC test had a multiple choice format had small correlations with motivation, test-preparation activities, time devoted to tested skills, time devoted to form-focused activities and achievement with the correlation coefficients being in the range from .13 to .15.

Test pressure positively correlated with all the types of impact, the correlation coefficients varying from .17 to .38, that is, from small to medium (see Table 6.8). It had the largest correlations with motivation, direct test preparation activities, achievement, time devoted to form-focused activities, $r=.38$, .34, .30 and .27, respectively. It had the smallest correlations with time devoted to language use activities, time devoted to non-tested skills, and time devoted to tested skills, $r=.17$, .23, and .24, respectively. It seemed unusual that the test pressure had a slightly smaller correlation with time devoted to tested skills than with time devoted to non-tested skills. Thus, these statistics showed that the test pressure did push overall student learning; however, it had stronger effects on students with regard to the tested content than the non-tested content.

Table 6.8 Statistics on Correlations between Perceptions of TOEIC and Its Perceived Impacts on Learning (N=1445)

		Motivation	Stress	Direct test preparation activities	Time devoted to non-tested skills	Time devoted to tested skills	Time devoted to language use activities	Time devoted to form-focused activities	Achievement
Test value	Pearson correlation	.443**	.107**	.303**	.334**	.251**	.172**	.308**	.491**
	N	1382	1392	1359	1383	1379	1388	1386	1374
Test difficulty	Pearson correlation	.088**	.369**	-.005	-.061*	-.012	-.122**	.010	-.066*
	N	1362	1372	1341	1363	1363	1370	1367	1359
Purpose-non-tested skills	Pearson correlation	.259**	.097**	.181**	.440**	.069**	.158**	.168**	.417**
	N	1406	1417	1384	1407	1402	1413	1411	1400
Purpose-tested skills	Pearson correlation	.270**	.065*	.320**	.205**	.340**	.205**	.313**	.351**
	N	1409	1420	1388	1409	1405	1415	1414	1402
Multiple-choice format	Pearson Correlation	.152**	.040	.130**	.079*	.139**	.058	.148**	.150**
	N	948	954	931	948	942	951	950	941
Test pressure	Pearson Correlation	.377**	.257**	.338**	.229**	.237**	.173**	.272**	.298**
	N	1397	1408	1376	1397	1395	1403	1401	1389

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

* . Correlation is significant at the 0.05 level (2-tailed).

In brief, the test factors had a larger number of significant correlations with the test impacts than the person factors did. The strength of the correlations was also higher. Test value, test pressure, and perception of the test as assessing actually tested skills seemed to make the most contribution to the formation of the test impact. Also, students' reported behaviours appeared to be consistent with their perceptions of the test design, e.g., the more they believed the test assessed the actually tested skills, the greater their time increase for them.

6.4.3 Perceptions of context factors and perceived impact of TOEIC on student learning.

The availability of resources and the teachers' influence variables had statistically significant correlations with more types of test impact than the time constraint variable did. However, the strength of the correlations was only in a small to medium range (see Table 6.9). Availability of resources had medium correlations with motivation ($r=.30$, $n=1409$, $p=.000$), direct test preparation activities ($r=.34$, $n=1386$, $p=.000$), and time devoted to language use activities ($r=.30$, $n=1414$, $p=.000$). It had small correlations with time devoted to non-tested skills, time devoted to tested skills, and time devoted to form-focused activities. This means that the more resources students had access to, the more they learned both for the test and for other purposes. However, availability of resources had no statistically significant correlations with stress.

Teacher's influence had small correlations with all the test impacts with the correlation coefficients ranging from .10 to .28. This means that teachers' influence did not have a large effect on students' reactions to the introduction of the TOEIC test. However, it is worth noticing that this variable had the largest correlations with motivation, achievement, and direct test preparation activities among the test impacts.

The time constraints variable associated with only a few types of test impact and the statistically significant correlations found were really small ranging from .10 to .16. This showed that time constraints were almost not involved in the formation of test impact. Overall, the context factors had only a small association with the impact of the TOEIC on student learning.

Table 6.9 Statistics on Correlations between Perceptions of Context Factors and Perceived Impacts of TOEIC on Student Learning (N=1445)

		Motivation	Stress	Direct test preparation activities	Time devoted to non-tested skills	Time devoted to tested skills	Time devoted to language use activities	Time devoted to form-focused activities	Achievement
Availability of resources	Pearson Correlation	.299**	.049	.335**	.224**	.240**	.303**	.265**	.331**
	N	1409	1419	1386	1408	1405	1414	1412	1400
Teacher's influence	Pearson Correlation	.284**	.155**	.273**	.215**	.229**	.173**	.243**	.276**
	N	1412	1423	1390	1412	1408	1418	1416	1404
Time constraint	Pearson Correlation	.091**	.121**	.110**	.027	.073**	.056*	.111**	.032
	N	1416	1427	1394	1416	1412	1422	1420	1408

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

* . Correlation is significant at the 0.05 level (2-tailed).

6.5 Differences between the Large City Universities and the Provincial University in the Perceived Impact of TOEIC on Student Learning

The independent-sample t-test was performed to compare the test impact scores for the large city universities and the provincial university. Overall, for most of the types of test impact, the provincial university had higher mean scores than the large city universities did (see Table 6.10). However, not all of the differences were statistically significant, and the effect size varied from small to medium only.

There were significant differences between the large city universities and the provincial university in motivation, stress, time devoted to non-tested skills, time devoted to tested skills, time devoted to language use activities and achievement. The difference in time devoted to tested skills had a medium magnitude ($d=.42$); all the other differences were small.

Table 6.10 Statistics on Differences in Perceived Impact of TOEIC on Student Learning between Large City Universities and provincial University

Test Impact	Location	N	Mean scores	SD	t	P (2-tailed)	Cohen's d
Motivation	LC	946	5.22	1.12	-5.89	.000	-0.32
	P	481	5.56	1.01			
Stress	LC	957	5.00	1.62	-5.91	.000	-0.33
	P	481	5.50	1.44			
Direct test preparation activities	LC	940	5.34	0.90	-0.90	.370	0.05
	P	465	5.39	0.99			
Time devoted to non-tested skills	LC	949	4.62	1.01	-7.50	.000	-0.42
	P	478	5.05	1.00			
Time devoted to tested skills	LC	952	5.47	0.76	3.67	.000	0.22
	P	470	5.28	0.95			
Time devoted to language use activities	LC	957	4.62	0.78	3.29	.001	0.20
	P	474	4.43	1.06			
Time devoted to form-focused activities	LC	954	5.17	0.82	-0.75	.455	-0.04
	P	477	5.21	1.08			
Achievement	LC	945	5.00	0.78	-5.94	.000	-0.34
	P	474	5.29	0.90			

Note. LC = large city universities; and P = provincial university.

6.6 Differences between the Large City Universities and the Provincial Universities in Factors Potentially Involved in Forming the Impact of TOEIC on Student Learning

6.6.1 Differences between the large city universities and the provincial university in students' perceptions of TOEIC.

Among the test perception factors, there were only significant differences in the means for test value, test purpose-non-tested skills, and test pressure (see Table 6.11). Provincial students considered the TOEIC test more valuable than large city students did, and the magnitude of the difference in the means was medium ($M=5.10$ and 5.52 , respectively; $SD=1.01$ and 0.96 , respectively; $t=-7.40$; $p=.000$, two-tailed; $d=-0.42$).

While large city students were almost neutral regarding whether the purpose of the TOEIC test was to measure the actually non-tested skills, provincial students slightly agreed that the test assessed these skills ($M=3.81$ and $=4.93$, respectively; $SD=1.34$ and 1.29 , respectively; $t=-15.14$; $p=.000$, two-tailed). The magnitude of the difference was large ($d=-0.86$).

Provincial students reported greater test pressure than large city students. This difference was statistically significant, but its magnitude was small ($M=5.56$ and $=5.30$, respectively; $SD=0.83$ and 0.81 , respectively; $t=-5.60$; $p=.000$, two-tailed; $d=-0.32$).

Table 6.11 Statistics on Differences in Students' Perceptions of TOEIC between Large City Universities and Provincial University

Test factors	Location	N	Mean scores	SD	t	P (2-tailed)	Cohen's d
Test value	LC	932	5.10	1.01	-7.40	.000	-0.42
	P	465	5.52	0.96			
Test difficulty	LC	913	3.05	1.03	1.29	.196	0.07
	P	464	2.98	1.03			
Test purpose-non-tested skills	LC	945	3.81	1.34	-15.14	.000	-0.86
	P	477	4.93	1.29			
Test purpose-tested skills	LC	946	5.36	0.91	-0.95	.341	-0.06
	P	479	5.42	1.03			
Test format	LC	485	5.81	1.22	0.73	.465	-0.05
	P	473	5.75	1.22			
Test pressure	LC	947	5.30	0.81	-5.60	.000	-0.32
	P	466	5.56	0.83			

Note. LC = large city universities; and P = provincial university.

6.6.2 Differences between the large city universities and the provincial university in students' beliefs about language learning.

There were significant differences in the means of most beliefs about language learning for the large cities and the provincial university, apart from the belief that overall improvement results in improved test scores. However, the effect sizes were mostly small. (See Table 6.12.)

The statistics showed that provincial students believed in the traditional methods more than large city students did ($M=5.53$ and 4.89 , respectively; $SD=0.95$ and 0.99 , respectively; $t=-11.61$; $p=.000$, two-tailed). This is the only difference that had a medium magnitude size ($d=-0.66$). Provincial students also believed less that language learning was for communication ($M=5.96$ and $=6.15$, respectively; $SD=1.05$ and 0.96 , respectively; $t=3.39$; $p=.001$, two-tailed; $d=1.9$).

The differences above suggested that provincial students had more traditional beliefs about language learning than large city students did. However, they believed in the necessity of learning all four skills to a larger extent than large city students did ($M=6.29$ and $=5.95$, respectively; $SD=1.00$ and 1.18 , respectively; $t=-5.45$; $p=.000$, two-tailed; $d=-0.31$).

Generally, provincial students also had a stronger belief in the value of learning test-taking strategies to improve test scores ($M=5.82$ and $=5.50$, respectively; $SD=1.06$ and 1.24 , respectively; $t=-5.07$; $p=.000$, two-tailed; $d=-0.28$).

Table 6.12 Statistics on Differences in Students' Beliefs about Language Learning between Large City Universities and Provincial University

Beliefs	Location	N	Mean scores	SD	t	P (2-tailed)	Cohen's d
Belief in traditional methods	LC	949	4.89	.99	-11.61	.000	-0.66
	P	472	5.53	.95			
Belief that language learning is for communication	LC	956	6.15	.96	3.39	.001	0.19
	P	479	5.96	1.05			
Belief in necessity of learning all four skills	LC	951	5.95	1.18	-5.45	.000	-0.31
	P	477	6.29	1.00			
Belief that overall improvement results in improved test scores	LC	951	5.60	1.20	-1.91	.056	-0.11
	P	478	5.73	1.17			
Belief that test-taking strategies are necessary to improve test scores	LC	955	5.50	1.24	-5.07	.000	-0.28
	P	479	5.82	1.06			

Note. LC = large city universities; P = provincial university.

6.6.3 Differences between the large city universities and the provincial university in students' perceptions of context factors.

There were no significant differences in scores for large city students and provincial students regarding all the context factors: teachers' influence, time constraints, and availability of resources (see Table 6.13).

Summary of Section 6.6

Most of the statistically significant differences between large city students and provincial students in the factors potentially involved in the formation of the test impact were small. There were almost no significant differences in the context factors. The largest differences were in the perception of the test value, the perception of the purpose of the test as assessing the actually non-tested skills, and the beliefs in traditional learning methods. Generally, provincial students considered the TOEIC test more valuable than large city students did. They also more strongly believed that it assessed the actually non-tested skills. Besides, they considered traditional learning methods to be more important than large city students did.

Table 6.13 Statistics on Differences in Students' Perceptions of Context Factors between Large City University and Provincial University

Context factors	Location	N	Mean scores	SD	t	P (2-tailed)	Cohen's <i>d</i>
Availability of resources	LC	951	5.11	1.00	-1.95	.052	-0.11
	P	473	5.22	1.01			
Teacher's influence	LC	953	5.14	1.20	-1.61	.109	-0.09
	P	475	5.24	1.16			
Time constraints	LC	954	4.89	1.43	1.69	.090	0.10
	P	478	4.76	1.41			

Note. LC = large city universities; and P = provincial university.

6.7 Chapter Summary and Discussion

This chapter reported findings from the student questionnaire. It first presented the descriptive statistics on students' perceptions of the impact of the TOEIC test on their learning and factors potentially involved in the generation of test impact. Next, it reported the results of the bivariate correlation analysis between the factors potentially involved in the generation of the test impact and the test impacts. Finally, it provided readers with the results of independent-sample t-test, showing the differences in the means between the large city universities and the provincial one.

The descriptive statistics showed that, generally, students perceived a positive impact of the TOEIC test on their learning although this influence was probably not very strong. According to students, the test motivated them to learn and made them spend more time on both the tested skills and non-tested skills, and both formed-focused and language use activities. It also engaged them in direct test preparation activities. Consequently, their achievement was also improved. However, it is noticeable that the reported time increase for the tested skills and formed-focused activities was larger than that for the non-tested skills and language use activities. Also, students only less than slightly agreed that the latter was increased.

Additionally, students appeared to have practical beliefs about language learning. They believed that language learning was for communication, and it was necessary to learn all four language skills. They also thought that both test-taking skills and overall improvement helped improve test scores. They seemed less supportive of the traditional learning methods.

Furthermore, students perceived the TOEIC test to be somewhat valuable (useful for employment and study and valid). They agreed that the test assessed the actually tested skills and doubted if it could assess the actually non-tested skills. The test did not seem too difficult for them, and its pressure was not very high. As for the context factors, students seemed to have enough conditions for learning, except the time constraints.

The bivariate correlation analysis revealed that the test factors, especially the perceptions of test value, test purposes and test pressure, were most responsible for the test impact on students' learning. The next most influential factor was possibly availability of resources.

Person factors seemed to have little connection with the test impact. Only beliefs about language learning had small significant correlations with the test impacts.

Independent sample t-test showed that the perceived test impact on the large city universities was stronger than that on the provincial one in terms of impact on motivation, time-devoted to skills, time devoted to language use activities, and achievement. However, all of these differences were small in effect size.

There were also significant differences in beliefs about language learning between the large city universities and the provincial one. Most of the differences were small. However, the two groups differed considerably in their belief about the traditional methods with the mean score for the provincial university being higher.

The two locations were possibly different in their perceptions of TOEIC. Again, opinions were generally stronger in the province, especially the perception that TOEIC measured the actually non-tested skills. There were no differences, however, in the context factors.

The student questionnaire data confirmed the complexity of test impact found in previous studies. In the current research context, all the three sets of factors – test factors, person factors, and context factors – were shown to be involved in the generation of the test effects. However, their involvements differed in strength. Test factors were the most associated with test impact in terms of the number of statistically significant relationships and their strength. These relationships supported Hughes' (as cited in Bailey, 1996) ideas about the mechanism of washback. According to Hughes, how participants perceived the nature of a test may influence their actions, and then their actions may influence their learning outcomes. The data showed an obvious consistency between students' reported actions and their perceptions of the purpose of the TOEIC test: Generally, they reported that they focused more on the tested skills than on non-tested skills, and made more use of formed-focused activities than language use activities. The other test factors such as test value and test pressure also had the strongest associations with test impact.

Some previous studies (Alderson & Hamp-Lyons, 1996; Watanabe, 1996) have showed that person factors play a central role in the generation of test impact on teaching. However, in the current study, the person factors appeared to have little association with test impact on students' learning. Students' beliefs had low correlations with their reported actions. This might show that the effects of the nature of the test overrode the effects of

personal beliefs. This might also show that the person factors used in the study were not comprehensive enough. Other person factors, for example, students' learning background and their language proficiency, need to be taken into account. These factors were, therefore, examined in Phase 2 of the study.

The results also somewhat supported the still limited findings about the role of school settings. There were statistically significant differences in many aspects of test impact between the large city universities and the provincial one. However, these differences were mostly small in size. As there were no statistically significant differences between the two locations in the context factors used in the study, these context factors could not explain those differences. This suggests that school settings might embrace other contextual aspects on top of availability of resources, influence of teachers or peers, and availability of time. Factors such as a school's policies, its English programmes, and textbooks might also have important roles.

Thus, the questionnaire findings showed the general patterns of the impact of TOEIC on student learning as well as revealed which factors were involved in shaping this impact, and which factors were most responsible for it. However, these results were only based on self-report data, so they might not completely reflect the real impact of the test. There were also a lot of unanswered questions. Some of them are *Why was the time devoted to non-tested skills also increased? Why did time devoted to form-focused activities increase more than time devoted to language use activities? Why was the test impact on provincial students stronger than that on large city students?* In order to fully understand and explain both the main trends and the variability, it is necessary to examine the phenomenon more closely. As indicated above, more in-depth data is needed, and more factors have to be taken into consideration. These can only be done through qualitative research methods. Hopefully, the case study in Phase 2 can help answer these questions.

PHASE 2 FINDINGS

The two chapters in this part report the findings from Phase 2. Chapter 7 presents the findings from the teacher case study, and Chapter 8 reports the results of the student case study.

To assist readers in understanding participants' perspectives and behaviours, a brief description of the English courses that they were teaching/doing and were observed by the researcher is provided here. These courses included Business English 3 at University A, English 4 at University B, and TOEIC 5 and TOEIC 3 at University C.

Business English 3 was one of the four General Business English courses at University A, which were also called TOEIC-oriented Business English courses. It used *Market Leader Pre-intermediate* (Cotton, Falvey, & Kent, 2012) as the course book. This was an integrated skills course book written for businesspeople and students of business English. It was accompanied by a work book called *Practice File*, which provided further vocabulary, grammar, and pronunciation practice. Besides *Practice File*, there was another workbook called *Practice Book*. This work book was compiled by the university's Foreign Language Department to give students further practice for the topics learned in the course book and a small proportion of TOEIC preparation. Course grades included scores from attendance, classroom assessment (mid-term speaking test, homework, quizzes, etc.), and the common final test, which was administered to all the students attending the same course at the same time. Before the TOEIC policy, the university taught four English courses: two General Business English and two ESP courses. After the TOEIC test was introduced to its language program, the two ESP courses were replaced by another two General Business English courses.

English 4 at university B was a test preparation course, which was preceded by three General English courses. It used *Longman Preparation For The New TOEIC Test* (Lougheed, 2007) as course book. The book aimed at improving students proficiency in English and test-taking skills (Lougheed, 2007, p. viii). It included information about the structure of the test, tips for taking the test, guidelines for teaching each part of the test, exercises focusing on each part of the test, and full practice tests. The course grades included scores from the mid-term listening test (20%), mid-term reading test (20%) and final test (60%). All of the tests had the TOEIC format and content. The mid-term tests

were compiled by classroom teachers and taken in class. The final test was compiled by the Language Centre and taken by all English 4 students at the same time. Like University A, before the TOEIC policy, University B taught two General English courses and two ESP courses. After the introduction of the TOEIC test, the ESP courses became optional and were replaced by one General English course and one TOEIC preparation course in the university's compulsory English programme.

TOEIC 5 was the last course in a series of test preparation courses taught at the Basic English Department of University C. The textbook used for the course was compiled by the university themselves using TOEIC preparation materials available on the market. It included three full practice tests and seven speaking topics with suggested questions for students to prepare for their presentations on the topics (see Appendix J). The course grade included scores from class attendance (10%), midterm speaking test (30%), and end-of-term test (60%). The midterm speaking test took place in class and involved group presentations on one of the topics included in the textbook. The end-of-term test was a complete simulated TOEIC test. TOEIC 3 was similar to TOEIC 5 in terms of the textbook content and course assessment, but it also included a language focus section with nine grammar topics and had a lower level of difficulty. Before the TOEIC policy, General English courses were taught instead of test preparation courses.

Observations showed that there were almost no differences in teaching and learning facilities at the three universities. Classrooms were equipped in the same way. Each one had a board, a projector, some microphones and loudspeakers. Desks and benches were of the same kinds and arranged in the same way, which actually was not very suitable for communicative language teaching activities.

Information was obtained from both the teachers and the students by semi-structured interviews. As the purpose of Phase 2 was to get in-depth information about test impact and explain the patterns found in Phase 1, questions similar to those in the questionnaires were used, but further details were also obtained by follow-up questions. Thus, participants were asked about the factors that might influence their teaching/learning such as their personal circumstances, educational background, teaching/learning experience, beliefs about language teaching and learning, and perceptions of the test as well as the influences of peers and other on their teaching/learning. They also reported how the TOEIC test affected their teaching/learning. In order to clearly see the impact of the test, they were

asked to compare their teaching/learning content and methods in their TOEIC and non-TOEIC classes. The non-TOEIC classes used for comparison were different for each teacher depending on their teaching experience and the universities language programmes. Usually, those were lower level classes (at University B) or classes they had taught at their universities before the TOEIC policy (at Universities A and C). However, as Phong and Quyen did not teach at University C before the introduction of the TOEIC test, they were asked to compare their current teaching with that in communicative English classes they had taught at foreign language centres. Thoai described how he would teach the same classes without TOEIC. Students were asked to compare how they learned English at university and at school.

Apart from the interviews, the teachers were observed in their current classes to find out what they actually did under the influence of the TOEIC test. After the observations, there were post-observation interviews about the moments of interest in order to understand why the teachers did what they did in the classroom. Concerning the students, they were asked to complete four weekly journal sheets during a four-week period to record their actual learning, perceived achievement, and feelings.

Chapter 7: TEACHER CASE STUDY

This chapter describes the results of the teacher case analysis. The case study of teachers was used to seek their perspectives regarding the following questions:

1. What is the impact of the TOEIC-as-exit-test on university English teaching?
2. What brings about the test impact on university English teaching?

The chapter includes two parts: a brief report of the results of the general data analysis and a qualitative presentation of three selected cases. The first part begins with the teachers' perceptions of the factors potentially involved in forming test impact on teaching, and then moves on to depict the actual impact of the test and offer an explanation of the way it occurs. The second part presents three teachers' stories in detail.

7.1 Demographic Information

Table 7.1 Demographic Information of Teacher Cases

Participants	University	Gender	Qualification	Age	English teaching experience at university (in years)	TOEIC teaching Experience (in years)
Phuong	A	Female	MA	Above 40	17	3
Thoai	A	Male	MA	20–30	6	6
Hoa	B	Female	MA	31–40	15	2
Lan	B	Female	MA	Above 40	19	NI
Phong	C	Male	BA	20–30	1.5	1.5
Quyen	C	Female	MA	20–30	2	4
Kim	C	Female	BA	Above 40	15	5 or 6

Note. NI = No information

Table 7.1 shows demographic information of individual teachers participating in the case study. Phuong, Lan, and Kim were the oldest and most experienced teachers. Thoai, Quyen, and Phong were the youngest and least experienced. Hoa was in the middle in terms of age and teaching experience. However, the teachers were not much different in

terms of TOEIC teaching experience, with the range being from 1.5 to 6 years. Most of the teachers held an MA; Phong and Kim only had a BA. The teachers from the same university taught the same English course, except for Kim, who taught a lower level course at the time of the research compared to her colleagues.

7.2 A General Analysis of Teacher Cases

The interview and observation data of the teacher cases were mainly analysed quantitatively with a small proportion of qualitative description of teacher's talking time and teacher's use of Vietnamese. The interview analysis involved factors potentially involved in teaching (teachers' beliefs about language teaching and learning, their perceptions of the TOEIC test, and the context factors) and the impact of the TOEIC test on teaching content and teaching methods. The observation analysis was done regarding TOEIC and non-TOEIC content and teaching methods (the use of participant organization, the use of teaching and learning activities, teacher's talking time, and teacher's use of Vietnamese).

The results of the interview analysis showed that all the teachers were advocates of the communicative approach. They believed that the purpose of language teaching and learning was communicative ability. However, for more than half of them, communicative ability was only the ability to communicate verbally, which included listening and speaking. Most of them considered direct test preparation, including teaching test-taking strategies, to be important for improving students' test scores.

The teachers perfectly understood the TOEIC test design. They knew it had only two sections – listening and reading. Many of them also mentioned grammar and vocabulary as the skills and knowledge it could also assess through listening and reading. More than half of them stressed that it did not assess communicative ability. It is not surprising as these teachers were those who considered communicative ability as ability to communicate verbally. The majority of them were concerned about the validity of the test, which they understood as the ability to measure overall communicative competence rather than only listening and reading. Despite their concerns about the test validity, the teachers supported the use of TOEIC as an exit test because it was most suitable for students among other international test options in several aspects, especially its difficulty. However, some of them stated that they would prefer a test that included speaking skill assessment. The test

was perceived to be very high-stakes for students, but not for teachers as teachers were not held accountable for students' test results.

The teachers generally had favourable conditions for teaching. They felt that they had enough teaching materials and equipment. Most of them only complained about the time constraint. They reported that their colleagues influenced their teaching through the exchange of ideas about teaching methodology, mid-term assessment, and teaching materials, but the influence was not substantial.

The interviews also revealed that the TOEIC test obviously impacted on all teachers' teaching, but its effects on individual teachers differed in extent. There was a general tendency of teaching to the test: focusing on the tested content more than on the non-tested content, spending more time on the tested content than before TOEIC, changing teaching methods due to the nature of the test and/or nature of TOEIC preparation course. However, this general tendency did not seem to be applicable for two of the teachers – Phuong and Kim. They only shared with the other teachers the fact that they spent more time on listening and, in Kim's case, reading as well. Time for the other skills was reported to be unaffected by the TOEIC test. Phuong and Kim also asserted that their teaching methods remained the same because their beliefs about teaching and learning had never changed.

Observation data generally confirmed what the teachers said. However, it revealed more clearly big differences between the Business English course and TOEIC preparation courses. At University A, most of the class time was spent on non-TOEIC content (from 72.55% to 80.20%), and a small amount of time was spent on TOEIC and non-TOEIC workbook (from 19.81% to 27.45%). By contrast, at the other universities, especially University B, the majority of time was devoted to TOEIC content (from 80.11% to 100%) (Kim was an exception). Moreover, teaching at University A was characterized by features of the communicative approach. In contrast, at Universities B and C, there was almost no communicative teaching in the majority of cases. Thus, at University A, there was some pair work and group work, but at Universities B and C, they were totally absent. At University A, there were a variety of communicative, language use activities such as pair/group discussions, role-plays, questions–answers, summarizing reading/listening texts, presentations, but none of them were used at University B, and only presentations and questions–answers were used at University C for non-TOEIC content. Also, at University A, instruction was almost entirely done in English, and students had the opportunity to

have meaningful communication with their friends and teachers. By contrast, at Universities B and C, Vietnamese was generally used most of the time, and teachers talked a lot, while students only gave single word replies to their teachers. There were some exceptions, however. Lan, although similar to most of the other teachers who were teaching TOEIC preparation, used English a lot. Kim did a lot of speaking activities in class and also used a lot of English. However, Lan said she spoke English much more in her general English classes. Certainly, Kim's teaching was determined by her strong beliefs about language teaching. However, the differences found in her teaching might also be due to the fact that she taught TOEIC 3, a lower level, where she had more time for activities other than TOEIC preparation. In addition, she was not observed long enough to gain conclusive evidence (for the amount of time that each teacher was observed for, see Table 3.15). From the general analysis of the teacher cases, course factors probably emerged as key contributors in shaping the patterns of test impact on teaching. (For more detailed results of the analysis, see Appendix O on the accompanying CD-ROM.)

7.3 Narratives of Teacher Cases

Three of the seven teacher cases were selected to be presented in this section: Phuong, Lan, and Phong. Phuong represented the teachers from the Business English course, while Phong and Lan – the TOEIC preparation courses. Although Phong and Lan taught a similar type of course, they were both chosen to show the role of the school settings (large city university vs. provincial university) and age, two factors of interest, in shaping test impact. Also, these cases best illustrated the different patterns of test impact on teaching and most clearly revealed factors involved in shaping it and their interaction. Last but not least, these cases provided the researcher with the richest and most detailed information about their teaching.

Each story begins with the participant's personal and educational background, their beliefs about language teaching and learning, and their perceptions of the TOEIC test and TOEIC benchmarks. Then it depicts the impact of the test on their teaching through the interview and observation data. It also discusses the roles of the influencing factors that emerged from their stories as well as their interaction to shape the test impact. It concludes with a short summary of the case.

7.3.1 Case 1: Phuong.

Personal and educational background

Phuong was in her late 40s and a very experienced teacher at University A. She got her first BA in Russian and had been teaching the language for two years when Russian was stopped being taught at the university. Like many other Russian teachers at that time, Phuong went back to university and trained to become an English teacher. After getting a BA degree in English, she came back to University A and taught English there. A few years later, she took an MA course in TESOL. When I conducted my research with her, she had been promoted to the position of Associate Head of the Foreign Language Department. At University A, she had taught General Business English as well as ESP courses for different departments. Since the introduction of the TOEIC benchmark, the university's English programme included only General Business English with a small proportion of TOEIC preparation. Besides teaching a little TOEIC at the university, she had had approximately three years of experience with it at a foreign language centre. She told me she had liked the teaching job since she was a child, and she seemed to be a very dedicated teacher. She often let students practice speaking in class although she knew that this meant a heavier workload: "Of course, my way of teaching is tiring. Others only explain quickly so that students understand, so they can finish the programme earlier. But I want students to speak a lot, so it takes a lot of time." (She told me she often had to teach one extra session at the end of courses). She talked to me about how she modified her teaching to better meet her students' needs through the insight she gained from her teaching experience.

In the past, I didn't like translation. It takes a lot of time. But now I've found that translation is needed for certain texts. From my experience, I know what students need, and what they often misunderstand, so I ask them to translate problem spots [words or phrases].

From her own research, Phuong found that students enjoyed making presentations and found the activity useful for improving their language ability, so she incorporated this extra activity into her class. She was also the first person in her department to use PowerPoint in her lessons to save time and accommodate weaker students.

Beliefs about language teaching and learning

Phuong was a strong advocate of the communicative approach. She was convinced that the purpose of language teaching was communicative ability. This was partly due to her own shortcomings as a language learner. The following is her explanation of why it was necessary to teach students to communicate:

While we were at school, we learned the traditional way: reading – translation, reading – translation, reading – translation. We didn’t practice the other skills. Now that we know our weaknesses in the past, we should help students avoid them. (...) I want students to master the four skills – listening, speaking, reading, and writing and achieve the communicative purpose, so that later when they meet foreigners, they can communicate; and when they read books in English, they can understand them.

To Phuong, “communicative competence is the four skills blended together.” She also believed that all four skills were equally important: “I don’t give priority to any skills, but I don’t have enough time. Usually I teach speaking, reading, and listening in class. I ask students do writing [at home].” However, she paid more attention to speaking and listening: “I usually focus more on listening and speaking. I focus on the skills that students are weak at.”

In her opinion, in order for students to achieve communicative competence, teachers had to “teach communicatively”. Therefore, she always required students to practice speaking at the beginning of each lesson. She also asked them to prepare for the reading section at home to reserve time for speaking activities because the instructional time was very limited.

Although she believed that “learning is not to pass exams, but to get knowledge” and did not like the way people learned only to pass exams, she also cared about preparing students for exams. She said, “I prepare end-of-unit tests in the same format as the final test to familiarize students with the final test.” She also thought it was necessary to gradually introduce TOEIC content and tasks into the programme to prepare students for the TOEIC test using Practice Book.

Perceptions of TOEIC and TOEIC benchmark

As for her perceptions of the TOEIC test, like other teachers, Phuong knew that it assessed the listening and reading skills, which involved grammar and vocabulary knowledge. She was unsure about the validity of the TOEIC test. In her opinion, the test might give

accurate evaluation of the majority of students' language proficiency. However, she noticed cases of discrepancy between test scores and real language ability.

I don't really know, but there are... Possibly it's accurate for the majority of students. But there are some cases when students said they had a test score of 400 or 500, but they can't speak when they make presentations. Although they have a [TOEIC] certificate and are exempt from the English subject, they asked to attend my class. I asked them to present, but they didn't speak as well as the good students in my class (Phuong).

Despite her doubt about the validity of the test and her concern that the test did not assess speaking and listening, Phuong still believed that the TOEIC benchmark was useful for students and suitable for graduation purposes. She explained, "When we require them to take the test, they'll have to focus more on listening and reading. (...) When they listen and read more, they'll improve their English ability." In addition, TOEIC was most suitable compared to other international tests like TOEFL or IELTS because it assessed language for the workplace, and its difficulty was suitable for students' language ability.

While the TOEIC test was high-stakes for students, teachers did not have to be responsible for students' test results. Phuong perceived her teaching tasks as teaching Business English and introducing the TOEIC test to students through some exercises in Practice Book compiled by her department. Students were responsible for their own test preparation. She argued that the limited instructional time did not allow teachers to prepare students for the test. She added that if students' entering levels were good, and they knew how to learn, they would be able to achieve the benchmark; otherwise, they should do more self-learning. It is necessary to note that this is also the department's policy, which was clearly manifested in its syllabus and course materials.

Test impact on teaching

Phuong admitted that the TOEIC test did affect her teaching, but its influence was only small because, as discussed above, she considered her main task teaching the integrated skills textbook Market Leader.

Actually this book [Market Leader] doesn't include TOEIC at all. It is integrated. The skills are integrated with each other, so we focus on teaching students the four skills – listening, speaking, reading, and writing.

For this reason, she did not feel stressed or pressured about the test. She only saw the impact of the test on her teaching in her gradually familiarizing students with the test through Practice Book and her spending more time on listening than before TOEIC.

For example, now that I've known that students will take the TOEIC test later, I spend more time on listening than before. In the past, I taught listening every now and then. But now I have students listen more.

According to Phuong, compared to the General Business English courses she taught before TOEIC, now her time allocation for some other skills had been changed, too. However, she attributed reasons other than TOEIC to the changes. For example, she taught speaking more than before because "the textbook was professionally designed. (...) It gives students opportunities and ideas to speak". Also, she wanted to make up for what students lacked: "Up to now, students have been speaking too little, too poorly, so I increase the time so that they can speak more."

Phuong said because she spent more time on speaking, she had to reduce the time for vocabulary and grammar. She explained that she chose to spend less time on grammar because students had learned a lot of grammar at school, while grammar in Business English was simple. Therefore, she asked students to learn grammar themselves, using the files at the end of the book, and controlled their learning by end-of-unit tests. As for vocabulary, she taught students how to learn it themselves at the beginning of the course and asked them to prepare for vocabulary at home. In class, she often asked students to give the meanings of new words, and only gave explanations herself when they misunderstood the words or did not know them.

As for teaching methods, Phuong reported that there were some changes in time allocation to teaching and learning activities, too, but again, these were not associated with the TOEIC test. For example, she said that she spent more time on group discussions, but this was "not due to TOEIC, but the textbook. And I've found that students' needs for speaking have increased." Time for some other activities such as information gap or survey activities, whether increased or stayed the same, was also determined by the textbook: "This textbook has those activities; I don't design more. Because the textbook has them, I let students do them."

As for translation, the time for it remained the same. Phuong also excluded TOEIC as a reason for this. She said she only translated when students did not understand. Her use of

translation was due to her own beliefs about language teaching and learning and students' needs.

Although Phuong reported some changes in the time allocation to teaching and learning activities, she asserted that the TOEIC test affected the content of her teaching, but not her teaching method, and she still taught the language skills in the same ways. She said, "I think my teaching method hasn't changed. The issue is only in what I'm teaching more." She explained that her teaching method had not changed due to her unchanged beliefs about language teaching and learning.

Observations supported what Phuong said about the little influence of TOEIC on her teaching. They showed that she spent most of the instructional time (80.19%) on non-TOEIC content (Market Leader textbook and Practice Files, Non-TOEIC self-designed materials and activities, end-of-unit tests, which dealt with the textbook content). She spent only 19.81% of class time on Practice Book, which included both TOEIC test preparation materials and other supplementary General Business English materials.

As for participant organization, Phuong spent 84.72% of her class time on whole class work, 12.28% - on pair/group work, and 2.99% on individual work. It is necessary to note that she reserved much more time for pair/group work than the other teachers I observed.

Phuong also used English most of the time in class. She gave students a lot of opportunities to practice speaking. Every class session began with students working in pairs, discussing the content they had learned in the previous session. After pair discussion, some students came to the front and presented their conversations or ideas to the class. Besides, there were a lot of discussions in the pre-listening/reading and post-listening/reading tasks (Nevertheless, it is important to point out that the majority of these activities were included in the textbook). Furthermore, unlike in TOEIC classes, students in Phuong's class did have meaningful communication with their classmates and their teacher. The following excerpt from her teaching shows these features.

Materials: Market Leader, page 22 (see Appendix K)

Task: Students worked in pairs to review the lesson learned in the previous session.

Episode 1: Teacher assigned and explained task

T: OK, Now let's continue. Unit 3. Now first one: review. Answer the question: What kinds of company you would like to work for. You have three options: family owned

company, multinational company and your own company. Work in pairs, make questions and answer the questions. What criteria do you depend on to answer the questions? What are the criteria? What?

S: Work environment.

T: What?

S: Work environment.

T: Yeah, work environment. What's next?

S: Salary and pay.

T: Salary. OK! What's next?

S: Promotion?

T: Promotion possibility. And the last one?

S: Job security.

T: Job security. OK. Talk to each other. Explain the reasons why you choose this kind of company to work for and you don't want to work for another kind of company.

Episode 2: Students worked in pairs.

Episode 3: Teacher called on some students to present their conversations in front of class.

T: Van. Mai.

S1: Hello, everybody. My name is Mai.

S2: Hello, everybody. My name Van. Hello, Mai. Today I (()) question. Which of company would you like work for?

S1: I like work for my company.

S2: Oh, and why?

S1: When I work for my company, I can do everything I want and I can be creative. And you?

S2: Uh. I like work for a multinational company.

S1: Why?

S2: Uh. Because I think when I work for the multinational company, I... uh I can work everybody. I can work a lot of people all around the world. Uh... and when I work for the multinational company, I can have high salary, and earn a lot of money, and I have a lot of flexible time. In... in the multinational company, I uh... um... uh... I...

Episode 4: Teacher and student interacted with each other and negotiated meaning.

T: OK. You have listened to your friends. One wants to work for herself. The other one wants to work for a multinational company. Do you have any questions to the two presenters? Any questions? (()) What should you do to get enough capital to set up your own company? You. You said that you want to work for yourself? (()) You want to operate your own company? Right?

S1: Yes.

T: What would you do? What would you do to get enough capital to set up your own company? Do you understand my question?

S1: Sorry, I don't understand.

T: You said that you want to set up your own company? Yes?

S1: Yes.

T: You work for yourself, right?

S1: Yes.

T: Yeah. How can you get enough money, capital, to set up your own company?

[The teacher asked the students some more questions and had a short discussion with the whole class about the issue.]

Note. (()) = unclear

As can be seen in the excerpt, the teacher and students used English all the time. The students had a chance to discuss questions in pairs. Some students presented an extended conversation in front of the class, although not spontaneous. There were some real communication and negotiation of meaning between the teacher and the students.

Phuong's comparison of the General Business English courses before and after TOEIC revealed several changes in the content of her teaching and her use of some teaching and learning activities. Only two of them were due to TOEIC. The other changes were caused by the new textbook, students' needs, and her teaching experience. Phuong asserted that her teaching method was unchanged because her beliefs about language teaching and learning always remained the same.

However, when asked to compare her present courses with the ESP courses before TOEIC, Phuong admitted big differences between the two. She said that the ESP courses only focused on the reading skill because the purpose was to help students read their specialized materials. The textbooks were, therefore, only made up of reading passages, which were followed by comprehension questions and vocabulary exercises. Instructional time was reserved exclusively for reading, grammar, and vocabulary. There were absolutely no listening and speaking: "Only occasionally did I require students to speak to make it less boring." Her content focus seemed to contradict her beliefs about the necessity of teaching all four language skills, especially listening and speaking, and her perceptions of students' needs. She attributed this content focus to the textbook: "I did what the textbook required."

Concerning her teaching methods, Phuong herself was surprised to realize that they were different in the two courses, too. As for grammar, she explained it more carefully in ESP

courses “because I focused on those things only.” Now she only reviewed the grammar knowledge that students had learned and organized it in a system for them. When teaching conditional sentences, for example, she did not explain each type carefully as in the past.

Teaching vocabulary was also reported to be different. In ESP courses, in addition to teaching guessing skills like in the General Business English courses, she explained words, had students repeat them after her, and told them how to use the words in sentences. Besides, she had students do more vocabulary exercises, which she prepared herself. For instance, students had to look for words in the reading texts that matched the given definitions or match given words with their definitions. However, she said she only dealt with vocabulary quickly in the current General Business English courses. Instead of explaining vocabulary herself, she required students to look it up in the dictionary at home. In class, she focused more on language use. She had students use the new words and checked if they used them correctly. Phuong also agreed that the differences in her teaching were due to the textbooks.

Explaining the differences between the current General Business English and the ESP courses in the way she taught writing, Phuong said:

The differences, if there were any, were due to the textbooks. (...) for example, this textbook requires students to write letters, so I teach them how to write letters. But the other textbooks..., for instance, the ESP textbooks that I talked about, covered conditional sentences (...), so I asked students to write those types of sentences. I only had them write sentences about given situations; I didn't require them to write a whole passage (...). That depends on the textbook. Teachers stick to textbooks to make it [teaching] conform to the requirements of the later exams.

Her last sentence revealed an important reason why teachers in Vietnam were so strongly controlled by textbooks: course assessment is usually closely based on the textbook content. This highlighted that exams had never stopped being important for teachers.

Case summary

Phuong was an older teacher at University A. She seemed to be very little influenced by the TOEIC test. Obviously, her teaching was shaped by an interaction of numerous factors including her beliefs about language teaching and learning, teaching experience, experience as a language learner, and personality (dedication); students' ability, needs and learning background; textbooks; and limited instructional time. For example, as she believed that the purpose of language teaching was for students to be able to communicate,

she gave students a lot of opportunities for language use practice. However, although she considered all four skills important components of communicative competence, she paid more attention to speaking and listening due to students' needs, their low speaking and listening ability, the textbook, and the TOEIC test in the case of listening. Also, because of the limited instructional time, she asked students to prepare for the reading section at home and learn grammar and vocabulary by themselves, so that they had more time for speaking practice in class. Thus, her decision on what to teach more resulted from an interaction of her own beliefs, students' needs and abilities, the textbook, the TOEIC test, and the time limitation.

The use of the communicative approach in Phuong's teaching of the General Business English Courses was clearly evident from the observation and interview data and was certainly a result of her beliefs about language teaching and learning and her effort to meet students' needs. However, the comparison of her teaching in ESP courses and current General Business English courses disclosed the most important factor in her teaching – textbooks. Textbooks influenced not only her teaching content, but also her teaching methods. They even overrode her teaching beliefs at times. Thus, although Phuong said that her beliefs about language teaching had never changed, she admitted that she taught the ESP courses and current General Business English courses very differently. She herself attributed these differences to the differences in the textbooks. Therefore, the minimal extent of the impact of the TOEIC test on Phuong's teaching was likely to be attributed to the overwhelming influence of the textbook. This strong influence of textbooks on her teaching, in its turn, could be explained by the common practice of teaching to mid-term and final tests.

7.3.2 Case 2: Lan.

Personal and educational background

Lan was in her early 40s and had a lot of teaching experience. She held a Postgraduate Diploma and a Master's degree in TESOL. After graduating from university, she became an English teacher at University B, and, like most teachers in Vietnam, she worked as an adjunct teacher at different universities in the city at the same time. She seemed to be very busy, so she preferred one long interview to two or three shorter ones. Before the introduction of the TOEIC test to University B, she taught both General English and ESP courses. After TOEIC became the language exit requirement, she continued to teach

General English courses (English 1, 2, and 3) and started teaching TOEIC preparation courses (English 4). English 1, 2, and 3 used an integrated skill textbook called Cutting Edge, and English 4 employed a TOEIC preparation textbook called Longman Preparation for The New TOEIC Test. Besides teaching TOEIC for the university's regular programmes, she also taught it in evening classes at the university's foreign language centre. When I carried out my research with her, I observed one of her English 4 classes.

In Lan's opinion, the teaching job was a little boring because teachers taught the same things over and over again, and textbooks were changed only after several years. Therefore, she said she designed her lessons differently depending on the types of students. She added something new to every course to make herself feel less bored. In fact, among the teachers I observed, she was the one who used extra materials the most. During my three observations, she showed students videos for fun twice. She said she had attempted to have students in English 4 work in groups and role-play the scenarios in videos she showed them, but her attempts were not successful. I also noticed that she picked up the grammar points often tested in the TOEIC to teach them separately by topic instead of following too closely to every textbook section.

Beliefs about language teaching and learning

Lan believed that the purpose of language teaching and learning was for students to communicate and better integrate into society. She considered communicative ability as the ability to do verbal communication, i.e., to understand spoken language and respond to it quickly. In her opinion, to achieve this purpose, students needed to have an environment where they could have a lot of contact with the language. For that reason, she often advised them to take part in English clubs and take a communicative English course at foreign language centres before a job interview even if their TOEIC test scores were high. She also held that it was necessary to increase instructional time for the English subject, but she knew this was impossible.

Lan was very exam-oriented. She considered her career goal "helping students to achieve the best exam results." She repeatedly related what she did with that goal: "My conscience always says: 'They have to pass!', so I try to teach differently from someone who thinks: 'It doesn't matter how I teach as long as I finish all the hours.'"

However, she perceived her main teaching task as preparing students for their mid-term and final tests, rather than the TOEIC test because the latter was still too distant.

Within 14 weeks, teachers have to strive for students' best mid-term and final test results first. (...) Now they are only in their second year. Usually, only in their fourth year will they prepare for the test. A test taken in two years' time is too distant to prepare for. So the immediate objective is students' passing their [mid-term and final] tests because if they don't pass these tests, they'll have to retake the course. If they retake a course, the possibility of failing is higher than the first time. And if students fail a course many times, they won't be able to get their degree.

It is necessary to mention that both the mid-term and final tests had TOEIC format, but they were easier than the international TOEIC test. The mid-term tests included two subtests, each accounting for 20% of the final grades. The first one focused on grammar, structures, text completion, and reading comprehension (single passages only). The second one was a listening test. The final test accounted for 60% of the term test scores and included the double passage tasks as well.

Lan seemed to have a practical mind. Although she believed that the purpose of language teaching was communicative ability, she thought that technical students did not need to learn all four language skills because they primarily needed to be able to read technical materials, and they had no environment to practice the language. Therefore, the reading skills and vocabulary knowledge are more important.

Of course, it's important for language majors [to learn all four language skills], but that's not necessary for technical students because they mainly learn to read reference books. So reading is the most needed. They mainly have to be able to read quickly, read for main ideas, and have a wide vocabulary. Also, if they want to communicate well, they need to have an environment. If they learn a language without language use, then they'll forget it.

Later, when I questioned the contradiction between her beliefs about the general purpose of language learning and her idea that technical students did not need to learn all four language skills, she cited the time limitation as another reason for her opinions.

Yes, that [communicative ability] is the purpose, but it's impossible to teach technical students all four skills. There is not enough time to focus on the four skills. So after graduation, most students have to take [communicative English] courses to pass job interviews. (...) You see that students learn only four periods per week. It's impossible.

Lan was also practical in her beliefs about test-preparation. She considered it important for students to solve sample tests and learn test-taking skills in order to improve their test

scores. She agreed that test-taking tips were very important and they could help increase test scores by 30%.

Practice a lot. Practice by doing sample tests and learn to manage your time. Because time management can help gain a lot more points. (...) And you need to know skills, for example, when listening, you should be able to recognize key words. The same is for the reading section. You don't need to read every word and translate every word. That would take too much time.

On the contrary, she believed that the theory about language teaching that she had learned was not applicable for test preparation.

[Theory] is not of much use. (...) The university also occasionally organizes workshops, but if we entirely follow those techniques that the instructors use, students can't get high scores. (...) If we use games or stick words [on the board] in the way we were trained to teach or have students run around and pick up words..., how can we have enough time to teach a reading text like that!

Perceptions of the TOEIC test

According to Lan, the TOEIC test had the purpose of evaluating test-takers' language ability, especially listening, reading, and grammar, but it could not assess their overall communicative ability:

The purpose of the test is to assess students' language ability, but, in my opinion, (...) it can't reflect everything like TOEFL because communication, that is, speaking, isn't included. Even if students get, for example, 900 or more than 800 for the TOEIC test, there is no guarantee that they can speak. That is to say, they're short of that part a lot. (...) It can assess students' ability to use English, but still can't assess communicative ability.

She added that a TOEIC certificate was not enough to get a job because students needed to pass employers' interviews. Despite her concern for the validity of the test, Lan agreed that the use of TOEIC as an exit test was necessary and useful. First, it could help evaluate students' language ability after several years at university. Second, it helped motivate students' learning: "Generally, when it's compulsory for students to have the certificate, then they will try to learn."

She also thought that TOEIC was suitable for graduation purposes because "if IELTS or TOEFL were used, they would be too difficult for students." However, she preferred a test that included speaking as well.

Lan also affirmed that teachers did not have to be responsible for students' test scores, so she did not feel the pressure either from the university or from students: "We aren't held

accountable for students' test scores like at school. Students mostly have to study themselves.”

However, in her opinion, all teachers had a conscience, so they always wanted their students to get good marks. For three weeks before a mid-term or final test, she often reviewed everything for students and solved sample tests to familiarize them with the test types, so that they could get good results. Again, she made it clear that she mostly aimed at the mid-term and final test: “But later, if they wanted to get an international certificate, they would have to make more effort.”

Impact of TOEIC on teaching

When asked about the impact of the TOEIC test on her teaching, Lan asserted that she had to teach to the test, i.e., teach the tested content, practice sample tests, give students an overview of the test structure, and teach time management strategies. She said:

For students to achieve [good] test results, I have to teach to the test. (...) For example, I solve a lot of [sample] tests. I have students do exercises on frequently tested areas. (...) [I tell them] that the test consists of 200 questions, (...) how many sections there are, (...), what is included in each section, how many minutes they should spend on each section, and how they should time their work at home, so that they will achieve the result. I have to be oriented towards the test!

In order to see the impact of the TOEIC test on her teaching more clearly, I asked Lan to compare her teaching in English 4 with that in English 1, 2, and 3. She said she taught the two types of courses differently. In English 1, 2, and 3, she covered a wider range of skills and teaching and learning activities: “In English 1, 2, and 3, I focus on basic grammar, basic vocabulary, and I teach listening and have students communicate. For example, I show them video clips, give them vocabulary, and then let them speak in groups.”

In English 4, the main focus was only on grammar and structures and test-taking strategies for the reading section: “Mainly grammar and structures. For reading, I teach them about the traps that they are often caught in in the double passages.”

However, she said that she did not teach those time management skills and test-taking tips which were specifically used for each listening part of the TOEIC test. Her explanation was that she did not have enough time while those skills were not needed for the mid-term and final test.

But I can't teach time management skills and tips for each listening part because I don't have enough time. Also, students don't need them because only two or three years after this course will they take the [international TOEIC] test, so they won't remember those things. Plus, the mid-term and finals tests mainly focus on grammar, structures and reading only. So my teaching is to serve those tests.

It turned out that the TOEIC-as-exit-test was not the only test that affected Lan's teaching. In fact, her responses showed that the influence of the mid-term and final tests was even stronger than that of the TOEIC-as-exit-test because they were more immediate. Clearly, the two assessments, exit language assessment and course assessment, had the same nature (content and format), so their impact on Lan's teaching blended with each other. However, the differences between them in the degree of difficulty and weight given to specific test components allowed us to see which source the influence on her self-reported teaching came from.

Thus, compared to English 1, 2, and 3, Lan's time allocation to the four language skills was perceived to have changed a lot. She now spent much more time on grammar, reading and vocabulary. Time for listening also increased, but only moderately. She did not teach speaking because "this [English 4] did not have speaking at all, and there is not enough time for students to do practice tests." She did not teach writing, either "because tests don't have writing." But she added that there was not writing in English 1, 2, and 3, either. Time for pronunciation remained the same. Nevertheless, when asked what she would like to teach if she had time and was allowed to teach whatever she wanted and whatever she considered important, she said, "speaking and pronunciation."

Besides the textbook, Lan said she used a lot of supplementary materials, which were sample tests or practice tests taken from other TOEIC preparation books. Explaining why she had to use extra materials, again, she attributed what she did to the final tests: "The university's textbook, Introductory, is too easy. It's easy, but (...) the final test is very difficult. If the final test was easy, too, then I wouldn't need to do that. So I gave them extra materials."

Lan also affirmed that there were differences between English 1, 2, and 3 and English 4 in her teaching methods. For *listening*, she associated the ways she taught with task types.

Different! For example, the ways they listen are different. [In English 1, 2, and 3] They listen to a passage, and then answer questions or fill in the gaps, that is, listen and play at the same time. It's different. But here the test has TOEIC format, so there are part 1, part 2, part 3, and part 4 task types. So the ways they listen differ.

Her responses showed that, generally, the steps in the teaching procedure remained the same and used for the same purposes: pre-listening, listening, and post-listening. The pre-listening activities always seemed to aim at accommodating students' understanding, dealing with vocabulary and content. However, the specific activities used were sometimes different due to the types of listening task. For example, in English 1, 2, and 3, Lan always taught students new words and had them read given questions or passages.

First, I give them new words, some new words. Second, I give them guided questions. Then they listen. If questions are already given, I ask them to read them first. If they have to fill in the gaps with given words, they have to read the passage first. Then they listen.

In English 4, she used other pre-listening activities depending on the task type in each part of the test.

For example, in Part 1, Photographs, I tell students what to look at in the photos or what words they should expect with such photos (...). I ask them to describe the photos first. For the Question and Response Part, I have to teach students types of questions, but usually I don't have time to teach all of them. If you hear this type of question, you have to select this or that word. For Part 3, I ask them to read the questions first and then answer. I also give them some new words found in the options, so that they can imagine [the content], and then they listen.

The perceived impact of TOEIC on Lan's teaching of *reading* was much clearer than on her teaching of listening. As she admitted herself, "reading is completely different."

As for reading, I don't give them new words. I have them read and, usually, time their reading. (...) But for the others [English 1, 2, and 3], reading is different. If the text is about travelling, for example, I ask questions before reading, for example, "Do you like travelling?" "What do you like doing in your free time?" I mean I ask questions around the topic first. (...) Next, I give them some words related to the reading text. Then they do the task. (...) But for this one [TOEIC], I don't have time to do that. I only time them. And at this stage, they must know the vocabulary. If they don't know it, I explain it after they finish reading.

As can be seen from the quote above, in English 4, Lan tried to familiarize her students with the real testing situation by skipping pre-reading activities, not explaining new words before reading, and timing students' work. Another reason for the difference, as Lan said above, was the limited time in the TOEIC preparation class.

Lan also revealed that due to the nature of the courses, different types of *grammar* were taught in English 1, 2, and 3 and English 4. In English 1, 2, and 3, the grammar was easy, but in English 4, the TOEIC grammar was difficult. Also, in English 1, 2, and 3, there were a larger variety of activities including language use one, while in English 4, there were

only formula presentation and form-focused activities. Here is Lan's description of how she taught conditional sentences in the two types of courses.

Differently! When I teach conditional sentences in English 1, 2, and 3, I give students examples or... group work. For example, for the second conditional, the first student says, "If I had time, I would travel." The second student says, "If I travelled, I would go to..." The third one says, "If I went to France, I would buy..." I mean they make a series of sentences, so that they can memorize the formula. But in English 4, I can't do that. In English 4, I have to present the formula and give them exercises to do right away. But, usually, I teach them inversion, too. I tell them how to use the inversion of *if* and ask them to pay attention to it because it is often tested. (...) It's totally different.

In English 1, 2, and 3, she also let students use the new grammar in both form-focused and language use activities after she presented it: "[They] practice speaking. After that, they do exercises. But their exercises are also very easy. They do the exercises in the textbook."

According to Lan, teaching *vocabulary* in the two types of courses was also "totally different." The first difference was that vocabulary in English 1, 2, and 3 was easy and taught by topic, while the vocabulary in English 4 was not:

In the other [English 1, 2, and 3], I mainly teach the vocabulary in the textbook. For example, if the unit is about travelling, then the vocabulary is related to the unit and the topic. But here it's not like that! (...) Here I try to teach idioms and expressions often used in TOEIC. (...) they don't belong to any topics. Very difficult.

The second difference was in the types of tasks and teaching techniques used. In English 1, 2, and 3, Lan taught vocabulary by explaining meaning or giving synonyms. She helped students remember new words in class through textbook exercises like gap-filling and true-false and through speaking and writing practice (although writing only involved single sentences). By contrast, in English 4, there was not such a variety of practice. Instead, Lan only had students do test type tasks, usually gap filling, asked them to highlight the new words, told them to memorize the words at home, and checked if they knew them in the next class session by asking them for the meanings of the words. Lan said, "Teaching [vocabulary] in English 1, 2, and 3 is more enjoyable because it's more varied than in this one."

As for *pronunciation*, Lan said that in English 1, 2, and 3, she taught all the difficult words, but in English 4, she could only teach the extremely difficult words. She explained, "If I taught every word like that, I wouldn't have enough time."

There was a wider variety of *teaching and learning activities* in English 1, 2, and 3, too. Lan said,

In English 1, 2, and 3, students are much more dynamic. For example, they stand up and speak, answer questions, go to the board and write, and so on. But in this class, we can't carry out those activities because the time is limited.

Time allocation to language teaching and learning activities were also changed. Lan said she spent more time explaining grammar rules in TOEIC class and taught them more deeply. She also said that she talked a lot more and used less English in class. Communicative activities such as survey, information gap, and group discussion, were completely absent. Their absence was attributed to several factors: class size, classroom arrangement, and especially time limitation which, in turn, was due to the broad and more advanced content of the final test. When explaining why she did not use information gap activities, Lan said:

Information gap activities are indeed used more in English 1, 2, and 3 because there is more time (...). Besides, the test pressure [in English 4] is greater, so I have to pay more attention to the test. But in the other courses [English 1, 2, and 3], what is learned is tested. Very easy. I only cover certain things: "Ah, this term we learn units 1, 2, and 3, so you are tested on those units only." But this one [TOEIC] covers everything! I'm also pressured by that. I have to try to finish all the lessons, the whole textbooks, and have students do lots of exercises so that they will get good results.

Thus, from Lan's report, the differences in teaching methods between English 1, 2, and 3 and English 4 were evident. In the former, Lan used a wider range of activities than in the latter, and she focused not only on forms, but also on language use. This variability was due to the differences in task types, the volume and degree of difficulty of knowledge to teach in General English courses and TOEIC preparation courses, and the necessity of accommodating students to testing situations. In other words, the nature of the TOEIC test (test design, test content) and nature of TOEIC courses (volume of knowledge, difficulty of knowledge, time limitation) were factors that reportedly influenced Lan's teaching methods.

Classroom observations confirmed what Lan said about her teaching in TOEIC preparation courses. Almost all class time was spent on TOEIC materials (95.02%). A very small percentage of time was used for videos and riddles "just for fun" (4.98%).

Concerning participant organization, there was no pair work and group work at all. All class time was spent on whole class work (74.74%) and individual work (25.26%). No

communicative activities such as information gaps, surveys, role-plays, or group discussions were observed, either. It is a little unusual that Lan used English in her class a lot more than most of other teachers who taught TOEIC preparation. However, she said that she used much more English in General English classes: “That’s not a lot! In English 1, 2, and 3, I use more English than that. In this class (...) I still use English for things students can understand, so that they can listen [to English]”.

She added that she mostly translated things into Vietnamese when she taught grammar and vocabulary. She cited the difficulty of vocabulary in the TOEIC test, the time limitation, and the large number of new words to teach as the reasons for using less English in English 4:

I use more [English] because the terms are not as difficult as here. In the TOEIC test, sentences or answer options are very difficult. (...) In the other [courses], I can give a simple definition, and students can understand. For example, to explain what addict means, I say, “An addict is a person who is interested in doing something,” and they will understand immediately. Right? But here if I give such a definition to every word, I won’t have enough time. Also, there are so many words!

The observations showed that the two types of teaching activities frequently happening in Lan’s class were teaching grammar and doing practice tests or sample tests. They were both very teacher-centred. The teacher talked most of the time, while students mostly listened and took notes. When teaching grammar, the teacher explained grammar rules and examples shown on slides. Students listened quietly, copied down what was shown on the slides, and then did exercises given in the teacher’s hand-outs. They rarely asked the teacher questions.

Solving practice tests often followed this general procedure: The teacher assigned questions for students to do within a time limit, she asked them to tell her their answers when the time ran out, and then she corrected them and explained the choices. Vocabulary and grammar were taught as they were found in the test. Texts were often translated into Vietnamese. Students only gave single word answers to the teacher’s questions like “A,” “B,” “C,” or “D,” or Vietnamese equivalents of English words. There was no extended, meaningful interaction between the teacher and students. Probably for this reason, Lan admitted that she did not know students’ names and their ability: “And with this learning/teaching style, how can I know their names? I don’t know their ability, either!” The following teaching excerpt can help illustrate these features of Lan’s teaching.

Materials: Longman Preparation Series for the New TOEIC test: Introductory Course, p. 270 (see Appendix L)

Task: Text completion task

Episode 1: Teacher assigned task to students.

T: All right. Test 2. *What part have we done?* Practice test 2. *Have we done page 263? All of the items?*

Ss: [unclear]

T: *Already. Page 268?*

Ss: [unclear]

T: 270. OK, 270, 272, 275. 270 to 275.

Episode 2: Students did the exercise individually for about 13 minutes.

Episode 3: Teacher checked students' answers; explained the reading text; and translated the text, questions, and answers into Vietnamese.

T: All right. Turn to page 270 first. "We are all aware that" [The teacher read the text]. To be aware of *or* to be aware that, nhận thức được rằng [translation], "the procedure for charging and reporting expenses for business trips taken on behalf of." All right. On behalf of. *Please highlight it.* Nhân danh, với tư cách là [translation]. *So những kỳ đi công tác* [translation], all right, "on behalf of the company have long been out of hand" - đã không kiểm soát được [translation]. *So, những chi phí để dùng cho cái mục đích đi công tác của công ty là không kiểm soát được* [translation]. "As a result _____." All right. As a result what?

Ss: [unclear]

T: A. 149, "as a result of recommendations from the Budget office, the following procedures will be adopted." All right. "Company personnel will no longer be given company credit cards to cover expenses while on out of town trips. Instead, all travel expenses, with the _____." All right. Follow "the" here is a noun. Which one number 150?

Ss: [unclear]

T: C. All right. "With the exception of airline ticket which will continue to be charged directly to the company, will be paid for out of pocket."

[Teacher continued this way until the last question of the passage]

T: *How many correct questions have you got? OK. Learn the highlighted words and phrases carefully.*

Note. Words and sentences in italics have been translated into English by the researcher.

As can be seen in the excerpt, the teacher controlled the whole activity. Students only gave the teacher one word answers such as A, B, and C when asked. Instruction was done both in English and Vietnamese. There was a lot of translation from English into Vietnamese. Vocabulary was taught only by giving the Vietnamese equivalents of English words.

Case summary

Lan was an older teacher from University B. She was a very test-oriented teacher, who considered helping students pass their tests to be extremely important. Therefore, the impact of the TOEIC test on her teaching content and teaching methods was powerful. This strong influence of the test was clear from both the interview and observation data. In English 4, she taught only what she perceived as needed for the test. The general teaching approach was also changed to make it more effective for testing purposes: increased teacher talking time, increased use of Vietnamese, increased translation, no more communicative activities such as information gap, survey, or group discussion. The ways she taught specific skills were also changed to meet the testing purposes. The clearest changes were found in teaching grammar, vocabulary, and reading. A smaller change was evident in teaching listening. The teaching of these skills had become more superficial and/or less varied. Teaching grammar and vocabulary, for example, aimed at forms rather than language use and employed a smaller range of techniques. Teaching reading aimed at accommodating students for the real testing situation with fewer activities and less interaction. Teaching approaches which were considered most effective for test preparation contradicted theory of language teaching in some ways. Compared to English 1, 2, and 3, her teaching methods were changed due to the nature of the TOEIC test and of the TOEIC preparation course. Because the same test was used for the language exit assessment and the course assessment, the two assessments co-produced effects on Lan teaching. However, Lan was more strongly influenced by the course assessment (mid-term and, especially, final tests), which was easier and gave different weights to test components, because it was more immediate than the TOEIC-as-exit-test.

7.3.3 Case 3: Phong.

Personal and educational background

Phong was a young teacher at the provincial university. He majored in Business English and graduated from university only two years before I met him. He had had less than two years of teaching experience at university, but as a second year student, he started working at various foreign language centres, where he taught both communicative English and TOEIC preparation courses. He said, generally, he liked teaching, but he sometimes felt dispirited by the challenges and troubles he encountered, for example, teaching naughty students or those who had to resit English exams. The unhappy feeling about teaching is

probably connected with his low expectations of his students: “Speaking about expectations, I don’t dare expect too much. Simply they are willing to learn when in class and do homework at home. I only require that much.”

As for his expectations of students’ results, he only hoped that after four years at university they could achieve a TOEIC score of 400 and communicate with foreigners about simple matters such as greetings and daily conversations.

Beliefs about language teaching and learning

Phong distinguished between the purposes of language teaching in general and at his own university: “I think at my own university (...), the purpose is a minimum of TOEIC 350 after students finish TOEIC 5. (...) But speaking about language teaching in general, the purpose is the ability to communicate well in English.”

Thus, he believed that the general purpose was the communicative competence, and his own university’s purpose was passing the TOEIC test. Communicative ability for non-English majors, in Phong’s opinion, was “simply making other people understand us when we speak English. It’s not necessary to have correct grammatical structures, but we need to use vocabulary somehow to express our ideas to other people.”

Phong considered the communicative approach the best teaching method, but not applicable to large classes. When asked about what should be done to help students achieve communicative ability, he said, “With large classes of 60 to 70 students, it’s impossible for me to help students communicate well.” Therefore, he said he only marked students’ mid-term speaking performance, tried to speak English a lot, and made his students use English a lot in class.

To help his students pass the TOEIC test, he said he attached special attention to test-taking strategies. He strongly believed that they are important to improve students’ test scores because as non-English major students, they couldn’t know all the grammar and vocabulary in the test. He said, “I give students test-taking tips, (...), because it’s impossible for them to know all the vocabulary and grammar in our university’s TOEIC test. So definitely, I have to teach them test-taking tips to improve their test scores.” He also thought that the effects of test-taking strategies were large:

[test-taking strategies] influence the results a lot. For example, some students who have a good grasp of knowledge, know a lot of vocabulary, know structures, but don't have test-taking skills, especially TOEIC test-taking skills, can't get high scores because they don't know how to allocate their time. (...) answering 200 questions in two hours exceeds students' ability.

Perceptions of TOEIC and TOEIC benchmark

Phong had mixed opinions about the TOEIC test. On the one hand, he did not think it could assess communicative ability:

Essentially, it's called a test for communication, but I don't think it can assess communicative ability because it doesn't include the speaking skill. (...) I don't think a person who can listen well and read well can speak English well because [speaking] is related to many other things, depends on grammar and pronunciation ability as well. (...) If a person who is cramming for the TOEIC test only practices listening and reading, I think that's not enough to communicate well with foreigners (Phong).

On the other hand, he believed it could accurately reflect learners' language proficiency because it was an international test, and only with good English could students meet the university's benchmark. These ideas seemed to conflict with each other at first, but probably what he meant was that the TOEIC test could accurately assess what it actually assessed, but not the overall communicative competence. He also held that the test was very useful for university students because it forced students to learn, and the large quantity of business vocabulary in TOEIC would benefit students in their future jobs.

He knew that the TOEIC test assessed the listening and reading skills, which included vocabulary and grammar knowledge: "The first section assesses the listening skill, students' listening skill. It's also a communicative skill, but it'd be better if the speaking skill were included, too. The reading section mainly measures students' vocabulary and grammar."

Phong did not feel stressed or pressured about preparing students for the TOEIC test because "our university doesn't hold teachers accountable. We mainly need to try our best with all our conscience." Nevertheless, he admitted improving students' test scores were very difficult, especially in large classes like his, where he could not control all of the students.

Test impact on teaching

When asked about the impact of the TOEIC test on his teaching, Phong asserted that the test certainly affected his teaching content and teaching methods. Compared to the communicative classes he had taught, in TOEIC classes, he spent much more time on listening, reading, and grammar while reducing the time for speaking, writing, and pronunciation a lot. Time for listening, for example, increased three or four times. He explicitly attributed these changes to the nature of the TOEIC test: “TOEIC has only two skills – listening and reading, so I only focus on these two skills, and I reduce and restrict the time for speaking and writing.”

Also, his time for grammar increased a lot more in TOEIC classes due to the nature of the course and the TOEIC test:

...in the communicative classes that I taught at language centres, there was very little grammar. Usually, even at the advanced level, the grammar was still simple. But in these TOEIC classes, even though the level is low, for example, TOEIC 1 and TOEIC 2, the grammar is very advanced. So I concentrate on grammar.

On the contrary, speaking accounted for the majority of class time in communicative classes. It was included even in reading and listening.

Phong’s responses showed that he not only focused nearly entirely on the tested skills, but he was also very keen on teaching test-taking strategies as he believed that they affected his students’ test scores a lot. In the following extract, he talked about the strategies he taught his students to help them pass the TOEIC test.

Non-English majors definitely can’t understand all the questions. So I direct them to listening for question words. Those are the most important words. And I give them some tips to avoid the traps in the TOEIC test. For example, if we hear a word that is included in the question, we can assume that it is a trap and should avoid it. Parts 3 and 4 are extremely difficult, especially for non-English majors. I don’t require students to understand all the content of the listening texts. They just need to listen for the main ideas, listen for the sounds that are similar to the answer options and select their answers. (...)

Concerning the reading section, the grammar..., I also teach students some tips to do the grammar, for example, the order of parts of speech in a sentence, so that students don’t need to look at the whole questions. Because students encounter a lot of new words in the test, they should guess the answers without looking at all the new words. They only need to guess.

It is clear from what Phong said that he aimed at helping students answer test questions correctly, often at the cost of understanding. In addition, he tried to familiarize his students with the testing situation by having them do practice tests in class and timing their work.

Some teachers only correct students' answers in class. But I make my students do practice tests exactly like... [in the testing situation], for example, spend five minutes on a passage, five questions – exactly five minutes, because in the TOEIC test, you have one minute for each question.

He reported that he used no extra materials in TOEIC 5 classes due to time limitation and high workload that third year students had: “In TOEIC 5, we don't have enough time to cover everything; students don't have time to do extra work. Also, in the third year, students have very heavy workload. Giving them extra materials would only burden them.”

Observation data supported what Phong reported. It showed that he spent 88.99% of his class time on the TOEIC content in the textbook, and 11.01% on the speaking topics. As described on page 157, each unit of the TOEIC 5 textbook was made up of a complete TOEIC practice test and a speaking topic with suggested questions. Clearly, the TOEIC content totally focused on the two sections of the TOEIC test – listening and reading. While teaching these two skills, Phong also dealt with the grammar, vocabulary, and a little pronunciation related to the practice tests done in class. Speaking was included as specified by the department, but, as Phong said, it was almost not taught:

In TOEIC 5, I almost don't teach speaking, that is, I asked students to prepare a presentation at home. At TOEIC 5, students certainly know how to answer the questions. If they don't know them, they can Google them on the Internet. Before the presentation day, some groups send me their presentations so that I can correct their grammar or add and remove ideas.

This is also what the other teacher who taught TOEIC 5 did. Obviously, this way of teaching speaking was agreed on in the department. Writing was not taught at all during the three observations, either.

The interviews revealed that the TOEIC test not only influenced the content of Phong's teaching, but also his teaching methods. There was a clear shift away from the communicative approach. He admitted in TOEIC classes he talked much more and used more Vietnamese. His use of Vietnamese in TOEIC classes was likely a result of the nature of the test or the test preparation course, which involved a lot of advanced grammar and vocabulary as he said elsewhere. He also did a little more translation in TOEIC classes.

This way of teaching was clearly evident in his observed lessons. What happened in Phong's TOEIC classes was that he assigned a part of a practice test for students to do within a time limit. After that, he checked students' answers, corrected them, and explained them. Vocabulary and grammar were taught as they were found in the test. Texts were always translated into Vietnamese. Students' work was mostly doing practice tests, giving single word answers to the teacher's questions like "A," "B," "C," or "D," or Vietnamese equivalents of English words. There was no pair work and group work at all. Instructions were done almost entirely in Vietnamese. Meaningful and extended communication in English between teachers and students or between students and students was almost absent. The following excerpt represents a typical piece of Phong's teaching.

Teaching materials: TOEIC 5, page 71 (see Appendix M)

Task: Sentence completion

Episode 1: Teacher assigned task

T: *Today [many in] our class are late. We can't listen now. Now take out your book and learn reading first. Turn your book to page 71. Page 71. Do questions 124 to 129, page 71. Finish page 71. Question 124 to 129, page 71.*

Episode 2: Students did the assignment individually for about 6 minutes.

Episode 3: Teacher checked students' answers; explained text, vocabulary, grammar, and answers, and taught test-taking strategies.

T: *OK. Let's correct the exercise. Question 123. This is a question of vocabulary type. Look at the four words skilled, earnest, tactful and reliable. Which ones do you know?*

Ss: [not clear]

T: *Skilled. One word. Any other words?*

Ss: [not clear]

T: *Reliable. OK, two words. I've told you before: For vocabulary items, you just need to look at the four words, right? Identify the words that you know, and then put them in the space to see if they fit in meaning. If they don't fit, select one of the rest at random. If they fit in meaning, choose them. OK? Among these four words, you know skilled and reliable. Try to put one of them in the space to see if it fits.*

S: *No*

T: *Why did you say "No" so fast? "InfoWorks technology has been dedicated" [Teacher read a sentence from the text]. Here there is a new word - dedicated. Underline it. What does Dedicated mean? OK. Write down the meaning: tận tâm, tận tụy [translation]. Cái công ty có tên là InfoWorks Technology lúc nào cũng tận tụy để offer - provide (cung cấp) những sản phẩm như thế nào đó với cái giá reasonable prices [translation]. You have "reasonable." Reason is...? Lý do [translation]. Reasonable - hợp lý [translation]. Cái giá hợp lý [translation]. We translate this as cái giá phải chăng. OK. What are the products like? Here you know two words. One is "skilled" meaning...? skilled. You have skill meaning kỹ năng. That's right. "Skilled" is có kỹ năng, là giỏi, có kỹ năng*

[translation]. *That's the first one. The second one is "reliable" meaning tin cậy* [translation]. *Try to put these two words here to see if they fit in meaning. First, skilled products. That doesn't sound right. Reliable products. OK?*

Ss: *Yes.*

T: *OK, Select that one. Now (()) then let's learn the other two words. First, earnest. Earnest. Write down the meaning: serious, nghiêm túc* [translation]. *This is an adjective that can be used to describe people. Serious is a personality. And tactful. Tactful is?*

Ss: [unclear]

T: *Tế nhị* [translation]. *That's right. A tactful person, that is avoid saying something that makes others upset. That's tế nhị or khéo xử. Tế nhị is khéo xử, khéo ăn nói. OK. Select D - Reliable.*

Note. The English words and sentences in italics have been translated from Vietnamese by the researcher.

The only exception from this kind of teaching was the pre-listening activities in Part 1 of the TOEIC test – Photographs, where test-takers have to look at a photograph, listen to three statements, and choose the one that best describes the photograph. Before students listened to the statements, Phong asked them display questions about what they saw in the photo using much more English than in the other activities. Only here did some meaningful, though still limited and not genuine, communication in English take place. The following excerpt from Phong's teaching help display these features.

Teaching materials: TOEIC 5, page 55 (see Appendix N)

Task: Photographs

Episode 1: Teacher assigned task and carried out pre-listening activity.

T: *Turn to the listening section, please. Page 55. Page 55. The other day, we finished question 6, right? Now let's continue to listen to questions 7 and 8. In the same way, before you listen, look at photos number 7 and 8, and tell me what you see.* [unclear] *This is a bookshelf, OK? Are there any books on the bookshelf?*

Ss: *Yes.*

T: *Yes, OK. There are many books on the bookshelf. What else? OK, please notice something on the bookshelf besides books. *Besides books, what else do you see on the bookshelf?* OK. Maybe some [unclear] two plants on the bookshelf. What else?*

Ss: [unclear]

T: *OK. Nothing else.*

Note. The English words and sentences in italics have been translated from Vietnamese by the researcher; the questions underlined were display questions.

Phong explained that he spoke more English for Photographs because the photos make it easier for students to understand what he said: "They can imagine, for example, what is in

the photo, imagine the words in their head, so when I say the words, they can recognize them. But in the other parts, they definitely don't know.” Phong's explanation revealed the fact that how much English he used depended on the difficulty or nature of the materials.

Phong reported that he changed the ways he taught language skills, too. In his opinion, the purpose of *teaching listening* differed between communicative classes and TOEIC classes. In the former, he wanted his students to understand 100% of the listening text and did not attach much importance to their being able to answer questions correctly. However, in the latter, students only needed to listen for main ideas, and their ability to answer questions correctly was vitally important. For that reason, Phong used different teaching procedures for the two types of classes. In communicative classes, he asked his students to listen and try to understand listening passages first and then answer comprehension questions. On the contrary, in TOEIC classes, he let his students read questions first, then listen, and answer the questions.

The reported changes in *teaching reading* were the same as in teaching listening and for the same reason. In communicative classes, he would use the comprehension questions to assess how much students understand, but in TOEIC classes, he only aimed at students' answering questions correctly, even without understanding. Therefore, he used opposite procedures for the two types of classes. This is his description of teaching reading in a TOEIC class.

Reading is the same as listening: The main thing is being able to answer questions, so I made students read the questions first. After that, it's not necessary for students to understand what the text is about or what the main ideas are, they only need to use key words to answer questions. The main purpose is to answer questions.

Phong later explained the reason why he did not place importance on students' understanding when teaching students how to do tests. According to him, the time allowed was too short (75 minutes for 100 questions), so it was impossible for students to completely understand passages to answer all questions.

The way Phong *taught speaking* was also said to be completely different. As said above, he almost did not teach speaking in TOEIC 5 classes. However, in communicative classes, he would teach students necessary grammatical structures and expressions in advance. Students would not prepare for the topic at home. Instead, he would bring up the topic in

class, and students would also practice speaking in class. This is how he described he would teach the topic of holidays in a communicative class:

For example, the topic Holidays (...). First, with a communicative class, I have to greet students; then I suggest ideas about the topic. For example, "Did you go out last Sunday?" "Did you go away in the summer?" (...) And then I bring up the topic: "Today we're going to talk about Vacations or Holidays." (...) Students will work in pairs and talk to each other. After that, pairs which are more confident will come to the front and talk in front of the class, and the other students will comment.

Although *writing* was rarely taught in both classes, some differences were reported. Phong said in TOEIC classes, he taught writing mostly in the reading sections. When there was a new structure, if time allowed, he would ask students to write sentences using the structure, and say them. However, in communicative classes, there were separate grammar sections, which were followed by writing activities. Writing could also be included in speaking topics: Students write a passage about a topic, for example, holidays, and then present the topic in front of the class.

Phong's responses to the interviews also showed that his *grammar teaching* in TOEIC classes differed a lot from that in communicative classes. He attributed the differences to the nature of the course: In communicative classes, one class session dealt with only one small grammar point, "but TOEIC is not like that because its grammar is very broad (...). In each session, there can be 10 different grammar points." Thus, in TOEIC classes, Phong only explained the rules quickly so that students could do tests.

For example, in communicative classes, one type of conditional sentence should be taught in one period, but in TOEIC preparation classes, all the five types of conditional sentences are learnt in only 15-20 minutes. [I] only introduce the structures for students to learn, and then show them, for instance, what tense this word [the word in the test question] should be used in, and in what type of conditional sentence.

However, in communicative classes, he would introduce a grammar point in a meaningful context and help his students discover the grammar rules by themselves.

If there is a projector (...), I can show some pictures illustrating [the meaning] of conditional sentences, second conditional, for example. I'll show a poor person on one side and a rich guy with a lot of money and a big house on the other side. I'll make sentences, for example, *If I were rich, I would buy a big house*. Then I'll let the students identify the structure of the sentences. (...) After students suggest the structure, I'll correct the suggested structure (...).

After that he would have them practice the grammar they were learning in a series of exercises from easy to difficult, from form-focused to communicative.

The exercises can be from easy to difficult. For example, at first I only ask students to match parts of sentences [...]. More difficult exercises would be asking students to put the verbs into spaces in the correct form. For more difficult exercises, I let students make up their own sentences.

Phong also revealed another reason for more communicative activities in communicative classes: the textbook. He said, “And usually such grammar points in a communicative book are very well designed, that is, if they are taught in the grammar section, they are covered in the speaking activities, too. They are included in the reading text, too.”

Thus, compared to communicative classes, there was a shift in grammar teaching from form-focused and communicative practice to no practice, from students’ actively discovering knowledge to students’ passively receiving knowledge.

From Phong’s responses, *vocabulary teaching* was also different for the same reason. If in communicative classes each lesson dealt with one topic, then in TOEIC classes, “vocabulary is huge, immense.” Also, TOEIC required a large capacity of vocabulary, but its requirement involved recognizing and understanding, rather than using. Therefore, in TOEIC classes, he only explained the meaning (using synonyms if possible; otherwise, using Vietnamese equivalents) and, at the most, gave the word families, so that students could recognize them if they encountered them later. In communicative classes, he first gave the word’s meaning in English, and then in Vietnamese if necessary. Then he gave students examples and asked them to apply the word in their own examples.

Because expanding vocabulary in TOEIC is very important. You must have a large capacity of vocabulary in order to do tests. Same for listening and reading sections. For TOEIC, the more vocabulary the better. But in the other [communicative classes], when you learn vocabulary, you must have a good grasp of it, and you must know how to use it, too.

The interviews also revealed that the way Phong *taught pronunciation* was partly different. The reason was that “in TOEIC classes, there is very little pronunciation. Students don’t need to pronounce words because there is no speaking test. We don’t need to teach pronunciation. I only teach difficult words with unusual pronunciation.” In communicative classes, he would let students listen to new words in a sentence or a passage first. After that, he would write the phonetic transcription of the words on the board and ask students

to repeat them after him. As teaching pronunciation was associated with teaching vocabulary, he would ask them to make sentences with the new words.

Case summary

Phong was a young teacher from University C. From both the interview and observation data, clear signs of teaching to the test can be seen in his teaching. He spent much more time on the tested skills while ignoring the non-tested skills. In addition, he put a great emphasis on teaching test-taking strategies. The general teaching approach was also changed to make it more effective for testing purposes: increased teacher's talking time, increased use of Vietnamese, increased translation, no more communicative activities such as information gap or survey. The ways he taught specific skills were also changed to meet the testing purposes. The clearest changes were found in teaching grammar, vocabulary, reading, listening, and speaking. The teaching had become more superficial. Teaching grammar and vocabulary, for example, aimed at knowing and recognizing rather than using. Teaching listening and reading had the purpose of answering questions correctly and in the shortest time rather than true understanding. Certainly, a number of factors were reported to influence Phong's teaching, for example, the large class size and students' ability. However, the nature of the test and the nature of the course were clearly the most influential factors. These so strongly governed Phong's teaching that they even predominated over his beliefs about communicative language teaching. Obviously, the ultimate reason that caused him to closely follow the textbook and teach to the test was his perception of his university teaching purposes, which became his own teaching goals – helping students to achieve the TOEIC benchmark.

7.4 Chapter Summary and Discussion

This chapter reported the findings from the teacher case study. It began with a summary of the results of the general analysis of all the teacher cases. The summary showed that the teachers were very similar in the factors potentially involved in shaping the test impact, namely their beliefs about language teaching and learning, their perceptions of the test, and their perceptions of the context factors. Despite those similarities, they were affected differently by the TOEIC test. Striking differences were found in the test impact between Business English and TOEIC preparation courses.

The stories of the three teachers, Phuong, Lan, and Phong, each representing one of the universities, more vividly illustrated and helped explain the patterns summarized in Section 7.2. Phuong's teaching mostly focused on non-TOEIC content and was characterized by features of the communicative approach such as the use of pair/group work and communicative activities, the use of English, and the interaction between teachers and students or students and students. By contrast, Lan's and Phong's teaching mostly covered TOEIC content. Their teaching was non-communicative, very teacher-centred, and superficial with a lot of teacher's explanations and translations, a much heavier use of Vietnamese compared to their non-TOEIC classes, and an absence of pair/group work and communicative activities. Their non-communicative and superficial teaching approach was explicitly attributed to the test design, especially the multiple-choice format, or the nature of the TOEIC preparation courses, e.g., large amount of advanced vocabulary and grammar to teach. The teachers' main focus was to help students answer multiple-choice test questions effectively, which led to targeting shallow understanding rather than language use. The purpose of answering test questions correctly was at times achieved at the cost of understanding. This again confirmed Smith's (1991) findings that multiple-choice tests caused multiple-choice teaching. Also, although the teacher questionnaire findings suggested that age was negatively correlated with teaching to the test, the case study showed that this was not the case. Lan, one of the oldest teachers and Phong, one of the youngest teachers, essentially taught in the same way despite some differences in their teaching styles, e.g., the extent to which English was used and the way of teaching grammar. Thus, the case study interestingly showed that in the current context, school locations were not as important as the nature of the course.

It was now clear that the impact of the TOEIC test on the two types of courses differed. However, the questions remained as to why the same language exit test affected them to different extents, and why in an exam-driven culture like Vietnam, this high-stakes test had only minimal effects on the teachers from University A. The three teachers' stories probably helped answer this question. They showed that the teachers' teaching was determined by a number of factors, e.g., beliefs about language teaching and learning, teaching experience, experience as a language learner, testing, availability of time, class sizes, students' ability, students' needs, and nature of courses, particularly textbooks. Among them, examinations were still a dominant factor. All three teachers were very concerned about tests and examinations, whose nature was most clearly connected with

their teaching content and teaching methods. Nevertheless, it is interesting that among the tests and exams students had to take, the mid-term and, especially, the final tests, were likely to most strongly determine teachers' teaching. The teachers often perceived their teaching tasks as helping students pass the course assessment, i.e., mid-term and final tests, rather than the TOEIC-as-exit-test. The reasons were that the latter was more distant than the former and had no consequences for teachers. Teachers at all the three universities were not held accountable for their students' TOEIC test results. The clearest evidence for the stronger effects of the course assessment was seen in Lan's case. She repeatedly emphasized that she mainly aimed at the mid-term and final test. Another important factor was textbooks. This was especially visible in Phuong's case as analysed above. Phuong did not see many differences in the ways she taught General Business English courses before and after the introduction of the TOEIC test because the textbooks were similar in nature. However, there were big differences between the current General Business English and the former ESP courses. She explained these differences by attributing her teaching to the textbooks. However, the effects of course assessment and textbooks turned out to be interrelated because, as Phuong confirmed, teachers stuck to the textbooks to better prepare students for mid-term and final tests. In other words, examinations had always remained a key factor in teachers' teaching, but in this study, the course assessment probably mattered more than the TOEIC-as-exit-test. When the TOEIC-as-exit-test matched the course assessment, as in the case of University C, the effects of the two possibly blended together. When they were different in nature, the course assessment appeared to have stronger effects.

Another interesting finding was that examinations and textbooks were so influential that they may have overridden some of teachers' beliefs about language teaching and learning. Although the teachers at Universities B and C also considered students' communicative ability as the goal of language teaching at university, and many of them believed that students should learn all four language skills, they only focused on the two tested skills, and their teaching was very far from being communicative. Phuong's teaching in the former ESP course at University A also reflected this fact.

In brief, the findings suggested that the impact of the TOEIC-as-exit-test on teachers' teaching was strongly moderated by textbooks and course assessment, which appeared to be interrelated because the former was often based on the latter. Other factors such as

teachers' beliefs, their teaching styles, their teaching experience, students' needs, and students' ability were likely to be only secondary players. When the course assessment and the language exit test were identical, their effects tended to blend together. Figure 6 illustrates the relationship among the TOEIC-as-exit-test, course assessment, textbooks, and teaching.

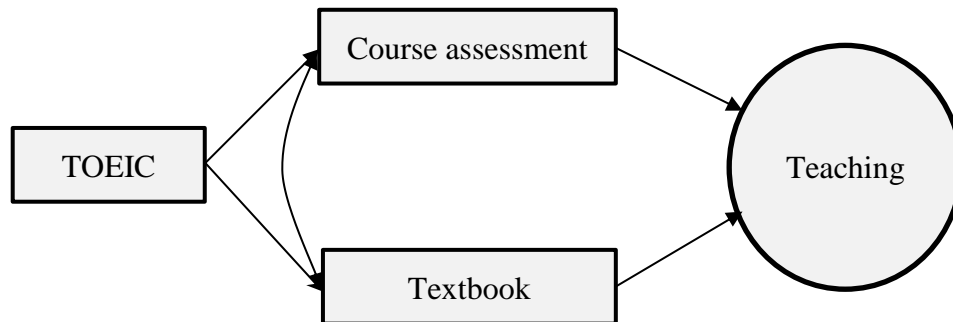


Figure 6 Relationship among the TOEIC-as-exit-test, Course Assessment, Textbooks, and Teaching

Chapter 8: STUDENT CASE STUDY

This chapter describes the results of the student case analysis. The case study of students was used to seek their perspectives regarding the following questions:

1. What is the impact of TOEIC-as-exit-test on university English learning?
2. What brings about the test impact on university English learning?

The chapter includes two parts: a brief report of the results of the quantitative analysis of the data and a qualitative presentation of three selected cases. The first part begins with a description of the students' perceptions of the factors potentially involved in forming test impact on learning, and then moves on to depict the actual impact of the test and offer an explanation of the way it occurs. The second part presents three students' stories in detail.

8.1 Demographic Information

A total of twelve students participated in the case study. Four students were recruited from each university. Eight of the students came from urban areas. The other four students came from rural areas, and their parents were all farmers. Most of them were third year students.

Based on their teachers' classification, there were five top students and five bottom students participating in the study; there was no information for two of the students from University B. The students also reported their achievement test scores including midterm and final term test scores, but they often did not remember them exactly. Some students only reported their GPA or the range of the scores. Generally, the students' test scores corresponded to their teachers' evaluation, but there were some mismatches in University C (Nga's and Huynh's cases). Nine out of the twelve students had taken the TOEIC mock test voluntarily or as a class test at different points in time. The scores ranged from 300 to 750. Hoang, Thien, Ngoc, three top students, had the highest scores. The students were also asked to evaluate themselves. However, the top students often underestimated their own ability with a rating at level five to six on a scale from 1 (*know very little English*) to 10 (*very good*). Only Hoang, who was considered by the researcher as among the very best, said that his language ability was a little above average. On the other hand, the two bottom students at the provincial university also evaluated themselves as average.

Table 8.1 Demographic Information of Student Cases (N= 12)

Name	University	Gender	High school location	Year at university	Proficiency group judged by teachers	Students' achievement test scores	Highest TOEIC mock test scores	Student's self-evaluation
Hoang	University A	Male	Urban	Third	Top	8.8	750 (1 yr ago)	5 - 6
Mai		Female	Urban	Third	Bottom	5	Nearly 300 (about 1 yr ago)	Below 5
Cuc		Female	Rural	Third	Top	6 to 7.5	300 (1 or 2 yrs ago)	Below 5
Khoi		Male	Rural	Third	Bottom	3< and <5	N/a	Below 5
Huyen	University B	Female	Urban	Fourth	N/a	7 to 8	380 (recent)	5 - 6
Toan		Male	Rural	Third	N/a	6< and <7	505 (recent)	5 - 6
Thien		Male	Urban	Second	Top	N/a	690 (recent)	Above 6
Thanh		Male	Rural	Third	Bottom	5 to 6	N/a	Below 5
Nga	University C	Female	Urban	Third	Top	5 to 6	N/a	5 - 6
Huynh		Male	Urban	Third	Bottom	5 to 8	400 (3 yrs ago)	5 - 6
Ngoc		Female	Urban	Third	Top	8 to 9	500< and <600 (a long time ago)	5 - 6
Luong		Male	Urban	Third	Bottom	6.0< and <7	300< and <400	5 - 6

Note. N/a = Not applicable

8.2 A Quantitative Analysis of Student Cases

The interviews with students and their weekly journal sheets were coded and analysed quantitatively. The interview analysis included factors potentially involved in student learning (students' English learning background and their beliefs about language learning, their perceptions of the TOEIC test, and the context factors) and the impact of the TOEIC test on learning (motivation, learning content, learning activities, and achievement). The weekly journal sheet analysis was done regarding the time the students spent on language skills, their learning activities, their use of learning materials, and their achievement.

The results of the interview analysis showed that although all of the students had the same seven year English programme at school, the majority of them had a very limited English background. The common features of high school learning were focuses on grammar and vocabulary, and the use of translation as a means of instruction. Also, listening, speaking and writing were often ignored.

Communication was shown to be a very important theme in students' beliefs about language learning. The majority of them believed that the purpose of language learning was to be able to communicate. This communicative purpose was strongly associated with future employment. Almost all of the students wanted to have communicative ability in order to get good jobs after graduation. It is interesting that the majority of the students understood communicative ability as the ability to communicate verbally only. Therefore, most of them considered speaking and listening, especially listening, the most important skills to learn at university. In addition, generally, students held that the best way to learn a language was to use it for genuine communication, especially communication with foreigners. In order to improve TOEIC test scores, however, many students thought it was important to practice listening a lot, and all of them acknowledged the importance of learning test-taking strategies. Thus, the best way to learn a language and the best way to improve test scores did not quite overlap.

The students knew the test design pretty well. All of them said that the TOEIC test assessed listening and reading. About half of the students mentioned that it also included grammar and vocabulary. Only one student said the test assessed speaking and writing as well. Nearly all of the students knew the test had a multiple choice format, except for one student who thought his university test included a speaking topic as well. The students

found the test valuable in some ways, but not the others. They somewhat agreed that it was useful because it motivated students to learn, and helped them improve their language ability, especially listening and reading, and as a result, it would also be useful in future employment. However, many students were aware that having a TOEIC certificate was not enough to get a good job; instead, they needed real communicative ability. Most of the students did not think the test could accurately reflect learners' language ability, either, because it did not assess overall communicative ability or because test scores were inflated by test-taking strategies and random guessing. In fact, even those students who thought it was relatively accurate had the same concerns as those who believed it was not. More than half the students also somewhat agreed that the test was suitable for graduation purposes, but many of them were concerned about its accuracy (or validity as termed by language testers) and its unsuitability for many majors. The TOEIC benchmarks seemed to be difficult for only less than half of the students, but the majority of them found the listening section more difficult than the reading one. All of the students were aware of the serious consequences of failing the test, and most of them admitted that the most important factor that motivated them to learn to the test was to be able to graduate from university. However, they did not experience pressure from teachers or parents. Some students experienced some sort of pressure from peers into test preparation. The students generally had favourable conditions for learning. Most of them had enough time and materials for learning.

The students' responses to the interviews showed that the TOEIC test obviously impacted on most students' learning, but its effects on individual students differed in extent and form. Ten of the twelve students reported the test impact on their motivation. Under the influence of the test, they were behaviourally, cognitively and/or emotionally more engaged in their learning. However, students were motivated not only by the TOEIC test, but also by their awareness of the importance of English for their future employment. Top students tended to perceive the test impact on their motivation to be minimal. One of them even completely denied it.

There was a general tendency of learning to the test. Generally, the students reported that they focused on what was tested or what was perceived to be useful for the test, and did not learn what was not tested or learned it very little. Overall, they spent more time on most language skills than before, but their time increase for the tested skills was perceived

to be much larger. The reason for devoting more time to speaking, a non-tested skill, was the students' desire to improve their communicative ability for their future employment or their personal needs of using English for communication in real life. Time for writing was not increased as much as speaking, and the increase was attributed to a variety of factors: learning for general language ability, learning for IELTS, learning for course assessment, and/or students' strengths or weaknesses.

The interviews also suggested that the TOEIC test influenced students' learning activities. The majority of them were involved in several direct test preparation activities, e.g., taking mock tests, learning test-taking strategies, doing practice tests, learning TOEIC extra materials, looking for information about the test and/or taking test preparation courses outside the university. Several students also used indirect test preparation activities, which included both form-focused and language use activities, to prepare for the test. They read or watched media in English, listened to English songs, and/or learned non-TOEIC materials, usually grammar material. The findings showed that the students from Universities B and C tended to be more involved in direct and indirect test preparation activities than those from University A.

Several students, especially the top students, said that they were engaged in certain learning activities without test preparation purposes in mind. They watched or read media in English, communicated with others, especially with foreigners, learned extra non-TOEIC materials, and so on. They explained that they did these activities to improve their English ability in general, to improve their communicative ability or just for entertainment.

The results of the students' journal sheet analysis were, generally, consistent with what they reported in the interviews. During the four weeks when the student recorded their learning, they spent the largest amount of time on listening (41.23%). The other skills that also received large attention were reading (20.30%), grammar (9.89%), and vocabulary (9.80%). Seven out of the eight students from Universities B and C were engaged in direct test preparation and/or indirect test preparation in their out-of-class learning time. By contrast, the students from University A did learning activities either without test preparation purposes in mind (e.g. to improve their communicative competence, to read specialized materials, to use English for their personal needs, and to get entertainment) or for their general English ability in the hope that this would be useful for TOEIC later. Seven students from Universities B and C also did some activities for other purposes. The

areas the students most frequently perceived improvement in were listening, grammar, and vocabulary.

It was noticeable that the top students tended to be less influenced by the TOEIC test than the lower ability students. They often aimed for real communicative ability. Also, the students from Universities B and C were more inclined to do direct and indirect test preparation activities than the students from University A. The latter usually learned for general English ability or purposes other than test preparation. (For more detailed results of the analysis, see Appendix P on the accompanying CD-ROM.)

8.3 Narratives of Student Cases

Three of the twelve student cases were selected to be presented in this section: Khoi, Thien, and Toan. As students' ability was shown to be an important factor that caused the major variability in the impact of the TOEIC test on student learning regardless of the courses they were taking, these students were chosen to represent three language ability levels: bottom student (Khoi), top student (Thien), and average student (Toan). They best illustrated the different patterns of test impact on learning and most clearly revealed the factors involved in shaping it and their interaction. Last but not least, these cases provided the researcher with rich and detailed information about their learning.

Each story begins with the participant's personal and English learning background, their beliefs about language learning, and their perceptions of the TOEIC test and TOEIC benchmarks. Then it depicts the perceived impact of the test on their learning through the interview and journal sheet data. It also discusses the roles of the influencing factors that emerged from their stories as well as their interaction to shape the test impact. It concludes with a short summary of the case.

8.3.1 Case 1: Khoi.

Personal and English learning background

Khoi was a third year student at University A. He was born into a farmers' family of seven people, who lived in a poor suburban district in Thua Thien Hue Province. His commune often suffered from natural disasters, so all students' school fees were paid for by the government. Khoi was a very quiet and taciturn student. He spoke quietly and often

answered me with very short or broken sentences, which were sometimes difficult to interpret, so I often had to ask him for clarification in the subsequent interviews.

Khoi experienced some financial difficulty. He was staying with his cousin, so he could save the rent, and his parents paid for his learning materials and some of the living cost. However, he still had to work part-time to support himself. Khoi said, “Sometimes I get enough money; sometimes I’m short of it.” At the time I was working with him, he had a job delivering airplane tickets to customers. He worked three afternoons a week and got about USD 50 per month.

He planned to work for a company for one or two years after graduation to get experience and then work with friends to set up their own business. He said he was working to save some money for that purpose. He preferred working for a foreign company but he said “that will depend on the conditions.”

Khoi studied English for seven years at school, but before he entered university, his English knowledge was almost zero. English was not taught in his primary school, which was small with only two or three classes and one teacher. However, when he was in junior high school, he was put together with students who came from the main primary school of the area and had learned English before. Therefore, he could not catch up with his friends and got discouraged. He said, “I didn’t understand anything... I mainly played. I sat around in class. Otherwise, I played football.” On top of that, he did not have to take the English test for the university entrance exam, so “I didn’t study [English]. I didn’t pay much attention to it. I focused on Maths, Physics, and other subjects.” When asked how he could pass the English tests without learning, he explained, “because I knew something, and my friends showed me the answers.”

Only at university did Khoi start to learn English seriously. Besides the university’s regular English programme, he did not take any extra English courses. He had never taken the TOEIC, official or mock tests, because he did not have money. He evaluated his language ability at level 4 or level 5 on a scale from 1 to 10. He said that, compared to his classmates, he was “weak.” He also said he could not follow the English programme at the university. When I asked him some simple questions about himself and his family in English, he generally understood them, but only gave me short and broken answers. Khoi did not like learning English because “there are so many difficult things,” “it’s so difficult!

Sometimes I can't translate... I can translate some things, but I can't translate some other things. I look up words in a dictionary for a while, then I feel too tired, I get discouraged." Khoi planned to take the TOEIC test at the end of year 4, right before graduation.

Khoi did not have enough time to learn English since he had to work part time to support himself: "In the evening, I spend one or two hours to learn. I can't study during the day." He had enough learning materials. However, he needed a recorder to record his own voice when he practiced reading because his laptop was not very good for that purpose.

Beliefs about language learning

Khoi considered English very important because it is necessary for employment: "It's very important! Because now job interviews are all in English. If I don't know English, I'll fall far behind others." Therefore, when asked about the purpose of English learning, Khoi said, "My purpose is generally to get a better job, to do business with foreign partners more easily." He wanted to be able to listen to and speak English. Khoi considered vocabulary and communication the most important parts of language learning because "it's sometimes OK to use wrong grammar as long as people understand us. But if we use vocabulary incorrectly, they won't understand us." Communicative ability for him was simply "when we speak, they understand us; and when they speak, we understand them."

Khoi thought that university students should learn all four language skills because "whichever skill we can learn is good for us." Nevertheless, he held that "communication, that is, listening and speaking, is more important."

He believed that the best way to learn a language was self-learning at home and learning by communicating with people. He explained, "I don't know what it's like to learn at language centres, but I think self-learning, making own efforts are better." He thought learning by communicating with people was useful because "when we make mistakes, people will correct them for us." This idea came from his friends:

A lot of people advise me so, Miss. Some of my friends are learning English with foreigners; they enrolled in classes or are learning at home with foreigners. Generally, they talk a lot in class. They chat. And then if they make mistakes, they'll be corrected. My friends have improved a lot.

However, Khoi could not follow his friends' advice because he did not have the resources to do so. Regarding the role of test-taking strategies in improving TOEIC test scores, Khoi

agreed that it was important, “Yes, indeed! Because some students are only average, but get very high test scores” thanks to TOEIC preparation courses. He proved his point by giving me a few examples of his friends. He had six or seven friends, who got more than 500 after taking TOEIC preparation courses, although their real language ability was not good. When I asked him how these students were compared to him, he said:

Only a little better, Miss, because I see that their speaking skill is also average... Uh, generally, they know some more vocabulary than me... Speaking is so-so; listening is not very good, either. But their test scores are very high.

He himself also intended to take a test preparation course before he takes the test.

Perceptions of TOEIC

Khoi knew that the TOEIC test assessed two skills – listening and reading, and it had a multiple choice format. He had never seen a test himself and never intentionally looked for information about it, but his friends told him so. He said: “I haven’t really tried to find out [about the test]. I hear about it from my friends. Because at my place, there is a student who is cramming for the test so I... I hear what he said.”

He thought that the test was difficult for him, especially the listening section. He considered the reading section easier because “for reading, sometimes, if I can translate a few words, I can guess the meaning of sentences, so I think it’s OK.” That was the strategy he used when he did classroom assessment tests: “Usually, for the tests in class, I translate a few sentences, I translate some words, and then I infer the meaning of the sentences and select an option.”

When I asked him whether the TOEIC test was useful for university students, he said it was “probably useful” compared to the Vietnamese National certificate. He thought it would motivate students to learn: “Generally, the listening and reading sections helps students train their listening and reading skills a lot,” and “when we listen well, we’ll like to communicate, so... so we’ll gradually improve our skills.” He believed that this would eventually be useful for his future employment because “now that we have integrated [into the world], we have to work with foreign partners a lot.” He explained that he considered it more useful than the Vietnam National Certificate because of its strictness: “Because we can’t buy [TOEIC certificates], we have to study to... pass the test. But with the other certificate, sometimes, if we’re too lazy to learn, then when we graduate, we can buy it.”

Khoi was also well aware of the serious consequences of not achieving the TOEIC benchmark. “If I don’t achieve the benchmark within eight years, my degree will be cancelled, I won’t get the degree” and “if I can’t graduate from university after four years, then... if I can’t get the degree, I’ll be at a disadvantage compared to my friends, I’ll lose a lot of job opportunities.” However, he did not experience any pressure from his parents or teachers about test preparation. He explained: “I’ve been independent since I was a child, so my parents don’t take care of my learning. They’re only interested in the results, not in my learning process” and “in class, teachers only teach us, so that we can pass the subject.” Therefore, the most important factor pushing him to learn for the TOEIC test was to get his degree.

Impact of TOEIC on learning

According to Khoi’s responses to the interviews, the TOEIC test had positively changed his attitudes to English learning. It motivated him to learn. In high school, he did not learn anything: “I just sat idle with my friends in class. There was a group at the back of the class. We often sat around and then played football in the school yard. The teachers scolded us a lot!” However, since he knew he had to take the TOEIC test, he felt that English was more important, he took English learning more seriously, and he learned it more actively: “In the first term I didn’t know what TOEIC was, so I didn’t study much. Also, there was no English in the first term. (...). But since I first failed the English subject, I started learning.” He was worried that he would not be able to get the TOEIC certificate as he could not pass even the course assessment. This is an example of his efforts: “Usually in the evening, I try to spend at least 15 minutes to listen or read stories or the like before I go to bed.” Later, in another interview, Khoi said that the TOEIC test was the major factor that motivated him to learn. Another reason was “to know something” for “work and communication.”

Khoi had not started direct test preparation yet, so he did not know how stressful it was, but he told me, “I’m worried a lot! Because I’m afraid that my degree will be held back.” He explained that he had not started test preparation “partly because I don’t have the time. I’m too tired after work, and in the evening I don’t want... Because my health is also bad. The weather here often changes erratically, so I... (...) I have sinusitis”; and “I don’t have the resources [money] to... take test preparation courses or do mock tests”; “I don’t have a means of transport. It’s too far”; “Courses often start at 5 and finish at 7, and there are no

buses home.” He only did exercises in the Practice Book for class requirements, and he knew that some of them followed the TOEIC format.

Although Khoi had not started direct test preparation, he said that the test had made him pay more attention to the tested skills. When I asked him whether the test influenced what he learned, he said, “Yes, it does because I used to learn vocabulary a lot... at the beginning. I didn’t use to listen, and... I listened very little. Now I focus on listening a lot”; and “I’ve also started practicing reading a lot.” He learned reading by reading short stories in computer software: “I right click the passages, it reads them to me. Then I read the passages and record myself. After that, I listen to the recordings.”

He said that he learned speaking and writing very little “because TOEIC is mainly reading comprehension and listening.” On top of that reason, the requirements for these skills in the course assessment were not very high: “This mid-term [speaking] test is singing a song in English” and the writing in the final test would only require students to complete the blanks in a passage with given information or data, so “sometimes if we make grammatical errors, teachers won’t take them seriously.” Later, he told me that if he had to perform a conversation for the mid-term test and if the teachers were not lenient, he would learn these skills more. He still learned these skills a little because “sometimes I also want to communicate”; and “when I learn reading and listening, I write what I understand. I just write, but I did not pay attention to grammar.”

Khoi reported that compared to high school, he spent more time on all the language skills. His time devoted to listening, reading, and vocabulary increased somewhere between “a little” and “moderately.” The time for pronunciation went up moderately together with the time for vocabulary and reading. However, there was only a slight increase in the time for speaking and grammar, and the time increase for writing was even smaller. Khoi spent more time on vocabulary “mainly [to prepare] for the reading section of the TOEIC test.” He did not pay much attention to grammar because no one explained it to him, and he could usually do random guessing for grammar test questions. Despite the reported time increases for all the language skills, the actual time he spent learning English was probably not much because, in high school, he almost did not learn anything. In fact, he said, he spent from only half an hour to an hour to study writing each week.

According to Khoi, since entering the university, he also spent a little time doing some learning activities that he never did in high school. These included doing grammar exercises (buying test papers to do at home), repeating after the tapes/CDs, and doing translations (“when I read stories, I translate them to understand”). He did not read newspapers or communicate with people in English because his language ability was still too low. He explained, “When I read them [newspapers], I don’t understand anything”; “I’m taciturn... Also, my vocabulary is still limited, so I’m afraid of communication; I mainly learn by myself”; and “...sometimes I made a lot of mistakes, so I feel that my self-pride is hurt.” Another reason was that he had no-one to communicate with. He did more translation because without translation he could not understand and he had been used to that way of learning since he was at school. He only read short stories and watched movies in English. However, he watched movies mainly for fun, hoping that he could learn something for English.

Concerning his achievement, Khoi said that his test scores had improved a little from above 3.0 to 4.8 or 4.9, but he still failed both term 1 and term 2. He also perceived some moderate improvement in the listening, reading, pronunciation, and vocabulary skills because he did the listening and reading sections in the final test better. His speaking and writing were improved a little.

Khoi’s Journal sheet analysis showed that he spent more than 24 hours to learn language skills during the four weeks (see Table 8.2). Most of the time was used for listening (30.45%) and reading (25.10%). The next largest amount of time was spent on writing (19.75%). He spent only around 7% of his time on vocabulary and pronunciation and did not learn grammar at all.

The three learning activities that he did during each of the four weeks were reading stories in English, watching videos teaching English on the Internet, and doing textbook tasks. All of these activities were done to learn English in general and prepare for course assessment. None of them were for direct test preparation. Accordingly, the three materials that he used all the time were the university’s textbook, software programmes (to read short stories), and videos. He also used a dictionary during these learning activities. While software programmes and videos were used to develop general English ability, the university’ textbook was used for both developing ability and preparing for course assessment. These results accord with what Khoi reported about his learning in the interviews.

Table 8.2 Time Khoi Spent on Language Skills

Skills	Time
Listening	30.45%
Reading	25.10%
Speaking	9.63%
Writing	19.75%
Grammar	0.00%
Vocabulary	7.12%
Pronunciation	7.94%

Case summary

Khoi was a bottom student from University A. As clear from his report, the TOEIC test had some effects on his learning. He focused on the tested skills such as listening, reading, and vocabulary more than the other skills and more than before. However, what he learned was still to develop his general English ability and pass the course assessment. He had not been engaged in direct test preparation yet. There were several reasons for not doing direct test preparation, for example, his financial difficulty, his lack of time, and his bad health. There were other factors that affected his learning, too. For example, he did not learn grammar because no-one explained it to him when he did not understand, he could compensate for grammar knowledge by random guessing, and teachers were lenient with grammar errors. He did not learn through communicating with others because his language ability was too low, and he did not have resources and an environment to do so.

Khoi wanted to learn English to communicate, that is, to be able to listen and speak, and to use English for jobs. He also had to meet the university's TOEIC benchmark for graduation. However, for him, the TOEIC test and communicative ability were different: Achieving the TOEIC benchmark did not mean having communicative ability, and having good communicative ability did not necessarily result in achieving the TOEIC benchmark. Therefore, for the time being, he learned English for both purposes. However, due to his low language ability and immediate needs, he aimed for the TOEIC test more than for communicative competence.

8.3.2 Case 2: Thien.

Personal and English learning background

Thien was 19 years old. He was a second year student at University B. He went to Le Hong Phong High School for the gifted and majored in Physics. His family lived in Ho Chi Minh City. His father worked as a motorbike taxi driver, who took children in the neighbourhood to school. His mother had a food stall in a market. They used to have a very hard time when the whole family of four people had to live in a 16 square meter house. Now their life “is a little better,” but still “slightly difficult.” Generally he had enough time and resources to study English. He only needed more books that taught test-taking tips so that he could study at home himself. Thien had two plans for his future. One was to stay at the university and become a lecturer. The other was to go to France to get a Master’s degree. He was actually preparing for the second plan by taking a French course.

Thien had a strong English learning background. He was considered a top student by his teacher. He started learning English as an extra subject in Grade 3 in his primary school. In junior high school, he studied the compulsory English Programme at school and, at the same time, took English courses at a language centre. In high school, he only followed the compulsory English programme because he had to focus on Maths, Physics, and Chemistry for his university entrance exam. Now at university, he started learning English again and was taking a TOEFL preparation course. He got a scholarship for the course from a language centre because he had won a consolation prize at the National Physics Competition and the first prize at the City Physics Competition in Grade 12. He explained that although he wanted to go to France for further education, he was still learning TOEFL because he would need all four English skills later, and the scholarship was a great opportunity for him.

Thien got two Vietnam National English certificates, Level A and Level B, when he was in Grade 8. His TOEFL mock test score was 60. He had also taken TOEIC mock tests at his own university three times. His scores went up from 525 to 680. According to him, his university’s benchmark was not high because of the low language abilities of the majority of students.

It was an average level because (...) at University B, the majority of students’ English proficiency was pretty low. They focus a lot on their majors, but their English for

communication was low. Many students who come from provinces did not have the opportunity to learn English, so their English level was low.

Thien evaluated his English ability at only 6.5 to 7 on a scale from 1 to 10 because his TOEFL speaking and writing were not good enough:

I suppose I can communicate in English well, but, indeed, when I learn TOEFL,... The TOEFL programme requires the listening, speaking, reading, and writing skills which are,... sort of academic, so I have to talk about a certain topic, but I can't speak fluently. I'm not used to it. I have to gradually practice it (...). If I have to write, I also feel difficult because I write little. I need to practice more.

Among the four language skills, he thought his speaking skill was the best; then came reading, listening and writing. He explained why his writing was the worst:

I, indeed, find my writing the worst because in high school, I didn't get to practice it much. The high school English programme mainly focused on vocabulary. Teachers mostly had us learn vocabulary, grammar. (...). But there was very little writing. When a skill is not practiced much, it is bad.

We chatted in English a little bit. Thien was fairly fluent. He had good pronunciation and seemed to have a developed vocabulary compared to many other students that I interviewed.

Thien said that the English programme at the university was lower than the one in his high school. Also, in English 4 (TOEIC preparation module), there were only two skills – listening and reading, but in high school the English programme covered all four skills.

There were two programmes: basic and advanced. At my Le Hong Phong School, the whole school studied the advanced programme, so the average level was high. But now at university, although the English programme is divided into four levels – English 1, 2, 3, and 4, I think even level 4 is lower than my high school programme. When I learn TOEIC, I can increase my vocabulary, but grammar is not much taught.

Beliefs about language learning

Thien considered his purpose of language learning as the ability to communicate with other people:

In my opinion, my purpose of English learning is to be able to communicate with people. I think that's the most important purpose. Because... For some other students, the purposes are different, for example, to study abroad. But my purpose is... Because I don't intend to study abroad [in an English speaking country], communication is enough.

However, it turned out that the ultimate purpose of his language learning was to use English for work. Later, when I asked him why he wanted to be able to communicate, he said, “I think my future working environment will be with foreigners, so communicating in English is very important.”

Thien understood communicative ability as the ability to communicate verbally: “When we listen, we understand; and we can speak to respond to people.” But when I asked him whether he thought reading and writing were also communication, he said, “reading and writing are also communication, but at a higher level” and he used an email to a customer as an example of communication in writing.

Despite his perceptions of communicative ability as speaking and listening, he believed that the most important parts of language learning were listening and vocabulary because these could effectively compensate for the gaps in our speaking ability. The following is his explanation for his opinion:

Because it’s possible that... for example, we don’t know how to respond to people, but if we can understand, we can use our hands and gestures and try to find synonyms. And if we have a large vocabulary, we can find synonyms and speak.

He added that when we communicated that way “we might make mistakes, but people still understand what we want to say.” This opinion might come from his own experience of communicating with foreigners. He told me he often practiced English by going to the city centre and talking to foreigners.

Therefore, he held that ideally university students should learn all four language skills, but he still put more emphasis on listening: “I think it’s best to cover all the skills, but, basically, I still give priority to listening. As I’ve just said, if we can listen, then we can respond, and we can express our ideas even with body language.”

However, when I asked him what skills he would need for his future job, he said he would need to invest in speaking a lot and need to have good pronunciation because he would have to make presentations. It was likely that initially Thien’s ideas about communication were only restricted to everyday conversations. My questions made his perceptions change a little.

Thien also believed that the best way to learn a language was to be hard-working and learn a little every day.

I think to learn [a language], first of all, you have to work hard, that is,... all the teachers that I've learned with said that you have to learn a little bit every day, not to learn everything at one go. For the technical subjects that I'm learning, if the next day is the exam, for example, then today... generally, I can learn everything in one day. But language is something you can't learn everything in one day. You have to divide it into small parts, and it'll be gradually absorbed day by day. And you can gradually master it.

He also stressed the importance of real language use. He said we should be brave enough to talk to foreigners.

[We have to be daring] because if we only learn a lot, but we can't apply what we learn, then it's useless. (...) Like when we meet foreigners, we have to pluck up the courage to strike up a conversation with them. They won't come and talk to us.

Thien considered direct test preparation very important for improving test scores: "I think we have to focus on doing a lot of practice tests and know the test structure, the test format, and the tips, then our scores will be high." He gave his own example to prove the importance of direct test preparation:

For example, when I took the TOEIC test for the first time, I didn't know what TOEIC was. I didn't know what it was like, how many questions it included. I was really dazed. But the second time, I already knew that the reading section had 100 questions, the listening and 100 questions, what this part was about and what the techniques were. So I just did the test.

Thien believed that thanks to his test-taking experience, his second test scores were higher than the first one. He thought that the experience could help improve test scores by 30% to 40%.

Perceptions of TOEIC

In Thien's opinion, the TOEIC test assessed the reading and listening skills: "It has two sections – reading and listening, so I think it assesses reading comprehension and listening ability. These two serve communication a lot." Later he added, "... as for speaking, we don't speak at all!"

The TOEIC test was not difficult for him: "Only occasionally do I encounter new vocabulary words and have to make guesses [about their meanings]." To him, both sections of the test had the same difficulty.

Thien strongly believed that the TOEIC test was useful for students because it assessed the two important skills for communication, it was suitable for students' proficiency compared to TOEFL, for example, and its topics were familiar to students.

I think it is very useful! It only focuses on reading and listening, which I think are important for communication. And students' ability, specifically the ability of those at University B, as I said, is average and fair. The TOEIC test includes short communicative sentences,... for example, for listening, we only listen to short sentences and responses, and listen about photos (...). Reading is very short announcements, advertisements, and so on, so it is very close to reality. But in the TOEFL test, the reading is very long. If average and fair students looked at it, they would be shocked right away! So I think the TOEIC difficulty level is suitable for students.

Thien believed that the TOEIC test could only partly reflect learners' language ability due to its scoring rules. He also thought that the TOEIC test was suitable for graduation purposes because it was suitable for students' ability.

Thien was well aware of the consequences of failing the test: "I think the consequences are also serious (...). It's so annoying when your knowledge of engineering is absolutely fine, but you can't graduate just because of English." However, he did not find learning for TOEIC stressful because his ability had already exceeded the requirement: "I find it very relaxing because I'm taking this English course [only] to maintain my English level." He did not have any pressure from his parents and friends, either.

Impact of TOEIC on learning

The interviews with Thien showed that the TOEIC test had minor impact on his learning. At first, when I asked him if the TOEIC test had a big influence on his learning, he said, "Not yet. I still study in the same way." He also said that he had been studying seriously all the time:

It's always been like that [learning English seriously]. Because I don't dare to stop learning. Only when I was in high school, did I not pay much attention to English. Now looking back, I think that's bad. I rather regret that. If in primary and junior high schools, I studied English at a level of eight or nine, then in high school, I only studied it at a level of six or seven.

Also, he had always been active in class: "If I'm at home, I will be a little lazy... It's up to me to study a lot or little, but I've decided that once I'm in class, I have to try to be more active."

He also explained that at school, he also took English learning seriously because of the pressure from his teachers and family, but at university, students were conscious of their own learning:

Generally, in high school, I studied all the subjects seriously. In the high school environment, teachers pressured us into a strict frame. But at university, teachers don't pressure us. Their job is teaching, and students' job is learning. It's up to students to do extra learning.

In Grade 3, my father wanted me to know English in advance, so he enrolled me in English classes.

He also added that his current perceptions of the importance of English were associated with opportunities for employment:

I think if two engineers graduate from university with the same degree, but the person who knows English, who can communicate in English or even in another language, definitely has more opportunities for promotions and possibly has higher salaries. I'm sure.

Later, he admitted some signs of learning to the test. At university, he studied English more because he understood the importance of both English in general and TOEIC in particular. He also said that he focused on the tested skills – listening and reading:

Yes... I mean... since I knew I had to take the TOEIC test, I've been focusing on its two main sections, which are speaking and reading. Oh, no I'm wrong, listening and reading. I have to practice those skills more. I have to practice listening more at home and do exercises.

Nevertheless, the time increase for listening and reading was reported to be only small. Due to TOEIC, the time for vocabulary was slightly increased, too: "I certainly have to learn vocabulary more because TOEIC vocabulary is new. I have to learn vocabulary a lot."

On the contrary, the time for writing was said to have been reduced moderately because it was not tested and not taught in class: "I didn't get to practice writing much." Also, contrary to most other students and even teachers, Thien did not think that TOEIC focused on grammar, so his time for it reduced a little. Probably, because he had learned a lot of advanced grammar at school, he found grammar not important for TOEIC.

I find that TOEIC doesn't require much grammar knowledge. It only has some very familiar types. Just looking at them, you can be absolutely sure of the correct answers. But in the past... in high school, there were more complex exercises. For example, different tenses were mixed together. They were trickier.

The time used to learn pronunciation remained the same because Thien thought that his pronunciation was good enough.

It is interesting from Thien's responses, though, that the time spent on speaking remained the same although it was not tested. The reason was that his purpose of English learning was to communicate with people. "I still have to communicate with people. I mean, although I don't study it [in class], I still have to communicate with foreigners."

Thien said that he was also involved in direct test preparation. He practiced past tests using the website suggested by the university: "In my first year, I discovered the website suggested by the university, so I used it more. It serves listening a lot." He had taken mock tests several times. He also knew some test-taking strategies for the listening sections, but he had not learned the strategies for the reading sections. However, besides the university's TOEIC preparation textbook and website he did not use any other test preparation materials.

The interviews revealed that there were some changes in the kinds of activities Thien did to learn English compared to high school. For example, instead of only listening to textbooks, he used a website to practice TOEIC listening, and he listened to music, too. Also, he did real language use activities. As said before, he often went to the city centre to talk to foreigners. He also practiced English at coffee shops.

In high school, I rarely communicated with people to learn speaking, but at university I go out to talk to foreigners. (...) There's a coffee shop. There everyone has to speak English. A month ago, I found another shop where there are foreigners. They speak English to us.

Besides, at school, he often read stories in English, but now he just started reading more difficult material.

In the past, I only read stories and textbooks. The other day, I could borrow a book about the United States. I read it... It is difficult. And I also read TOEFL texts, which are very long. I have to sit and do [reading tasks] for a long time.

He said he was motivated to learn this way because he was aware of the importance of English and TOEIC: "I think this is because of my purpose. Once I have understood the importance of English and TOEIC as well, (...) I have to learn more. So I have to change." Some of the changes could be directly associated with TOEIC, for example, listening to

TOEIC materials on a website. Changes in reading were also attributed to the difficulty of TOEIC, but TOEFL seemed to play a larger part:

As for reading, I think [the changes] were partly due to TOEIC and partly due to TOEFL. For example, a TOEIC passage is only 10 lines long with five or six questions. But a TOEFL passage is very long and I have to use different techniques. I have to practice reading more to understand more.

Thien also watched TV, such as Disney or Discovery channels to have fun and practice listening for TOEFL and TOEIC. He denied any connection between speaking activities and TOEIC: “[Changes in] speaking] are not [due to TOEIC].”

However, Thien asserted that his ways of practicing language skills still remained the same. For example, he did not often look up words in dictionaries when he did a reading task. Instead, he used guessing skills to work out the meanings of the words and reading passages. He used the same procedure when he did listening exercises. For speaking, he wrote a passage, then memorized it, and practiced delivering the speech in front of a mirror. In general, no associations between the test design and ways Thien practiced language skills were found.

As for vocabulary, a change was reported. Because of the much larger number of words to learn, he had to divide them into small parts to learn every day:

Now TOEIC vocabulary is more than before. In the past, in a unit, there were only 5 to 10 new words. But now, in a TOEIC session of English 4, there might be up to 20, 30 new words. Then I have to divide them into parts, so that I can learn them effectively. If I learn all of them at one time, then I can remember them today, but forget them all in the next day.

However, he still learned vocabulary words by reading them aloud and copying them many times to memorize them. Sometimes he also learned interesting sentences with the new words.

Thien was not sure about the influence of TOEIC preparation on his mid-term and final test scores because he was exempt from English 1 and 2 for his high entrance mock test result, and English 3 was not a test preparation course. However, he thought that his listening skill had improved a lot “because in class, we practice listening a lot.” Reading was also improved “because in class we practice only reading and listening,” but slightly less than listening. Vocabulary was improved thanks to TOEIC, too. But it was only a little

better because the number of TOEIC words to learn was large, but they were not used frequently, and, therefore, difficult to remember.

There are so many words, but I encounter them only once. For example, in high school, you had one new word in Unit 1, but you could see it again in Unit 2 and in mid-term or final tests... or in 15 minute tests, one-period tests. When we see the word many times, we can remember it longer. But now, for example, you hear a word in a listening text, you can be sure that you will never see it again in your tests. If it does, the probability was only one to two percent. (...) so it's very difficult to remember it.

Speaking was improved a little, too, but was not due to TOEIC: "It is better, but not due to the learning in class" because "in class, English is not used at all. All the speaking is in Vietnamese. So is writing." Writing remained the same for the same reason: "Because there isn't an environment to write. There is no pressure, environment to write." Pronunciation also remained the same; grammar had become worse.

The results from Thien's *journal sheet analysis* showed that he spent only 13 hours on language skills over four weeks. The largest percentage (about 45%) was devoted to listening, and the second largest – reading. Next came vocabulary (about 14%). Time spent on writing and grammar was extremely small – less than 2%. These results supported what Thien said about his learning.

Table 8.3 Time Thien Spent on Language Skills

Skills	Percentage of time
Listening	45.16%
Reading	24.06%
Speaking	8.45%
Writing	1.81%
Grammar	1.23%
Vocabulary	13.87%
Pronunciation	5.42%

As he recorded in his journal sheets, Thien used a variety of materials to learn English during the four-week period. Two of them were to prepare for TOEIC, two for TOEFL, and other materials were for grammar, speaking, and writing.

Concerning learning activities, Thien did both direct test preparation and indirect test preparation during the four weeks. For direct test preparation, he did exercises in the TOEIC textbook and TOEICpedia.com. For indirect test preparation, he watched TV for example, Disney and Discovery, which he believed could help him improve listening for TOEIC and TOEFL. He also did things not related to test preparation, for example, listening to music for fun and practicing listening in general.

Case summary

Thien was a top student from University B. From his self-report data, it can be seen that the TOEIC test had some effects on his learning. Thus, due to TOEIC, he focused more on reading and listening than before, and he was engaged in both direct test preparation and indirect test preparation. However, the test impact was perceived to be small. Generally, the time increase for the tested skills was only a little. Besides course materials, which were compulsory, he only used a website to practice TOEIC listening (he used it in only one week of the four week period that he recorded his learning for). He had not learned many test-taking strategies. There were almost no associations between the nature of the TOEIC test and the way Thien practiced his language skills. The reason for the small extent of the impact of the test was that his language ability already exceeded the test requirements. Also, he was motivated to learn English by his awareness of the importance of English for employment more than by the TOEIC test. Therefore, he focused on developing real communicative ability. He was engaged in various language use activities, especially talking to foreigners, because he believed that the best way to learn a language was through real language use. Thien's learning was a result of an interaction of a variety of factors: his beliefs about language learning, his learning background, his language ability, his perceptions of the test, and the nature of his English course.

8.3.3 Case 4: Toan.

Personal and English learning background

Toan was 20 years old and a third year student at University B. He was the second son in a family of five people, who lived in a poor suburban district of Ho Chi Minh City. His parents made a living by growing medical herbs and selling them to medical shops in the city. Toan did not have favourable conditions to study. He was in some financial difficulty. He did not have enough time to study:

I used to [work part-time], but this term, I've stopped working because this term... now I have only Monday evening, Wednesday morning, and Sunday afternoon free. On the other days, I have to study all day, in the morning, afternoon, and evening.

He also felt he lacked learning materials. His place was inconvenient for learning: "I have to listen on headphones. Only when my roommates are out can I use the speakers to listen."

Toan had little English background. He started learning English in Grade 6. However, because his family was in the most difficult social-economic area of the city, English was not paid attention to:

I lived in the most difficult social-economic area of the city, near Cu Chi Tunnel. There English isn't paid attention to. It's only taught for the sake of being taught. Right in junior high school, we learned English as a mere formality. Teachers... like... gave us marks for free. (...). In high school, the situation was the same. When the high school graduation exams came, the teachers proposed to the Department of Education that English should be replaced by Physics because, essentially, the whole school was bad at English. But the Department of Education rejected the proposal because Ho Chi Minh City disagreed. Then the teachers began to drill us in basic grammar, so that we could take the test. The drill continued for more than a month. So the school's English test results were low. (...) Only more than 40% of students had above average scores.

However, Toan got 7 for that exam. His English learning experience at school was restricted to vocabulary and grammar. He memorized word lists and grammatical formulas by copying them many times.

In high school, in Grade 12, we learned the English book written by the Ministry of Education. Each day, in the lesson, there was a passage. The teacher asked us to look at the passage and pick out all the new words. Then she put them all together [in a list] and translated them. She wrote the English words on one side, the part of speech and Vietnamese equivalents on the other side. Then we learned them at home. The teacher made us write each word ten times to memorize it. That was vocabulary. The second part was grammar. Each lesson was about one point such as the simple present tense, the simple past tense, the simple future tense, the present perfect, the past perfect. The teacher made us copy down the formulas $S + V + O$, then she gave examples, and we copied the formulas five times to remember them. (...) The next day, she gave us a test. She read the English words. We wrote the English words, the part of speech and the Vietnamese words. Then the teacher asked us, for example, what the formula for the simple present was. We wrote the formula and then gave examples. (...) Usually we used the examples we'd learned in class, which the teacher made us learn by heart. We were afraid of making spelling mistakes if we wrote our own examples.

Besides the English programme at school, Toan did not have any extra language classes because his family had financial difficulty.

When I interviewed him, Toan had done three General English modules – English 1, 2, and 3, and he was doing English 4, a TOEIC preparation course, at the university. He had also been taking TOEIC preparation courses in language centres outside the university since his second year. At the language centre he was studying at, each class session was divided into two parts: two periods for reading and two periods for listening. The reading teacher used Longman Preparation Series for the New TOEIC Test. He taught reading skills and grammar starting at the beginning level. The listening teacher let students do a lot of listening exercises and taught them the trick of test taking.

Toan had taken TOEIC mock test several times, and his highest score was 400. He planned to continue doing test preparation and take the test a few months later. He had never taken any other English certificate exams. Toan evaluated his own English ability at 6 on a scale from 1 to 10. His achievement test scores were usually above 6 and below 7.

Toan liked learning English because it helped him discover not only a new language, but also a new culture:

That's a second language, which has lots of mysteries. I like it. I can learn new things. It helps me gain more knowledge. I find that learning English is not only to know another language, but also... like... we know more about other people's culture.

Toan was a very hard-working and ambitious student. His career goal was to become a lecturer at University B. To realize his goal, Toan was studying hard to improve his scores of many subjects to 8.0. He also wanted to work for a company, but just to earn his living. He preferred the teaching job at the university. He had always wanted to become a teacher since he was a child. Toan also planned to get a second degree in Engineering to widen his knowledge, which would benefit his future job.

Beliefs about language learning

Toan's ultimate purpose of English learning was to have communicative ability for his future jobs: "I think today, if we graduate without English, we probably can't get a job. I'm trying to be able to communicate in English. I'll definitely need it to get a job." However, he found that learning for communicative ability and learning for a TOEIC certificate were different, so his short-term objective was to get the TOEIC certificate for graduation:

Now I'm learning to get the TOEIC certificate... to complete the study programme and get a degree in Civil Engineering. After that, I will learn to know, to communicate in English. Learning to communicate will probably be different from learning for TOEIC.

When I asked him why he thought that the two types of learning differed, he said, "Because I find that preparing for TOEIC is mainly learning skills and tips to get high scores. But it's not learning to be able to communicate with others."

Toan understood communicative ability as the ability to communicate verbally:

Communicative ability is... For example, when we talk to a partner, if we have the ability to express our ideas, talk about our work, and the partner can understand us, and the two parties can discuss the matters fluently, then our communicative ability is satisfactory.

Toan believed that the most important parts of language learning were listening, then grammar, and finally vocabulary. Listening was the most important because it was necessary for communication, and he was weak at it. Grammar was also important because grammatical errors could "change the meanings and have negative effects on the listeners." Vocabulary was important because "without vocabulary nothing can be done."

Toan held that university students should learn listening and speaking because they were necessary for communication and for their future work:

Miss, I think in the university programme, we only need to learn listening... listening... and speaking. [...] because... I'm learning a TOEIC programme, I find that... Listening and reading are, indeed, TOEIC programme. But listening and speaking are more important to know, to improve communicative ability, and to communicate with others. Only these will improve your language ability.

When I asked him about skills he would need for his future jobs, he said, he would need all four skills, but he still attached more importance to speaking and listening.

To achieve communicative ability, Toan believed that learners should be in an environment where they had to use English for communication: "The only way to be good at a foreign language is to communicate with foreigners a lot, take part in many clubs, English clubs. That's the best way."

However, Toan had not done any of these activities. One of the reasons was that his English ability was low:

Once I sneaked to the student association. They spoke English only and spoke continuously! I looked at them from the office, and I thought if I came there, I'd stand quietly, and I'd feel very ashamed. I'm afraid. I don't dare to join them.

As for the TOEIC certificate, the best way to learn was to practice listening and reading a lot and to learn grammar, vocabulary, and test-taking strategies.

Miss, the best way to improve TOEIC scores is... For the listening section, the only way is to practice listening a lot. And we have to know the ways to avoid traps. (...). And for the reading section, doing exercises regularly to get used to the grammar, memorize the vocabulary, and to do questions fast. Only by practicing a lot can you do that.

Thanks to practicing test-taking tips, Toan's TOEIC mock test scores had increased from 350 to 400 within more than one and a half months.

Perceptions of TOEIC

Toan knew that "the TOEIC test assesses listening ability and reading only," but he thought that grammar, vocabulary, and pronunciation were needed to do the test.

For the reading section, we have to learn knowledge about grammar and vocabulary. For the listening section, we need vocabulary, we have to understand the meanings, and... listening.... generally, we need to know pronunciation.

He also knew that the test had a multiple choice format. Toan found that the TOEIC test was "very difficult" for him, and the listening section was "extremely difficult" because of the fast speed, the word liaisons, and the pronunciation which was different from his teacher's. He said he was "ploughing extremely hard" for TOEIC.

Toan did not think the TOEIC test was useful for university students: "I don't think it's useful for university students. University students should learn communicative English. It's more suitable, more applicable." When I asked him to clarify why he thought what he learned in TOEIC couldn't be put into use, he elaborated that TOEIC was not as useful as communicative English because only the listening skill would be of much use for work:

Not as [applicable as] communicative English... It can be used, but when you work later, it's not as useful. When we work, we mainly communicate, but TOEIC... We can use the listening skill, but the reading skill is less used.

Toan also believed that a TOEIC certificate could be useful for job applications if only the score was high, above 700. He said, "If I were an employer, I would just accept applications when applicants had a test score of more than 700." He considered a

benchmark of 450 too low for employment. In his opinion, a person with a score of 450 “can work for a domestic company, but he can’t work for an international company.” However, he said he understood the reason why the university had adopted such a low benchmark.

I understand that the university wants to help their students improve their language ability. Traditionally, technical students are bad at foreign languages. This initial benchmark has been set up at a low level in order to experiment. It aims at making students pay more attention to the importance of foreign languages to study [harder].

Toan was aware that failing to meet the TOEIC requirement would have serious consequences for students. He knew that a lot of students were suspended from their studies because they did not achieve TOEIC 400 in the middle of year 4. Therefore, he was much pressured about the TOEIC benchmark. However, he did not experience the pressure from his parents because “my parents don’t know. They’re only farmers.” Only his current English teachers often pushed him and his classmates to study for TOEIC, the other teachers only aimed at helping students pass their course assessment.

Impact on learning

Toan admitted he was stressed about the TOEIC test because of the heavy workload he had.

In years 1 and 2, I didn’t feel stressed because I thought if I practiced a lot, I would achieve it. But now, in year 3, I have to study too much! I have to study a lot [for my major], and at the same time, for a TOEIC certificate, so that I can continue my major. So it [TOEIC] is as heavy as a second major.

Test preparation took a lot of his time away, and he had financial issues, too: “I have to spend a lot of time to study. It takes my time away. And the money issue is stressful.” He had to spend USD160 on his current two-month TOEIC preparation course. He also had to spend money on transport from home to the language centre. The test-taking fee was USD29 for students at University B.

Toan agreed that the TOEIC test motivated him to learn: “It urges me to learn in order to meet the progress requirements of the programme.” It even made him study more seriously than in English 1, 2, and 3. The reason was that the assessments for the first three modules were easier and the teachers were lenient: “In English 1, 2, and 3, to prepare for the final tests, I bought the past papers and solved them. In the real tests, the same types [of

content] are tested” and “The listening test is usually just for teachers to give us more marks.”

Toan reported that compared to high school, he spent much more time on listening, vocabulary, and pronunciation.

Now [I spend] more time [on listening], and I listen according to my plan. In the past, the teacher also told us to buy CDs; at home I also listened but stopped after a while. And I also rarely listened. Now I have to make plans. Every day after school, I have to listen to the unit’s materials to do the exercises. I complete all the exercises in order to switch to other subjects.

He needed a lot of time to learn vocabulary because “I often forget vocabulary. I learn it today and then forget it tomorrow.” Time for pronunciation was increased to the same extent as vocabulary because they were learned together.

Although reading and grammar were also needed for TOEIC, the time increase for these skills was perceived to be smaller than that for listening, vocabulary, and pronunciation. The reasons were that Toan had already learned them a lot in high school: “I read a lot before. (...) In high school, we mainly read and did the exercises in the textbooks”; “I only review grammar. To review it, I don’t need much time because, for example, I already know a tense, but if I forget something, I just review it a little bit.”

He explained that he reserved more time for these skills because they were tested in the TOEIC test and difficult to him. He only used the materials of university and language centre to learn them.

Toan also said that his time for speaking and writing remained the same. The first obvious reason was that they were not tested in the TOEIC test and were not included in the English programme. The second reason for not practicing speaking was the lack of an environment for language use. Toan said, “[I didn’t spend more time on] speaking because I don’t have the conditions, the environment to practice. Also, every day, in class, we don’t speak much. Writing is also not much related to TOEIC.”

The impact of the TOEIC test on Toan’s content of learning was mixed with the impact of the course assessment because they were, basically, the same in nature. When I asked if the course tests influenced his learning, Toan said, “They’re also related to this [TOEIC].”

Toan's responses to the interviews showed that he was almost totally engaged in direct test preparation. He was taking test preparation courses both at the university and outside it. He had taken mock tests a few times. He only used test preparation materials to learn English. He did TOEIC exercises every day.

Besides direct test preparation, he reported being involved to a small degree in some language use activities, but they were not related to the TOEIC test. Thus, he rarely read newspapers in English. He spoke English to his friends a little, but just for fun. He watched movies in English a lot, but this activity was not intentionally for learning English in general or TOEIC in particular: "My roommate watches them, so I join him. He majors in Economics, so he 'ploughs on' with English a lot. He practices, and I watch them with him."

The interviews also revealed that there were several slight changes in the ways Toan practiced language skills; however, many of them were not associated with the test design. It was likely that because Toan was more motivated to learn English than before, he learned more, but he just learned the way it happened to be. Toan admitted that he was not aware of any learning strategies. Some exceptions were the changes in his ways of practicing listening and reading skills, which he learned for the TOEIC test. In high school, he listened and looked at the scripts in his textbook at the same time, but now he listened without scripts and only used them when he wanted to check his understanding. Obviously, the textbook design and the teachers' teaching influenced the way Toan practiced listening in high school.

In high school, I mainly listened in class. The teacher brought a CD player into the class. In the book, there was a passage. The teacher let us listen, but we only listened and looked at the same time. We listened to the CD and looked at the textbook. We only listened that way and then answered the questions below [the text].

It is also clear that the way Toan practiced listening now corresponded to the test tasks. Probably, the difficulty of the TOEIC test and the types of test tasks caused the change. There was a change in the procedure he used to do reading exercises, too. He said, "In high school, I read the whole text, then looked at the questions, and looked at the text again. But now, I mainly read the questions first, and then find [the answers]." He cited the limited time allowed for TOEIC test questions as the reason for the change in his reading: "Now I don't have enough time to read that way [from the beginning to the end]. I only read

questions, then answer options, and infer answers from the text above. I look for similar words.”

Concerning the perceived impact of TOEIC on his achievement, Toan said the test preparation improved his listening, reading and vocabulary a lot. His mock test scores increased from 325 to 505 during a period of one and a half months. He got more correct answers than before. His grammar and pronunciation were also improved a little. However, the quality of his learning might be questionable. When I asked him whether he could understand reading texts more than before, he said, “That’s not reading comprehension. I mainly read (...) to answer questions only.” I did not immediately realize what he meant by “That’s not reading comprehension,” but later, when I was transcribing the interview, the sentence caught my attention. Therefore, I asked him to clarify. He explained, “I use the strategies I’ve been taught to do exercises quickly and effectively instead of translating sentences. The purpose is to get the most correct answers in the shortest time only.” On the one hand, the fact that Toan stopped translating every sentence he read into Vietnamese was certainly a positive effect of TOEIC on his learning strategies, but, on the other hand, the goal and habit of reading to answer questions correctly without understanding were detrimental to the development of his language ability.

As for speaking and writing, the two non-tested skills, they were not improved at all. His achievement test scores just increased by less than one point during the four terms of English.

The following are the results of Toan’s journal sheet analysis. Unfortunately, they could not accurately reflect the true amount of time he spent on the language skills because he did not record the time for learning vocabulary for three weeks (instead, he wrote the number of words he learned every day), and he did not record his learning time in week 3, when he learned only listening and vocabulary, because he did not learn English much, being mostly busy with his internship. However, the results still supported what Toan said in many respects. Although he did not record his time for listening in week 3, the percentage of time spent on listening was still the largest (39.19%). Next came the amounts of time for reading and grammar (30.15% and 21.62%, respectively). The time spent on vocabulary for one week accounted for about 9% of the total time, so it can be inferred from this figure that the actually time should be large.

The journal sheets also showed that Toan only used TOEIC preparation materials to learn English. Those materials were Longman Preparation Series for the New TOEIC test: Intermediate Course, Longman Preparation Series for the New TOEIC test: Introductory Course, and 600 Essential Words for TOEIC.

Table 8.4 Time Toan Spent on Language Skills

Skills	Percentage of time
Listening	39.19% ^a
Reading	30.15%
Speaking	0.00%
Writing	0.00%
Grammar	21.62%
Vocabulary	9.04% ^b
Pronunciation	0.00%

Note. ^aWeek 3 was not recorded; ^btime recorded for one week

Apart from direct test preparation activities, Toan recorded some other activities such as listening to music (every week), watching movies (one week). These two activities were mainly for entertainment and to a small degree to learn English in general. He also participated in a conference with English being a working language, which he assumed to be good for his listening skill.

Case summary

Toan was a student from University B, whose mock test score nearly met the language benchmark. The impact of the TOEIC test on Toan's learning was apparently a result of an interaction of various factors. The first ones were the test factors. The high stakes of the test strongly motivated Toan's learning. He was mostly engaged in direct test preparation activities such as taking mock tests, taking direct test preparation courses both inside and outside the university, and doing TOEIC exercises every day. He only used TOEIC preparation materials to learn English. His perceptions of the nature of the test entirely determined his content of learning. Thus, he only learned the skills he believed to be tested

or useful for the TOEIC test such as listening, reading, grammar, vocabulary, and pronunciation. He almost totally ignored the non-tested skills including speaking and writing. The nature of the test was also somehow associated with the ways he practiced listening and reading. However, he learned some tested skills more than the others due to the difficulty of the skills and his English background. He learned listening the most because it was extremely difficult for him. He spent less time on grammar and reading because he had learned them a lot in high school. His overall language ability was also an important factor in his learning. Like Thien, he believed that he needed communicative English for his future employment. However, while the other student had the ability to focus a lot on communication, Toan only tried to cope with the benchmark and intended to learn for communication only after graduation. Due to his limited ability, he did not dare to participate in English club. The context factors were also somehow responsible for how Toan learned. He did not communicate with foreigners because he did not have an opportunity to do so. Toan's beliefs about language learning, on the contrary, played a minor role in his current learning. Although he believed he should learn communicative English, and he perceived that learning for TOEIC was not useful for developing communicative competence, he was still totally engaged in test preparation temporarily ignoring his long-term objective. Thus, his beliefs about language learning were overridden by the TOEIC test. This is clearly evident in what Toan said: "If the university abolished the TOEIC test, I wouldn't learn TOEIC. Instead, I would learn communicative English."

8.4 Chapter Summary and Discussion

This chapter reported the findings from the student case study. It began with a summary of the results of the quantitative analysis of all the student cases. The summary showed that the majority of students had limited English backgrounds. They all shared the beliefs that language learning is for communication, and the best way to learn a language is by communication. They also knew the test design very well, except for one student. They generally believed that the test was useful, but many of them believed that real language ability was needed to get a good job, not a language certificate. The test was shown to have influenced most of the students' learning, but the extent and form of its impact differed greatly across individual students. The top students did not seem to have been affected much, while the lower ability students tended to learn to the test more intensively. Also,

the students who were doing the TOEIC preparation courses at their university were more likely to be involved in direct and indirect test preparation in their out-of-class time than those who were doing the Business English course.

The stories of the three students, Khoi, Thien, and Toan, each representing an ability level, more vividly illustrated the patterns summarized in Section 8.2. Thien, with a TOEIC mock test score exceeding the exit requirement, was not pressured about the test at all. He did a little test preparation, but he learned a lot for other purposes. Khoi, a bottom student from University A, was pressured about the test, but he was still learning for his general language ability and not involved in direct test preparation. Toan, an average ability student, most intensively learned for the test and to the test. His case gave the clearest example of both positive and negative effects of the TOEIC test on student learning. On the one hand, Toan was highly motivated by the test. On the other hand, his learning was only superficial aiming at getting high test scores by using test-taking strategies rather than developing real communicative ability.

The stories also helped to see the interaction among influencing factors and explained the test impact on student learning. They showed that the three students' learning was determined by a number of factors, e.g., test factors, beliefs about language learning, English learning background, language ability, needs, and possibly the nature of courses. However, some factors were probably more influential than others. Certainly, the test factors were among the most influential. Its impact was evident in most of the cases analysed. It motivated the students to learn, and its design directed what they learned and, to a lesser extent, how they learned. However, the impact of the TOEIC test may have been strongly moderated by students' ability, which resulted in major variability among students. All of the students had three different goals in their English learning: passing course assessment, achieving the TOEIC benchmarks in order to graduate from university, and having communicative ability for their future employment. The first goal was the most immediate and the easiest to achieve (Most of the students in the case study did not seem to worry much about this goal). The last one was the most distant and probably most difficult to achieve. The second one was in the middle in terms of both distance and difficulty. Depending on their language ability, the students chose to pursue different goals when they were still at university. Thien's high language ability, which exceeded the language exit requirement, gave him the freedom to learn for communicative ability, his

ultimate goal, through genuine communication, the way he believed to be most effective. By contrast, for the time being, Toan and Khoi, who had lower language ability, almost entirely aimed at the TOEIC test and, possibly, course assessment in Khoi's case because they were more immediate than communicative ability. Also, despite his beliefs about communicative learning, Toan was involved in test preparation activities that he knew were superficial learning and not useful for developing communicative competence. Toan's comment that he would learn communicative English instead of TOEIC if his university abolished the TOEIC policy revealed that the TOEIC requirement, to some extent, constrained students from learning for communication.

The nature of the course also seemed to be responsible for the variability among the students. Among the three students, Khoi, a student from University A, had not been involved in direct test preparation yet. By contrast, both Thien, and Toan, who came from University B were engaged in direct test preparation to different extents. Their behaviours reflected the common patterns reported in Section 8.2. At Universities B and C, where test preparation courses were taught, all the students, regardless of their ability, were involved in direct test preparation. However, at University A, where students learned business English, the low ability students had not been involved in direct test preparation, although they were in the same year as the students at the other two universities. Instead, they were still learning for general English ability with more emphasis on the skills tested in TOEIC. All the students from University A, including the high ability students, intended to do direct test preparation only closer to their test-taking time. Therefore, the non-TOEIC learning programme might have delayed the impact of the test on some students' learning. However, this should be interpreted with caution because other factors may have been involved. Khoi said he had not started taking test preparation course because of his personal circumstances. Thus, no conclusive evidence was found due to the small number of cases.

Course assessment was not shown to be clearly connected to the majority of the students' learning. At Universities B and C, its effects might have overlapped with those of the TOEIC-as-exit-test because, to the students, they were essentially the same. Where the two kinds of assessment were different, course assessment did not seem to strongly influence students' learning because of its easiness or teachers' leniency in marking test papers. Thus, it did not take Toan much time to prepare for the final test in the General English

courses. Khoi did not learn writing because teachers would be lenient with students' grammatical errors. Generally, students did not seem to be as worried about the course assessment as the TOEIC-as-exit-test.

The TOEIC test and students' ability were probably the major factors that determined the degree and the main forms of test impact on students' learning. The nature of the course nature might have also been important. However, there were other factors that possibly created finer variability among students of the same patterns, e.g., learning to the test or not learning to the test. Those included students' beliefs about specific aspects of language learning, their English background, their perceptions of the comparative difficulty of the listening and reading sections of the TOEIC test, and conditions for their learning. For example, Toan exclusively learned to the test and perceived that grammar was needed for it, but he did not learn grammar much because he had learned a lot of grammar at school. All of the students paid more attention to listening than to reading, although both were tested, because listening was more difficult for them, and they did not learn listening much at school. Khoi and Toan did not learn by communicating with people partly because they did not have an environment to do so.

Chapter 9: DISCUSSION

This chapter begins with a summary of the findings to the research questions. Then it discusses the major issues emerging from the findings. The study aimed at answering the following research questions.

1. What are policymakers' perspectives on the TOEIC-as-exit-test policy?
2. What is the impact of the TOEIC-as-exit-test on university English teaching and learning?
3. What brings about the impact of the TOEIC-as-exit-test on university English teaching and learning?

9.1 Summary of Major Findings

9.1.1 Policymakers' perspectives on the TOEIC-as-exit-test policy.

Policymakers' perspectives on the TOEIC-as-exit-test policy were sought through semi-structured interviews. They were asked about their beliefs about the purposes of language teaching at university, their rationales for choosing TOEIC as a university language exit test, and the intended impact of the TOEIC policy. The findings showed that their perspectives reflected the view of the importance of English prevalent in society. The theme of English for employment was prominent in all the three topics they were asked about. Twelve out of the fourteen participants talked about students' ability to use English for communication in the workplace as the purpose of language teaching at university. The same number of participants gave job-related rationales for choosing the TOEIC test to be a language exit test. Thus, the TOEIC test was selected because the policymakers felt that it assessed the language in the workplace, a TOEIC certificate would give students an employment advantage when applying for a job due to its prestige, and that employers required a TOEIC certificate. The major intended impact of the TOEIC-as-exit-test was also to motivate students to learn and make them meet the requirements of the labour market.

Besides students' future employment, other factors were taken into consideration when the decision about the TOEIC-as-exit-test was made. One of them was test quality. The majority of the policymakers were concerned with the poor language testing and assessment practices in Vietnam, so an international test, which was considered to be

prestigious and to have high quality, should be used. Among the major international English tests, TOEIC was chosen because policymakers were constrained by some other factors such as test-taking fees and test suitability to students' language ability. Two of the participants admitted that the TOEIC test was not ideal, but most suitable for the existing situation.

9.1.2 Actual impact of TOEIC on teaching and factors involved in test impact generation.

The TOEIC test obviously had an impact on teachers' teaching, but the impact varied in form and intensity across the types of course and individual teachers. The questionnaire data showed a tendency of teaching to the test among the majority of teachers. This included direct test preparation, an increase in time devoted to the tested skills and a decrease in time devoted to the non-tested skills since the introduction of the test. There was also a shift away from communicative teaching with time increased for form-focused teaching and learning activities and decreased for communicative activities compared to before TOEIC. Several factors were found to contribute to shaping this test impact on teaching. The most influential ones were test factors, especially test value, test purpose, and test pressure. One of the context factors, opportunities, was also considerably involved in the generation of test impact. Teachers' age and teaching experience also had some strongest relationships with test impact variables; however, they negatively correlated with the tendency of teaching to the test. Surprisingly, teachers' beliefs did not play an important role in creating the general patterns of test impact on teaching. Contrary to our expectation of the remarkable differences in the test impact between the large city and the province, only significant difference in direct test preparation was found, and this impact was much larger in the province.

The teacher case study confirmed that the TOEIC test impacted on all teachers; however, it also disclosed that the test influenced different courses and individual teachers in different ways. The teachers at University A, where Business English courses were taught, were much less influenced by the test than those at the other two universities. They did very little direct test preparation. Although they also paid more attention to the tested skills than before the introduction of TOEIC, the narrowing of the curriculum was not clearly evident. Their teaching methods were much more communicative in nature than the other two universities. By contrast, at Universities B and C, where TOEIC preparation courses were

investigated, the tendency of teaching to the test revealed by the case study was very strong. Time was exclusively for test preparation except for a little time for speaking presentations in university C. The teaching methods were completely non-communicative and superficial with an emphasis on forms and understanding and recognizing knowledge, not language use for communication. There seemed to be exceptions in Kim's teaching, but due to the limited observation time, the findings related to Kim's teaching were not conclusive. The case study also revealed that these differences between the two types of course were probably due to the fact that the teachers followed the textbooks very closely, and they were also likely to be influenced by the course assessment more than the TOEIC-as-exit-test. It turned out that the course factors were the dominant factors in determining the form and intensity of the test impact. Teachers' beliefs about communicative language teaching were generally overridden by them. This explains why these factors were not shown to be associated with test impact by the questionnaire results. As shown in Chapter 7, there were differences among teachers of the same course types, too. These were shown to be caused by person factors such as experiences as a language learner, personal styles, and personal beliefs about language teaching and learning. However, the case study findings were unlikely to support the negative relationships between age and teaching experience and the tendency of teaching to the test found in Phase 1. Despite their age difference, teachers of the same course type taught in very similar way.

9.1.3 Actual impact of TOEIC on learning and factors involved in test impact generation.

The findings of the study showed an obvious impact of the TOEIC test on student learning, but this impact differed across individual students and, possibly, types of course. The questionnaire results showed that, generally, the test had positive influence on student learning although the influence was not very strong. This positive impact was shown in students' motivation, their time increase for all the language skills, and the improvement of their language ability compared to high school. However, there was also a clear tendency of learning to the test. Students did direct test preparation. The time increase for the non-tested skills was considerably smaller than that for the tested skills, and the time increase for language use activities was only minimal. Students were also a little stressed and pressured, and some of them had negative attitudes towards preparing for TOEIC. The questionnaire also showed that test impact on student learning differed between the large city universities and the provincial one in most aspects mentioned above, but the degrees

of the difference were mostly small. In addition, the questionnaire revealed that the most influential factors in test impact generation were test factors, especially test value, test purposes, and test pressure. Availability of resources played the next most important role. Person factors were little involved in shaping test impact.

The case study results generally supported the questionnaire findings, but they gave a deeper insight into student learning. They not only confirmed the impact of the TOEIC test on students' motivation, but also showed that this impact operated in tandem with the impact of students' awareness of the importance of English for employment, which was very common in society. They also displayed more clearly the ways students were motivated. It is now clear that students were not only more behaviourally and emotionally engaged in their learning than before, but several students also started to show some signs of cognitive engagement, e.g., planning their learning or exerting more mental effort to understand and remember. Additionally, the case study revealed more about how students prepared for the test. They were not only involved in direct test preparation, but also in a little indirect test preparation, which included both form-focused and language use activities (the latter involved only the receptive language skills). This might explain the time increase for both types of activities found in the questionnaire data. Besides test preparation, some students did other learning activities without a test preparation purpose in mind because of their awareness of the importance of real communicative ability for employment.

The findings from the case study also disclosed important variability in test impact on learning among students. The higher ability students, whose mock test scores were higher than the universities' TOEIC benchmarks, were not affected by TOEIC as much as the lower ability students. They generally worked for the development of their communicative ability or other purposes. The lower ability students, by contrast, did not do much for communicative ability. Most of their time was for the TOEIC test and/or general English ability. Also, there was an observed difference in test impact on student learning between the types of courses. All the students at Universities B and C were involved in direct test preparation to different extents. By contrast, at University A, none of the students had started direct test preparation yet. As discussed in Chapter 7, this difference suggests that the impact of the TOEIC test on some students in the Business English courses might have been a little delayed compared to the impact on those in the TOEIC preparation courses. In

other words, course factors might have also moderated the relationship between the test and student learning.

Besides students' ability and types of course, other factors such as students' English background, their strengths and weaknesses, and their beliefs about language learning were also found to influence individual students' learning. For example, although some students also perceived grammar as useful for TOEIC, they spent less time on it because they had learned it a lot in high school. Thien did not learn pronunciation much because he thought his pronunciation was good enough. Mai learned writing more than listening because she was better at writing, and she believed it could help her gain marks in the final tests.

9.2 Discussion of Major Issues

9.2.1 Intended and actual impact, positive and negative impact.

This section examines whether the intended impact of the TOEIC used as a language exit test in Vietnamese universities was achieved. It also discusses the direction or value of the actual impact, i.e., whether it was positive or negative. As a judgement of the direction or value of test impact should be made with references to “our view of the desirable outcomes of language learning” (Bailey & Masuhara, 2013, p. 304) and specific audience (Watanabe, 2004), the impact of the TOEIC test is evaluated against different stakeholders' beliefs about the purposes of language teaching and learning at university. It is clear from the findings that the opinions about the desirable learning outcomes of all parties including policymakers, teachers, and students converged on students' ability to use English for communication in the workplace. Therefore, the impact of the test is considered positive if it positively contributed to students' achieving this ability. As the current study investigated only the test impact on teaching and learning, this section only focuses on the influence of the TOEIC test on these areas.

9.2.1.1 Test impact on teaching.

As discussed in Section 4.5, impact on teaching did not seem to be a main concern of the TOEIC-as-exit-test policy. The shared intention was to improve teaching, but what “improved teaching” meant was not always clearly stated. Out of the eight comments, only five were rather specific. Three comments were about the necessity for teachers to improve their own skills and knowledge to meet the new challenges of teaching TOEIC, e.g., to improve their own TOEIC test scores or learn how to teach test-taking strategies to

students. Two comments indicated the intended direction of change in teaching methods: making teaching communicative. As almost all of the policymakers asserted that the purpose of English teaching at university was students' communicative ability, it is assumed that "improved teaching" happened when the TOEIC-as-exit-test policy drove teachers towards communicative teaching.

The findings showed that the test impact on teaching differed a lot depending on the type of course, but it is likely that the TOEIC policy did not achieve the purpose of making teaching communicative. In TOEIC preparation courses, there was a drastic narrowing of the curriculum desirable for the development of communicative ability. Teachers almost completely ignored the non-tested skills, i.e., speaking and writing, despite the fact that all the teachers and students in the case study considered speaking vital for communication. This contraction of the teaching content was obviously linked to the test design. The participants in the case study explicitly cited the fact that the TOEIC test did not assess speaking and writing as a reason for totally ignoring them or paying only little attention to them. The narrowing of the curriculum found in this study supported the results of the majority of other studies on impact of high-stakes testing both in general education (Cruz & Brown, 2010; J. Smith & Kovacs, 2011; M. L. Smith, 1991) and language education (Alderson & Wall, 1993; Green, 2006). Reviewing 49 qualitative studies in general education, Au (2007) also suggested that "the nature of high-stakes-test-induced curricular control is highly dependent on the structures of the tests themselves" (p. 258). However, there was one difference between the current study's findings and those of others in terms of the contraction of the curriculum. In other contexts, there was a narrowing of an official curriculum, i.e., teachers chose to teach only certain areas of the official curriculum which are tested. However, in the current study, there was a contraction of a curriculum desirable for the development of communicative ability, and the contraction was decided on by policymakers at the university level. The course purpose was to prepare students for the TOEIC test, and thus, the course materials, which represented the official curriculum, covered only the two tested skills in University B and the two tested skills plus some speaking in University C.

In addition, the teaching methods became non-communicative. There was a total absence of communicative activities. No genuine communication in English happened in class. Lessons were very teacher-centred with the teacher talking most of the time and the

students only sometimes giving single word answers to the teacher's questions. Most instruction was done in Vietnamese. There was also worrying evidence that teaching became superficial aiming at forms and understanding and recognizing knowledge rather than language use. A lot of attention was paid to test-taking strategies, so that students could answer test questions correctly sometimes even without true understanding. This kind of actual impact on teaching was certainly unintended and negative. These negative effects on teaching were, again, linked to the test design. Teachers naturally aimed at teaching students to do test questions most effectively, but the multiple-choice format of the test did not require them to teach their students to perform authentic language use activities. This link between a multiple-choice test and "deskilled" teaching was also found in Smith (1991): "Because multiple-choice testing leads to multiple-choice teaching, the methods that teachers have in their arsenal become reduced, and teaching work is deskilled." Another important reason for this superficial, non-communicative teaching was the nature of the course. Test preparation textbooks did not include authentic language use activities, so teachers did not teach them. Also, the teachers in TOEIC preparation courses often complained that the high volume and advanced nature of the knowledge to teach, especially vocabulary and grammar, and the limited instructional time impeded them from organizing communicative activities and using English in class. Two of the provincial teachers also attributed the shift away from communicative teaching to the large class size at their university. Again, this negative impact of high-stakes tests on teaching methods was also discovered in other research (Au, 2007; Cruz & Brown, 2010; Diamond, 2007; Gerwin & Visone, 2006). Cruz and Brown (2010), for example, found that teachers used less student-centred activities to prepare students for Texas high stakes exams.

However, as shown in Chapter 7, the test impact varied across the types of course. There was almost a lack of test impact on teaching at University A. Although both of the teachers there reported paying more attention to listening and reading due to the introduction of the TOEIC test, observations showed that their teaching was much more communicative than their counterparts' at the other two universities. There was no clear evidence of the narrowing of the teaching content. Phuong, one of the two teachers, even said that now she spent more time on speaking than before. This difference was due to the textbook that the university adopted after the introduction of the TOEIC test. The Head of the Foreign Language Department of University A asserted that the introduction of TOEIC had made teaching in Business English classes more communicative than in the former ESP classes,

and so the test impact on teaching was positive. He also indicated that this happened due to the new textbook. This was partly true. However, the positive changes would be better attributed to the department's choice of textbooks than to the TOEIC test. The use of the new textbooks was not an inevitable consequence of the introduction of TOEIC. University C, for example, had another choice, which was using TOEIC preparation textbooks. Moreover, University A could have used Market Leader even without the TOEIC test. Although teaching at General English courses at University B was not investigated, the test may not have influenced them, either.

9.2.1.2 Test impact on learning.

Impact on student learning was clearly the major intention of the TOEIC-as-exit-test policy. The two types of intended impact most often mentioned by the policymakers were impact on students' motivation and their ability to meet the requirements of the labour market.

First of all, it is undeniable that the policy had had a positive impact on students' motivation although it was not too strong. This was evident in both Phase 1 and Phase 2 results. There were big differences in student learning compared to high school, when many of them did not learn English at all or learned only a little, did rote learning, and passed the subject thanks to the marks given "for free" by their teachers. Now, students had much better attitudes towards learning, they understood the importance of English for their future employment, they worked harder, and took a deeper learning approach. In other words, they changed both their attitudes towards English learning and their learning behaviours. These changes were definitely positive, and, in this sense, the TOEIC test had made students "truly learn" as one of the policymakers said. However, the type of motivation created by the language exit requirement was in many cases only extrinsic. Some of them learned just because of the threatening consequences of not getting the degree. Explaining what was the most important factor that motivated him to prepare for the TOEIC test, Khoi said, "I'm only afraid of not getting the degree, but... [I don't like learning English]." Most other students would prefer learning communicative English: "If the university abolished the TOEIC test, I wouldn't learn TOEIC. Instead, I would learn communicative English" (Toan). This is the reason why some teachers (Hoa and Kim) asserted that the TOEIC motivated some students but just pressured others.

Second, the TOEIC test was likely to improve students' learning outcomes. Students perceived that they had improved their language ability, especially in the language skills they spent more time on than before. However, it was also probable that their improvement was not large enough to make them meet the requirements of the labour market, that is, to use English for communication in the workplace. A lot of teachers and students were aware of this. Luong said, "I think if we learn the university programme, we only get enough to take the TOEIC test. As for communication, we probably can't speak very well." It was also believed that the TOEIC certificate with a score of 400 or 450 could not really give students advantage when they applied for a job. Toan said, "If I were an employer, when an applicant had a test score of more than 700, I would just accept his application, but... [not to mention the other rounds of assessment]." Thus, the intended impact on students' learning outcomes, to some extent, was achieved and could be considered positive. However, a lot more should be done for students to be able to use English for communication in the workplace.

There was evidence of unintended impact, too. First of all, there was generally an ignorance of the non-tested skills. As said above, compared to the fact that many students did not learn any English at school as found in the case study, the fact that students learned the tested skills was already an improvement. However, from the communicative teaching perspective, this is definitely a narrowing of curriculum.

Second, there was a superficial kind of learning that aimed at coping with the test rather than on developing communicative competence. Luong, even argued that students did not need to learn reading to prepare for the TOEIC test because

We focus on what the question asks about and look for the answers [in the text] as the teacher has taught us. We look for the place with similar words, we translate that part cursorily, and we know the answer.

This conclusion about students' superficial learning seems to contradict what was said about the influence of the test on their cognitive engagement. However, the problem is that, on the one hand, students needed to truly learn to know and understand, for instance, grammar and vocabulary, in order to pass the test, but on the other hand, they can use test-taking tips to guess answers without truly understanding the reading or listening text, and they did not need to learn to use the language for real communication, either.

Thus, what students learned and how they learned under the influence of the TOEIC test were not enough to lead to communicative ability and sometimes were detrimental to its development. The TOEIC test, on the one hand, did help improve students' learning; on the other hand, it constrained them from improving their communicative ability.

9.2.2 Discrepancy in perceptions of TOEIC validity between policymaking level and implementation level.

The high validity of the TOEIC test was one of the most common rationales for using TOEIC as a language exit test. However, there was an interesting difference in the perceptions of this validity between the policymaking level (MOET officials and Heads of Department at universities) and the implementation level (teachers and students). As found in Chapter 4, the policymakers interviewed generally referred to the validity of TOEIC as the ability of the test to measure what it actually measured due to its strictness and standardization. However, the case study showed that only a couple of students perceived the test validity that way. When discussing the validity of the TOEIC test, the majority of them had in mind its ability to assess students' overall communicative competence. Therefore, while nine out of the eleven policymakers who commented on the test validity held that the TOEIC test was valid, eight out of the twelve students straightforwardly said it was not, and one student said it was only relatively valid. All nine students cited the same reason for their opinions. They talked about the mismatch between high test scores and real communicative ability, particularly speaking ability. They generally saw the TOEIC test and communicative ability as two separate things. In their opinion, the reasons for this mismatch were the absence of the speaking skill and the multiple-choice format, which resulted in test score inflation by random guessing, the use of test taking strategies, and test preparation. Teachers were very similar to students. The case study revealed that, although the teachers seemed more cautious when judging the validity of the TOEIC test than their students were, most of them also mentioned the mismatch between test scores and overall communicative ability due to the absence of the speaking and writing skills as well as the multiple-choice format. Thus, there was a wide divergence of perceptions of TOEIC validity between the policymaking level and implementation level. However, teachers and students were not wrong. A study (Liao et al., 2010) on the correlations of test scores measured by the TOEIC Listening and Reading tests and TOEIC Speaking and Writing tests published on the ETS website showed that "each test measures distinct aspects of English language proficiency that cannot be adequately assessed by the other

tests” (p. 13.11). It suggested that test-takers should take all four tests to have a full understanding of their overall language proficiency. These results were confirmed by a later study (Liu & Costanzo, 2013) also published on ETS Website.

Due to this discrepancy in the perceptions of the test validity, the rationales for choosing TOEIC as a language exit test might not be fully justified from the implementation level’s perspective. In fact, two out the eight students who were asked about the suitability of TOEIC for graduation purposes said that it was definitely not suitable because it could not assess overall communicative ability. Three of them, who a little reluctantly agreed that it was generally suitable, expressed their concerns about the same validity issues of the TOEIC test (see Chapter 8). Cuc, for example, thought the test was not suitable to be a language exit test because it did not accurately reflect learners’ communicative ability while “the university learning outcomes are that students will have English proficiency to work and communicate.” Clearly, this student pointed out the mismatch between the test construct and the purpose of its use. From a professional point of view, this mismatch goes against the ILTA Guidelines for Practice (International Language Testing Association, n.d.), which advise that users of test results should “Make certain that the test construct is relevant to the decision to be made” (p. 5). Universities, therefore, should take into consideration the match or mismatch between the test construct and its use. They should also make clear to their students, employers, and the wider society what their language exit test could actually measure, and whether it could ensure that graduates, who have passed the test, could perform the real world tasks that they are supposed to.

On the other hand, the ILTA Guidelines for Practice also state that “The test developer’s understanding of just what the test, and each sub-part of it, is supposed to measure (its construct) must be clearly stated” (p. 1), and the test developer should “Refrain from making any false or misleading claims about the test” (p. 4). Therefore, the TOEIC test designer should make clear to students, teachers, universities, and the wider society, who will be directly or indirectly affected by the test, and what the test can measure. They should warn universities that using only the listening and reading tests cannot assure an assessment of overall communicative competence. Misleading information should be avoided. For more than 30 years, the TOEIC test has been called Test of English for International Communication. The examinee handbook for the listening and reading test (ETS, 2012b) says, “The scores indicate how well people can communicate in English

with others in business, commerce, and industry” (p. 2). However, only recently have the speaking and writing skills been included in the assessment.

In addition, as shown in the questionnaire data, there were a number of significant positive correlations between students’ perceptions of the test value and their motivation and the time increase for learning. Similarly, the more teachers believed in the test value, the more positive they felt about test preparation and engaged in behaviours that were likely to be useful for improving students’ test results. Therefore, in order for the language exit test policy to produce beneficial impact on teaching and learning, the test chosen should be perceived as good by teachers and students.

9.2.3 Differences in test impact between large city universities and provincial university.

Some studies (McCarthy, 2008) have found differences in impact of high-stakes tests on teaching due to school settings. Research into teachers’ beliefs and practices has also showed that “contextual factors can have powerful influences on teachers’ beliefs and, in effect, affect their classroom practice” (Fang, 1996, p. 53). However, school settings have rarely been an explicit research focus of impact studies compared to test factors and person factors. Therefore, one of the purposes of the current study was to explore this variability. With an expectation of significant differences between large cities and provinces in terms of teaching equipment, resources, teacher training, and student population, a large city and a province were chosen as research sites for the study. Following is a discussion of the results related to the variability in the impact of the TOEIC test on teaching and learning between the two locations.

9.2.3.1 Differences in test impact on teaching between large city universities and provincial university.

The findings revealed that the differences in the impact of the TOEIC test on teaching between the large city universities and the provincial university were not as striking as those between types of courses – Business English and TOEIC preparation. The Business English courses at University A were very different from the TOEIC preparation courses at Universities B and C in terms of teaching content and teaching methods. On the other hand, the TOEIC preparation courses at Universities B and C were very similar to each other. As analysed in Chapter 7, despite minor differences, Lan’s (University B) and Phong’s

(University C) teaching had a lot in common. Most of the time was spent on TOEIC preparation materials, and the most common activities were doing practice tests. Teachers talked a lot. They explained grammar and vocabulary and translated everything or almost everything into Vietnamese. Students, on the other hand, produced only single word answers to teachers' questions. These findings were unexpected by the researcher and different from those in McCarthy's study (2008). The different findings were due to a complex interplay of different factors.

First of all, many others' studies (McCarthy, 2008; Shohamy et al., 1996) have shown that stakes of a test are a driving force behind teachers' teaching, and differences in test stakes lead to differences in test impact. As reviewed in Chapter 2, the different stakes of the EFL and the ASL tests (Shohamy et al., 1996) made a big difference in their impact on teaching. In McCarthy's study (2008), although the stakes of the NCLB were supposed to be high for all teachers, its *de facto* stakes were different for teachers at the high-income schools and teachers at low-income schools due to the school settings, students' ability in particular. Teachers in high-income schools felt safe because their students always made Adequate Yearly Progress. On the contrary, teachers in low-income schools were under great pressure to prepare students for the test; otherwise, they might lose their jobs. This difference in stakes caused strong impact of NCLB on teachers at low-income schools, but no impact on teachers at high-income schools. However, in the current study, the stakes of the TOEIC-as-exit-test were low for teachers at both locations – large city and province. At both places, teachers were not held accountable for students' test results. Therefore, in the current study, test stakes played a much smaller role in shaping test impact than in other studies, and probably could contribute only a little to the differences between the two locations.

Second, other studies (Alderson & Hamp-Lyons, 1996; Fournier-Kowaleski, 2005; Watanabe, 1996) have showed that teacher factors played an important role in shaping test impact. The differences in their beliefs, educational background, past teaching experience, personality, and teaching styles can make them teach differently under the influence of the same test. Alderson and Hamp-Lyons (1996) found that the TOEFL influenced different teachers to different degrees and kinds. Similarly, Watanabe (1996) also pointed out the nature of the exam could not entirely explain teachers' teaching methods. Instead, the differences among the teachers could have an important influence on their teaching.

However, in the current study, teachers at the two locations generally differed very little. They were all advocates of the communicative approach. They believed that language learning was for communication, it was necessary to learn all four language skills, the communicative approach was an effective way of teaching language, and the best way to learn a language was to use it for communication. The person factors, therefore, were unlikely to contribute much to the difference between the large city and the province.

Third, the context factors in the two locations did not differ much, either. Questionnaire data showed no significant differences in the constraints and opportunities factors. Teachers perceived the impact of the time constraints and class size on their teaching in the same way. They enjoyed the same availability of resources and the freedom to use them. Observations also showed that classrooms in all the universities in this study were equipped in the same way. All the teachers said they had enough teaching materials. This similarity in these context factors seems to contradict how people often depict them. Probably the reasons for no differences were that language teaching does not usually require as much special equipment as other subjects. Also English teaching and learning materials are now widely available in Vietnamese cities. Kim, a provincial teacher, said, “Now [learning resources] are more than sufficient. The only thing is whether we want to use them or not.” There was a visible difference in the class sizes. Each class in the large city universities included a maximum of 40 students, but in the provincial university 60 to 70 students. However, as said above, the questionnaire results showed that the provincial teachers perceived the impact of class sizes on their teaching in the same way as their counterparts in the large city, so this was unlikely to be an important factor in shaping the differences between the two locations.

There was probably considerable difference in the student population between the two locations. The students in the provincial university were likely to have lower learning ability than those in the large city university as the university entrance scores of the former was much lower than that of the latter. A certain proportion of the students in the provincial university generally had worse attitudes towards learning than those in the large city university. Observations showed that in each provincial class, there were a considerable number of students who had bad attitudes towards learning. They came to class very late. In class, they did not listen or learn. Instead, they sat at the back of the class and slept or played with their mobile phones. This phenomenon was pretty rare in the large

city university, where some students might be late too, but they did not often show the “don’t care” attitudes as those in the province. This difference in students’ attitudes did not show in the student cases, probably, because the teachers only introduced students with good attitudes to the researcher. Quyen explained that she did not select the ones who had bad attitudes towards learning because she was afraid that they might not cooperate well with the researcher. However, although the teachers in University C occasionally complained about students’ attitudes, they often mentioned the impact of students’ attitudes on their mood and enthusiasm for the teaching job, not their teaching content or methods. Phong said,

...occasionally, when I have some difficulties in teaching and get weary, I feel discouraged. (...) For example, (...) departments such as Electrical Mechanics and Civil Engineering include only boys, and most of them are especially difficult to deal with. Teaching those classes is very exhausting.

Differences in students’ ability are supposed to lead to differences in the test impact on teachers’ teaching as shown in McCarthy’s study (2008). However, the teachers at the two locations did not perceive the test difficulty for their students differently. The reason was likely to be the differences in the benchmarks the universities used. The large city universities used a higher benchmark (TOEIC 450) than the provincial one (TOEIC 400). Also, the former required students to take the international test, while the latter only administered to their students an internal test, which was perceived by both teachers and students at the university as much easier than the international test. Kim, a provincial teacher said, “At the moment, we still use the internal TOEIC. The international TOEIC is very tough!” Luong, a bottom student at the same university was confident that he would pass the internal test: “I’ll probably pass because the university’s centre will be... like more lenient, it’s easier to pass. If we take the test at the international centre, it is probably difficult.” Therefore, students’ ability was not an important factor, either.

While person factors, test factors, and context factors were unlikely to influence the test impact on teaching a lot, course factors, i.e., purpose, textbooks, and mid-term/final tests, were shown to be very important. Teachers at the three universities generally followed course objectives, textbooks, and course assessment very closely. They mainly taught what was included in the textbook. Extra materials were not used or rarely used in most of the classes observed. Their methods were also determined by the textbooks. They mainly used the activities provided in them. Explaining why she spent the same amount of time on

information gaps activities before and after TOEIC, Phuong said, "...I have students do them when they are included in the textbook. I don't design more [of these]." The teachers at Universities B and C were also constrained by the nature of the course materials. Some teachers (Phong, Quyen, and Lan) said that they could not let students do language use activities when they taught grammar because of the large amount of grammar knowledge and the advanced nature of the grammar points to teach in limited instructional time. It was likely that course factors dominated and suppressed all the other factors. Consequently, the differences between locations became much less apparent than those between types of courses. Thus, this phenomenon was a result of the interaction of many different factors. If course characteristics are also considered part of school settings, then school settings do still make a difference in the test impact; however, the location is not an important factor.

9.2.3.2 Differences in test impact on learning between large city universities and provincial university.

Unlike the impact of the TOEIC test on teaching, the test impact on learning was shown by the questionnaire data to vary in many aspects across the locations. The effects of the test on the provincial students were generally stronger than on the large city students although the size of the differences was mostly small. Provincial students appeared more motivated and also more stressed than the large city students. In addition, they also perceived larger improvement in their language ability compared to high school. However, it is interesting that their time increase for the tested skills was perceived to be smaller, while their time increase for the non-tested skills was larger compared to the time changes of the large city students. On the other hand, their time increase for the language use activities was also smaller than that of the large city university students. It also seemed unusual that no significant differences were found between the two locations in direct test preparation activities and time devoted to form-focused activities.

The differences mentioned above could be explained by several reasons. First of all, on the one hand, the bivariate correlation analysis showed that students' behaviours were, to some extent, consistent with their perceptions of the test factors. Generally, the more valuable they perceived the test to be, and the more pressured they felt, the stronger the test impact on them. The more strongly they believed the test assessed the tested skills, the greater their time increase for them. Similarly, the more strongly they believed the test assessed the non-tested skills, the greater their time increase for them. On the other hand,

students at the two locations differed in their perceptions of the TOEIC test related to test value, test pressure, and test purpose-non-tested skills, with provincial students' opinions being stronger. These differences in their perceptions of the test might explain why provincial students were more motivated and stressed than the large city students. They might also explain why provincial students increased the time they spent on the non-tested skills to a larger extent than the large city students did.

There were no significant differences in students' direct test preparation activities probably because activities like practicing test questions, learning test-taking strategies, using test preparation materials, and looking for information about the test are prevalent ways of test preparation. These activities were also found to be used by students for test preparation in other studies such as Ferman (2004), Stoneman (2006), Booth (2012).

Although provincial students thought more favourably about the traditional learning methods (the importance of learning vocabulary, grammar rules, and translation), their time increase for form-focused activities did not significantly differ from that of large city students. This lack of difference was likely due to students' background. The case study showed that learning grammar was the aspect most affected by students' learning background. Many students said that they did not learn grammar much because they had learned it a lot when they were in high school.

The other factors including person factors and context factors were not likely to contribute much to the difference in the test impact on learning between the two locations because the correlations between these factors and test impacts were lower than those between the test factors and the test impacts.

There was a possibility that the test impact on learning was stronger for the province than for the large city because all the provincial students were doing a TOEIC preparation course, while approximately half of the large city students were doing a Business English course. However, the case study could not confirm or reject these differences between the two locations. Unlike test impact on teaching, test impact on learning seemed to depend a lot on students' ability, a person factor. Therefore, the small number of cases could not show clear patterns of differences between locations.

9.2.4 Mechanism of test impact on teaching and learning in the Vietnamese context.

Research has identified a number of important factors that come into play in shaping test impact on teaching and learning such as the status of the language and stakes of the test (Shohamy et al., 1996), teachers' beliefs and teaching styles (Alderson & Hamp-Lyons, 1996), school-settings (McCarthy, 2008), and the nature of courses (Read & Hayes, 2003). However, in the current study, these factors sometimes operated in different ways from those in the other studies. Also, the mechanism of test impact on teaching appeared to operate differently from that on learning. This section discusses the ways test impacts on teaching and learning were generated in the current research context.

9.2.4.1 Mechanism of test impact on teaching.

The way test impact on teaching occurred in the current study probably exhibited more differences from, than similarities with, those in other studies. Like many other studies into test impact on teaching (Li, 1990; O, 2001), this study found that the test design was among the most important factors in test impact generation. It was likely to govern both the content and methods of teaching although it had a strong influence on certain teachers, but only slightly affected others. Where the effects were strong, teachers generally spent more time on the tested skills and spent less time on the non-tested skills or ignored them completely compared to before the TOEIC policy. There was also evidence that the changes in teaching methods were associated with the test content, test tasks, test format, and test preparation materials. There was a tendency to drift away from the communicative approach to traditional teaching methods in TOEIC preparation classes as described and discussed in Section 9.2.1.1. Teachers spent more time on form-focused activities and less time or none at all on communicative activities because these were not included in the textbooks, and preparing students for a multiple-choice test did not require them to do so. They used English less than before TOEIC because they were afraid students could not understand TOEIC advanced grammar points. Teachers also changed their ways of teaching language skills to accommodate students to the test tasks and ensure they perform the best in the real test. This often led to more superficial teaching which aimed at understanding and recognizing rather than the ability to use language to express meanings. Sometimes teaching aimed at answering test questions correctly at the cost of understanding.

However, unlike many other studies such as Shohamy et al (1996) and Tsagari (2011), in the current study, the status of English and the stakes of the TOEIC-as-exit-test did not seem to play dominant roles in shaping test impact on teaching. English, apparently, has a very high status in Vietnamese society. All the participants in the study placed a great emphasis on its importance for employment. The TOEIC test was clearly high-stakes because it was used for gate keeping purposes. It would give students access to the outside world of employment or prevent them from that world depending on how well they performed in it. However, in the Vietnamese context, the test did not seem to be high-stakes for teachers. At all the three universities, teachers were not held accountable for students' test results. Instead, "the university only requires teachers to have appropriate conduct, come to class on time, finish class on time, and prepare for their lessons carefully before class" (Phong). They only slightly felt the pressure to prepare students for the test from the external expectation, which mainly came from students. They taught to the test because as teachers they naturally wanted their students to succeed. Therefore, while the teachers in Shohamy et al's study (1996) "feel that the success or failure of the students reflects on them" (pp. 309-310), the teachers in the current study were pretty calm: "I don't feel pressured (...). Generally, I also want my students to pass (...), but I only do everything I know and I can. The rest is their responsibility" (Quyen).

Additionally, the role of teacher factors in this study was very different from that in other studies. Teacher factors were found important in other studies (Alderson & Hamp-Lyons, 1996; Fournier-Kowaleski, 2005; Watanabe, 1996). Watanabe (1996) suggested that "teacher factors, such as educational background, personal beliefs and teaching experience may outweigh the possible effect of the entrance examinations" (p. 318). However, in the current study, these person factors did not seem to contribute much to the general patterns of test impact on teaching. Regardless of their personal differences, teachers who taught the same type of course were very similar to each other and were very different from teachers of the other courses in terms of content of teaching and teaching methods as discussed above. These factors only made smaller differences among teachers of the same courses. As shown in Chapter 7, Phuong and Thoai were much more communicative than the other teachers, but Phuong employed more pair work and group work and less individual work than Thoai due to her individual style. She asked her students to prepare for their lessons at home, e.g., look up new words in the dictionary and read reading passages at home, so that they could have more time for interaction with others in class to

develop their oral communication skills. Similarly, Lan used English much more than most other teachers who taught TOEIC preparation courses although she shared with other teachers the tendency of using less English in TOEIC preparation courses than in general English courses. Like other teachers, she used English less in TOEIC preparation classes because the vocabulary and sentences were much more difficult and there were far more new words. However, she still tried to use English to talk about things students could understand so that they had opportunity to listen to English.

As the test stakes for teachers were low, the course factors including course purpose, course textbooks, and course assessment became dominant and suppressed the other factors. As the findings showed, although having the same stakes at all the three universities, the test had very different impact on University A, where Business English was taught, and Universities B and C, where TOEIC preparation courses were taught. The tendency of teaching to the test was much stronger in the latter. The first reason was that teachers followed the course purpose and textbooks very closely. The fact that teachers relied a lot on textbooks was also prevalent in other parts of the world. There has been evidence that new exams often affect teachers' teaching content through teachers' heavy reliance on new textbooks (Cheng, 1997; Chen, 2002b; Wall & Alderson, 1993). The second reason was that the course assessment was more immediate than the TOEIC-as-exit-test. The fact that teachers did not feel pressured about the TOEIC test did not necessarily mean that they did not care about exams at all. In fact, all of the teachers in the case study were concerned about exams, which was clearly illustrated through the case of Phuong, Phong, and Lan in Section 7.3. However, as the TOEIC test was more distant, teachers appeared to attach more importance to course assessment, especially the final tests. Therefore, where the course factors, namely course purpose, course textbook, and course assessment, were the same as the language exit test, there was possibly no difference in the impact of the two on teaching. But where they were different, the impact of the course assessment on teaching was presumably stronger than that of the TOEIC-as-exit-test. The effect of the course factors probably overrode the influence of the person factors, including teacher's age and teaching experience. As discussed above, despite their age and teaching experience differences, the teachers of the same course type shared a lot of common features that distinguished them from the other teachers. The negative relationships between teacher's age and teaching experience and the tendency of teaching to the test found in Phase 1 could have been due to the fact that teachers at University A were

generally much older than the teachers at both Universities B and C with the age means being 42.17, 33.84, and 35.12, respectively.

Not many other studies looked into the role of test value in shaping test impact on teaching. However, the questionnaire findings suggested that it also made one of the largest contributions to test impact. Unfortunately, case study data could not clearly confirm or reject this.

9.2.4.2 Mechanism of test impact on learning.

Test impact on learning has been very little investigated compared to test impact on teaching. Therefore, not much has been written about the factors shaping test impact as well as how they interact with each other. This section discusses the mechanism of test impact on learning operating in the research context with reference to findings from other studies when possible.

The study confirmed general findings on test impact on learning that it is very complex and involves an interaction of numerous factors. However, it showed that, in the research context, certain factors were more important than others. The most important factors that determined the general patterns of the impact of the TOEIC test on students' learning were likely to be test factors (including test stakes, test value, test pressure, and test design), students' ability, and course factors.

The current findings suggested that test factors had the strongest and greatest number of relationships with the test impacts. The stakes of the test were certainly very influential. All of the students in the case study were well aware of the serious consequences of failing the TOEIC-as-exit-test, and most of them said the most important factor that motivated them to prepare for the TOEIC test was to graduate from university. The stakes of the test clearly contributed to the intensity of the test impact on learning. Other test factors such as test value and test pressure were also found to influence the intensity of the test impact. Questionnaire data showed that they positively correlated with motivation and many other types of test impact on learning. The test design, especially test purposes (skills tested) and test format, also influenced students' learning. Both the questionnaire and case study data revealed that the perceptions of the test purposes most strongly governed what students learned. A narrowing of the learning content was clearly evident. As seen in Chapter 8, students tended to spend much more time on the tested skills than the non-tested skills, and

they explicitly linked this behaviour with the test purposes. This relationship between the test design and students' content of learning was similarly evident in other studies (Li, 1990; Xie & Andrews, 2013). The case study showed that the test format and test tasks slightly influenced how students learned. Some relationships between these features of the test and students' learning strategies were found in the study. For example, some students started to skim and scan reading text or only listen to key words instead of reading every word and translating text into Vietnamese. Many students took up a superficial learning approach which aimed at answering questions correctly by using test-taking skills even at the cost of real understanding.

The current study also confirmed the dominant role of students' ability in shaping test impact on learning evident in other studies (Ferman, 2004; Pan & Newfields, 2012; Shih, 2007; Shohamy, 1993). According to other researchers' findings, the lower the students' ability is, the more intense the test impact. Ferman (2004) further suggests that the average ability student was most affected by testing with their anxiety level being the highest and their parents being much more involved with the test. Supporting those findings, this study showed that students' ability was the factor that caused great variability in the intensity of test impact on student learning. Regardless of the university where students were studying and the course they were taking there, those students whose ability had already exceeded the requirements like Hoang or Thien did not care about the TOEIC test at all or did very little for it. By contrast, lower ability students such as Toan were working intensively for the test. These findings also proved Green's (2006) model of test impact, where the intensity of test effects depends on the interaction between the importance or stakes of the test and its difficulty. When a test is high-stakes, but is easy, there will be no test impact on participants. On the contrary, test impact is the most intense when the test is both high-stakes and challenging. Person factors other than students' ability appeared to be little involved in shaping test impact on learning. As shown in the case study, students' beliefs about the purpose of language learning, the skills they should learn, and the way they should learn it often contradicted what they actually learned and how they actually learned.

As shown in Chapter 8, the nature of the course may also have contributed to the variability of test impact on student learning. All of the students in the TOEIC preparation courses were involved in direct test preparation. By contrast, all the students in the Business English courses had not started direct test preparation yet. The high ability

students did not bother to prepare for the test until closer to the test taking time. One of the low ability students felt that they would not benefit from direct test preparation due to their low ability. This suggests that test impact on some non-TOEIC students might have been delayed compared to TOEIC preparation courses. However, due to insufficient evidence, this conclusion is only tentative.

Test factors and students' ability were likely to define the general patterns of test impact on learning. The course might have contributed to those patterns, too. Other factors such as students' beliefs about language learning, their English learning background, their perceptions of the comparative difficulty of the reading and listening sections of the TOEIC test, their own strengths and weaknesses, the availability of resources, and personal circumstances only appeared to create finer variability of test impact among students who were influenced in the same pattern.

The mechanisms of test impact on teaching and on learning shared some common features. In both of them, the test design, i.e., test purposes, test content, test tasks, and test format, influenced the "what" and the "how" of their behaviours in the same directions. The test value may also have been one of the greatest contributors to the test impact generation. Additionally, their beliefs about communicative language teaching and learning were usually overridden by other factors. However, there were interesting differences between the two. First of all, the stakes of the test played very different roles in the two mechanisms. They were high for students, but low for teachers. Therefore, stakes were one of the factors that dominated test impact on learning, but they were likely to contribute little to test impact on teaching. Second, while the variability in test impact on teaching was mainly caused by course factors, the variability in test impact on learning was mainly determined by students' ability. As discussed above, teachers of Business English courses behaved very differently from teachers of TOEIC preparation courses. On the other hand, teachers of the same course behaved in a very similar way. As for students, regardless of what course they were doing, the language exit test was likely to be more important for them than the course assessment. However, the impact of the language exit test on higher ability students was much less intense than on lower ability students. Probably the test impact on average students was the most intense like in Toan's case. However, due to the small number of case study participants, conclusive evidence was not found.

Many of the specific features of the mechanism of the impact of TOEIC on English teaching and learning at the three universities may have been associated with the socio-cultural and educational context in Vietnam. This context is characterized by the high status of English in society, its importance for the economic development of the country as well as for individuals, the practice of not holding university teachers accountable for students' achievements, and teachers' heavy reliance on textbooks. As evident in Chapter 7, teachers' reliance on textbooks in turn might be attributed to the close relationship between textbook content and course assessment. Another possible reason for this habit is that, in Vietnam, schools' syllabuses and textbooks are rigidly controlled by the central authorities. All schools have to use the only set of textbooks written and published by MOET's Educational Publishing House. School teachers have to closely stick to their school's common teaching plan. University teachers certainly have more flexibility, but they might also be influenced by the prevailing habit among teachers.

9.2.5 A critical look at the TOEIC-as-exit-test policy.

The TOEIC policy in Vietnam is somewhat different from high-stakes testing in many other places in the world. As discussed in Chapter 4, it seemed to be a situational response to the MOET's requirements for publicized learning outcomes rather than a well-thought out strategy intended for curriculum innovation or educational reform. It is a result of both top-down and bottom-up decision making. While the MOET required universities to publicize their learning outcomes, it did not specify the use of any language tests as a measurement of the outcomes. Universities themselves decided on whether to use a language exit test or not, and which test to use for the purpose. Some changes had been made in line with the introduction of the TOEIC-as-exit-test. For example, universities had modified their English programme to suit the language exit test. These included adjusting the instructional time for the English subjects and changing the course materials and course assessment. Certain universities had carefully assessed students' proficiency to establish the most suitable benchmarks for their students. However, the TOEIC policy did not seem to have clearly formulated objectives. Except for the intentions to motivate students to learn and make them meet the requirements of the labour market, which were shared by the majority of the policymakers, other intended impacts were often vaguely stated. For example, only two out of the five policymakers who said they wanted to use the TOEIC test to influence teaching said they intended to make teaching more communicative. The other three only talked about improved teaching in general. While

other high-stakes testing policies around the world often directly target instruction to bring about improved learning, the TOEIC policy seemed mainly to be aimed at students.

As for the mechanism of change, the common idea was to use the TOEIC test as a language hurdle for students to overcome and consequently motivate them to learn. Except for that general idea, there seemed to be a lack of specific assumptions about the linkages between the use of the test and the intended outcomes, i.e., what aspects of the test would influence teaching and learning, how they would influence the “what” and the “how” of teachers’ and students’ behaviour, and what factors might assist or hinder the intended impact of the policy. There did not appear to have been thorough analysis of the Antecedents, Process, and Consequences as suggested by Wall (2000), either. Students’ and teachers’ viewpoints might not have been taken into account. Although these stakeholders generally supported the policy, they often preferred a test that would really assess students’ communicative competence.

Due to the lack of clearly stated objectives, an analysis of the Antecedents component of the implementation process, and a model of change which showed the logical paths from the test use to the intended outcomes, the TOEIC policy met with several issues. First of all, there was a lack of a necessary alignment among the language teaching goals, teaching practice, and testing. At University A, there was an agreement between the course and the language teaching goals, but these two did not go in line with the language exit test. University B had the same problem as University A with their first three General English courses. However, for their TOEIC preparation course, the teaching goals did not match both the course and testing. At university C, the language teaching goals were different from both all the TOEIC preparation courses and the language exit test. Second, the test selected to be the language exit test was not an authentic or performance-based test as suggested by researchers.

The issues mentioned above, unfortunately, have led to a negative unintended impact on teaching and learning as discussed in Section 9.2.1. First of all, there was a narrowing of the desirable curriculum. There was also evidence of a detrimental impact of the test on teaching and learning methods, which was contrary to the policymakers’ intentions. In TOEIC preparation classes, teaching and learning shifted away from the communicative approach and became superficial with a focus on forms and remembering and recognizing knowledge rather than language use. This unintended impact was shown to be linked to the

test design, i.e., the test purpose, test tasks, and test format. This confirmed Smith's (1991) findings: "Because multiple-choice testing leads to multiple-choice teaching, the methods that teachers have in their arsenal become reduced, and teaching work is deskilled" (p. 10).

In addition, although the tested skills were endorsed by students as necessary, and the test was generally considered doable and somewhat valuable as Xie and Andrews (2013) suggested, most students were concerned about the test validity and found preparing for TOEIC not very useful for their ultimate goal – developing their communicative competence, especially the speaking ability.

However, the TOEIC-as-exit-test also undeniably brought about some positive effects on student learning. Many students were obviously motivated by the test. The majority of them became aware of the importance of English, changed their attitudes towards learning, and worked harder or "truly learned" as one of the policymakers said. Their language ability was also perceived to be slightly improved. Compared to high school, where a lot of students did not pay attention to English; listening was completely or almost completely absent; and teachers gave students marks for free so that they could move to the next grade, these changes were real improvements.

While most other studies on high-stakes tests (M. L. Smith, 1991; Tsagari, 2011) found high levels of stress and pressure on teachers and students, this study showed that this level was low for teachers and not too high for students. The teachers in Israel were driven to teach to the EFL test because they felt that "the success or failure of their students reflect[ed] on them" (Shohamy et al., 1996, pp. 309–310). However, the teachers in the current research context only needed to complete the curriculum as specified by their department: "No pressure at all! Very relaxing! As long as I teach the right curriculum, at the right pace, and the right number of hours as assigned by the Department" (Kim). Students were generally not too stressed because the benchmarks were selected by their own university to suit average students. All of the students in the case study were certain that they could achieve the benchmark, no matter how difficult they found it. The provincial students, whose learning ability might be lower than the large city university students, also seemed relaxed probably because their benchmark was lower, and the internal TOEIC test their university used was perceived to be much easier than the international one. Luong said, "I'll probably pass because the university's centre is... like more lenient with students. It's easier to pass. If we take the test at the international centre,

it is probably difficult.” Nevertheless, the stress level, of course, depended on students’ ability. As found in Chapter 8, high ability students like Hoang and Thien did not worry about the test at all, but the lowest ability students like Mai were highly stressed.

Unlike high-stakes testing in many other contexts, the TOEIC used as a language exit test in Vietnamese universities was generally welcomed by the majority of students and teachers. Despite some validity and usefulness issues, both teachers and students thought that the TOEIC policy was necessary and generally useful. The main reasons for these attitudes were that they believed the test, to a certain extent, helped enhance students’ motivation and improve their language ability, which was considered to be beneficial for their future employment. As pointed out in Chapter 1, with the government’s open-door policy, Vietnam has been more and more deeply integrating into the world’s economy, and people have been really keen on working for large international companies, where they can earn much higher salaries. English, therefore, enjoys an extremely high status in society. Due to this perception of the importance of English, teachers and students believed the TOEIC policy benefited students despite their general concerns about the validity and purposes of the test as well as their negative opinions about preparing for TOEIC and its usefulness for developing communicative competence. Toan, for example, considered the benchmark very necessary. In supporting the policy, he said:

Essentially, the university set up this benchmark to do good for students because, although students from University B are known for being good at technology, they can only work for Vietnamese people, they can’t work for foreigners. The purpose of this policy is for students to understand who they are and to determine who they will work for later (Toan).

In a social context, where English is seen as very important, and students’ English proficiency is generally low, it is not surprising that the use of such a test is generally thought of favourably. Compared to before the TOEIC policy, the test, at least, has brought about some positive, although not ideal, changes in students’ learning. It makes them learn more and improves their English ability, although still not to the extent that they can use the language for real communication.

From the viewpoint of ideal language teaching and learning, the test apparently had more negative than positive impact on teaching and learning. The TOEIC policy had not achieved the purpose of language learning for communication yet. However, considering the Vietnamese context, especially students’ limited learning background, the changes in

students' motivation and achievements were already an improvement. This suggests that a language exit test is necessary in the Vietnamese context, at least, for the time being. However, modifications needed to ensure the real success of the policy. Implications of the study for policymakers regarding these modifications are presented in Chapter 10.

9.3 Chapter Summary

This chapter gave a short summary of the research findings and discussed the major issues emerging from the data. First, the TOEIC policy clearly had certain positive effects on student motivation and their achievement. However, it also brought about unintended, negative effects such as the narrowing of teaching and learning content, non-communicative, superficial teaching methods, and superficial learning strategies. A lack of impact was also observed in certain cases. Next, the chapter discussed the discrepancy in the perceptions of the validity of TOEIC between the policymaking level and the implementation level. It argued that policymakers should justify their choice of the language exit test by taking into consideration the degree the test construct matches the purpose of its use and take responsibility to make clear to students, teachers, employers, and the wider society what graduates who achieve the required scores can do with their English. Then the chapter examined the differences in test impact between the large city universities and the provincial one. It suggested that the difference in the location was not as important as the difference in the type of course in the case of test impact on teaching and individual students' language ability in the case of test impact on learning. After that, the chapter dealt with the mechanism of test impact on teaching and learning in the Vietnamese context. It discussed the contribution of influencing factors to shaping the test impact and pointed out how and why certain factors operated differently from other research contexts. The last section of the chapter addressed the TOEIC-as-exit-test policy in Vietnam. It discussed its success and issues that remained to be dealt with.

Chapter 10: CONCLUSION

10.1 Purpose, Methods, Findings, and Significance

This study was set out to examine the impact of the TOEIC test on teaching and learning in Vietnamese universities and explain the mechanism by which it operated. It intended to achieve the research objectives through triangulation of quantitative and qualitative methods. First, two sets of questionnaires were used to identify the major patterns of test impact on teaching and learning. Then a case study was conducted to gain more insight into the phenomenon. Although the main focus of the study was teachers and learners, the policymakers' perspectives' on the TOEIC policy were also taken to obtain a more holistic understanding of the issues and better evaluate the success of the policy.

Overall, this study has succeeded in depicting the general patterns of the impact of the TOEIC test on teaching and learning in the three universities as well as its variability. It showed a general tendency of teaching and learning to the test, but the extent and form of the test impact varied greatly across different groups of teachers/students due to an interaction of various factors. It has also been able to identify the major factors involved in shaping the test impact and revealed how they interacted with each other in the research context.

The study has revealed many insights into the phenomenon of test impact in the Vietnamese context. On the one hand, it confirmed many findings from other studies, and, on the other hand, disclosed features specific to the Vietnamese context. It supported other researchers' findings about the complexity of test impact (Alderson & Hamp-Lyons, 1996; Spratt, 2005). It also confirmed Hughes' (as cited in Bailey, 1996) ideas about the mechanism of test impact by showing a clear connection between the participants' perceptions of the test design, especially the purposes and format of the test, and what and how they taught/learned. It again demonstrated the role of learners' ability in determining the intensity of test impact on learning mentioned by other researchers (Ferman, 2004; Green, 2007a; Pan & Newfields, 2012). It also confirmed the beliefs that the value or the positive/negative direction of test impact not only depends on test characteristics. The findings suggested that the TOEIC test had brought both positive and negative changes to the classroom although teachers and students were critical of some aspects of the test design, i.e., the lack of assessment of the speaking skill and the multiple-choice format.

However, the study showed that in the current research context, several influencing factors played different roles from those in other contexts, and consequently the mechanism of test impact also operated in a distinctive way. While many other studies show similar roles of the stakes of the test in question in shaping test impact on teaching and on learning, in this study, test stakes were likely to dominate student learning, but have limited effect on teachers' teaching. A probable reason was that the TOEIC test was high-stakes for students, but it was distant and had no consequences for teachers. As a result, the mechanism of test impact on teaching operated differently from that of test impact on learning. In the case of test impact on teaching, the influence of the test may have been first filtered through the course factors, i.e., course purpose, textbooks, and assessment. When the course factors were consistent with the test design, the test impact on teaching was very strong. When they were different from the test design, they became the main factors that governed teachers' teaching, and the test design only played a subservient role. It is also interesting that while other studies revealed the great importance of person factors such as teachers' beliefs about language teaching and learning, their teaching styles, and their teaching experience (Alderson & Hamp-Lyons, 1996; Watanabe, 1996), this study showed that these factors were only of secondary importance. For the teachers who taught the TOEIC preparation courses, what they did under the influence of the test usually contradicted their beliefs about communicative language teaching. Figuratively speaking, these factors were perhaps only the secondary filter that created variability among teachers of the same teaching tendency.

Concerning student learning, the influence of the TOEIC test appeared to have been first filtered through students' ability. A secondary filter may have been the course nature, students' beliefs about language learning, their learning background, their strengths and weaknesses, their personal circumstances, and the availability of resources. It is interesting that in many cases students' beliefs about communicative language learning were also likely to be overridden by the test factors as teachers' beliefs were when teachers taught TOEIC preparation courses.

Also, contrary to expectations, school settings did not seem to play a major role in shaping the test impact, especially on teaching. The case study showed that the teachers at Universities B and C taught in similar ways. The questionnaire findings suggested that test impact on learning at University C was a little stronger than at Universities A and B in

several aspects, but the case study showed that when the course factors were excluded, no clear differences were seen between University B and University C. One possible reason was that, the test impact on student learning was, to a large extent, moderated by students' ability, a person factor, so the small number of participants in the case study may have prevented the researcher from obtaining conclusive evidence.

The findings also revealed some issues related to the use of the TOEIC-as-exit-test. Compared to high school, the test was likely to have brought about some positive changes in student learning, e.g., in their motivation and achievement. However, from the point of view of communicative language teaching, a lot of issues remained to be dealt with. Those were the narrowing of teaching and learning content; the superficial, non-communicative features of teaching and learning; the validity concerns; and issues related to the planning and implementation of the language exit test policy. Many of the features of the impact of the TOEIC test on teaching and learning at the three universities appeared to be determined by the socio-cultural and educational context in Vietnam.

The use of both quantitative and qualitative methods in this study allowed the researcher to see not only the general patterns, but also subtle variability among individual teachers and learners as well as the factors causing this variability. This use of mixed methods has assisted in more fully uncovering how the influencing factors interacted with each other. It has also enabled her to discover new angles, e.g., the role of the course factors in teachers' teaching, which would have been omitted if only examining the questionnaire data in the way planned. This has again confirmed the value of using a mixed methods design in order to thoroughly understand a phenomenon.

This study also has practical significance. The findings on the actual test impact, both positive and negative, and the way test impact operates in the research context could give policymakers empirical evidence to make more informed decisions on language teaching and assessment policies. The study supported previous researchers' arguments that in order to use high-stakes tests to bring about beneficial effects on teaching and learning, there should be an alignment among teaching goals, teaching activities, and testing. Also, a model for the desirable change should be considered when the policy is made, and the policy should be implemented accordingly. Attention needs to be paid to the Antecedents, Process, and Consequences of the model of the implementation process.

10.2 Limitations

As validity and reliability are the most important research qualities, the best efforts were made to ensure them. However, this study could not avoid some limitations due to practical constraints. Its biggest limitation was the lack of a baseline study to investigate how teachers taught and how students learned without the TOEIC policy. To find out the influence of the TOEIC test on teaching/learning, the questionnaire and case study participants were asked to compare their current teaching/learning with that in non-TOEIC classes in the case of the teachers and in high school in the case of the students. Basically, the conclusions about the changes in the participants' behaviours were made based on their self-report data. According to Argyris and Schön (1978), people's espoused theory may not be compatible with their theory-in-use. What they tell others they would do in a given situation may not match what they actually do with or without them being aware of the differences between the two. The reason for the lack of a base line study was that it was impossible for the researcher to find universities that were about to implement the TOEIC policy to carry out such a study. Additionally, although it is common practice for teachers in Vietnam to teach different types of courses at different institutions at the same time, observing the teacher cases in their concurrent non-TOEIC classes were out of the question. Most of them did not teach non-TOEIC classes anymore at the time the research was being conducted. Also, as this study was a one-person PhD project, observing and interviewing seven teachers in their regular classes and working with twelve students over a three-month period was already a heavy workload for the researcher.

However, the impact of this drawback was, to a certain extent, diminished by some strategies. First, close rapport with the participants was established, so that they felt safe and had no reasons for not telling the truth. This was done through a commitment to confidentiality of information and a friendly, respectful, and non-judgemental attitude towards the participants. Second, there was a triangulation of data collected by different methods. The questionnaire findings were verified by the in-depth interviews, which were supposed to be generally trustworthy due to the rich, specific, and detailed information that the participants provided. Although there were no observational baseline data on how teachers taught in non-TOEIC classes, the observations in their current TOEIC/TOEIC oriented classes showed what they actually did under the influence of the test. The observation findings were highly consistent with the teachers' self-report data about their current teaching. Similarly, students' records of their actual learning in journal sheets

generally matched their responses to the interviews. The observed contrast between Business English and the TOEIC preparation classes, which were of a very different nature, gave sufficient proof for the impact of the TOEIC test on teachers' teaching. The Business English course, to some extent, could be considered a kind of baseline for comparison.

The second limitation of the study was a lack of peer checking of the qualitative data analysis. Due to the lack of time and resources, inter-coder reliability was not established. However, the coding was done thoroughly in a reiterative process by the researcher herself. Other measures such as triangulation; member checking; and thick, rich description were also used to ensure the credibility of the qualitative data. The fact that the participants made very few corrections to the summaries and transcripts sent to them for member checking (usually for reasons other than the inaccuracy of the summaries and transcripts) suggested that the conclusions obtained were generally credible. Also, concerning the case study analysis, more weight was given to the qualitative approach than the quantitative one, so inter-coder reliability was not too vital.

In addition, there were some issues with the questionnaires. To enhance their reliability, parallel forms of some items were used. For example, the test difficulty was measured by three parallel forms: *The TOEIC test is easy for me*; *the listening section of the TOEIC test is difficult for me*; and *the reading section of the TOEIC test was difficult for me*. However, it seemed that a small number of students did not pay attention to the reverse wording of these forms, which resulted in their contradictory answers to them. Probably due to this issue, the internal consistency of the test difficulty variable was a little below .7 although the component items (parallel forms) still loaded together. Luckily, this issue only occurred with this variable.

Several composite variables in the questionnaires retained for analysis did not completely meet the conventional criteria. There were four variables consisting of only two items. Ten of the composite variables had an alpha value which was a little lower than the conventional value of .7 (see Table 3.11 and Table 3.12). However, it is argued that these variables were still suitable for analysis because the items making up the variables conceptually belonged to each other. Also, they accounted for only a small percentage of the total number of variables, and many of them were not the ones that contributed the most to the main findings of the thesis.

Another issue related to the data collection instrument was the journal sheet design. First, there should have been boxes for students to tick the reasons why they learned the language skills and did the learning activities, e.g., TOEIC, TOEIC and other, and other. The journal sheet only provided spaces for students to write whatever reasons they had for their behaviours. Therefore, the information provided was often not clear enough to categorize or not usable. Consequently, it took a lot of time to analyse students' responses and ask them for clarifications. In a few instances, the information had to be coded as unclear. Second, responding to Question 3, some students seemed to exclude homework. The example included in the journal sheet to guide students might have misled some of them into including only language use activities although they were instructed to record whatever activities they did to learn English. As a result, the journal sheets might have not completely reflected the activities that the students actually did. Students' journal sheets should have been analysed during the piloting process to find out these flaws before the data collection. Nevertheless, the most important information was still captured, and it was shown to be generally consistent with what students reported in the interviews.

Additional analysis of variance for the three universities and t-test for the two types of courses (Business English and TOEIC preparation) should have been included into the thesis to show more clearly the differences among the three universities and the role of the course factors in shaping the test impact. If this had been done, more solid conclusions about the patterns and mechanism of test impact could have been made. However, due to the time constraint and the word limit of the thesis, this was not possible.

The case study students from University A had not been involved in direct test preparation yet or very little involved in it. Therefore, there was limited information about how students from this university directly prepared for the test, and whether they would do it in the same way as students from the other universities. These findings also suggested that compared to the TOEIC preparation courses, the Business English course seemed to cause a delay in the test effects on student learning. This might not be true for the whole student population of the university because the number of the case study participants was small, and, consequently, the similarity in their behaviours may have happened by accident. However, this was likely to be true for at least some students.

Also, those students in University C who were labelled "bottom" by their teachers might not be true bottom students. The teachers at University C did complain about students who

had bad attitudes towards learning and/or were desperate about the TOEIC benchmark. The classroom observations also showed that there were students who seemed to be very indifferent to learning. However, all the student participants seemed to have good attitudes towards learning and were confident that they would pass the test. Therefore, the test impact on real bottom students at University C might not have been captured.

Finally, it is necessary to note that the findings might not be generalizable to all the other universities in Vietnam. The study involved only three large universities among the hundreds of them, but Vietnamese universities differ widely in terms of size, facility, training programmes, and teacher and student populations.

10.3 Implications

10.3.1 Implications for test developers.

The findings from this study have shown that important decisions on high-stakes testing policy can be made based on the reputation and perceived quality of a test, especially its validity. However, different groups of stakeholders' opinions about the quality of a test might not converge as shown in the case of TOEIC, and consequently a language assessment policy might not be justified. Misuse of a test may also produce detrimental impact on teaching and learning. Therefore, to prevent test misuse, test developers need to make clear to test users, test-takers, and the wider society what their tests can measure, how accurately they can measure the constructs they intend to measure, and how test scores can be interpreted. Test-takers and test users then can make informed decisions about what test to use and how to use it.

10.3.2 Implications for policymakers.

This study has suggested that, in the current situation, a language exit test is necessary for universities in Vietnam, and it is possible to use such a test to bring about positive changes to language teaching and learning. However, some issues should be carefully dealt with to ensure its success.

Implications for universities

First of all, the principle that testing should be aligned with teaching activities and teaching goals needs to be observed. When the goal of language teaching at universities is students' communicative ability, then the teaching and the language exit test should also be

communicative. The TOEIC, unfortunately, was not perceived by the majority of teachers and students as a test for communication despite what its name suggests. At all three universities, the teaching goal was mismatched with the teaching activities and/or testing. The practice of teaching TOEIC preparation courses for the most part of the English programme at university also seemed questionable because, as the findings showed, the teaching activities in these courses were very distant from being communicative. Therefore, authentic, performance-based tests should be used as a language exit test instead of multiple-choice tests.

Second, an authentic, performance-based test still cannot ensure success of the language exit test policy. Other factors such as sociocultural and educational context, stakeholders' concerns, factors that facilitate or hinder the implementation of the testing policy should be taken into consideration. This could be achieved by applying a model of implementation, which includes three components: Antecedents, Process, and Consequences (Wall, 2000). A model of the influences of high-stakes testing on teaching and learning, which shows the linkage between a test use and improved instruction, should also be used (D. W. Chapman & Snyder Jr, 2000). This model could help clarify how the implementation process would occur and what should be taken into account to facilitate it.

Third, the findings from this study have suggested that in the Vietnamese context, factors such as textbooks, course assessment, and teachers' and learners' perceptions of test validity and test usefulness were likely to have the strongest influence on shaping the test impact. Therefore, in addition to using a performance-based test, these factors should be taken into consideration to fortify the positive impact of a language exit test. For example, a textbook and course assessment that are compatible with the communicative teaching approach would help drive teachers to teaching communicatively. A test that is perceived by teachers and learners as valid and useful would have stronger impact on their teaching/learning.

Finally, universities should conduct a needs analysis to know what the actual demands for English at businesses are, which language skills are needed, what language requirements for recruitment are, and how useful language certificates are for students' job applications. The result of a needs analysis would enable them to make appropriate decisions about language teaching and assessment. For example, they would be able to construct their

syllabus with appropriate teaching objectives, instructional time, textbooks, and teaching methods. They can also select a suitable language test to use as their exit test.

Implications for MOET

It is clear from the study that universities experienced some constraints when they made decisions about a language exit test. They had to consider a number of factors such as students' ability, test-taking fee, test difficulty, test quality, and resources for designing in-house tests. None of the existing tests would completely meet their needs. As many participants said, although IELTS and TOEFL can assess test-takers' language ability more comprehensively, they are too difficult for students, and the test fees are too high. The Vietnam National Test, on the other hand, has been shown to be unreliable and not valid. It is generally not accepted by businesses and society any more. In-house tests also involve validity and reliability issues due to teachers' lack of expertise in test design and teachers' leniency. One possible solution could be developing a Vietnamese test, which could have a lower testing-taking fee and a more suitable purpose and assess more appropriate skills than international tests. However, due to the lack of resources and expertise, it would be impossible for an individual university to develop a high quality test. Also, the lesson from the Vietnam National Certificate test has shown that it is hard to keep the test as strict and standardized as an international test if it is designed and administered by an individual university. Therefore, this could only be done at the central level, where concerted effort can be made. This issue seemed to have been addressed in the National Foreign Languages 2020 Project. One of the project objectives was to establish a language testing centre that would produce tests for different training levels in Vietnam. It is now urgent to accelerate this process.

10.3.3 Implications for teachers and learners.

This study has also revealed that most teachers and students were well aware of the drawbacks of the TOEIC policy. Although they generally welcomed the policy, accepted the test considering it most suitable among the available international tests, and saw the benefit of a language exit test to student learning, they still pointed out its lack of ability to assess overall communicative ability. They were also aware of the fact that only learning to the test could not help students develop that ability, which they needed for their future employment. Most of them preferred a test which would include all four language skills, especially speaking because they believed speaking was the most important skill for

communication. Therefore, teachers and learners should raise their voices about students' needs and issues related to the TOEIC policy, so that policymakers can make appropriate modifications to their policy.

10.4 Recommendations for Further Research

This study has been only exploratory in nature. There should be other studies that aim at confirming the roles and interactions of the factors involved in shaping the impact of the TOEIC test on teaching and learning which were found. Structural equation modelling could be used for this purpose. It would help confirm the degree of contribution of the influencing factors to the test impact and their order of influence suggested by the case study findings.

As students' ability was shown to be a factor that determined the major patterns of test impact on student learning, future surveys should include a measurement of this factor and make it a focus of analysis in order to obtain more evidence for the mechanism of test impact on learning.

Research needs to be conducted with employers to find out their real needs and requirements for English ability and language certificates. This would give policymakers necessary information to make decisions about their language teaching and assessment policies.

In brief, high-stakes testing has never been a panacea for all educational problems. A lot of effort needs to be made concerning other aspects of education such as teacher training, syllabus, classroom assessment, and environment for language practice. However, the research findings have confirmed the possibility of using a high-stakes test to improve instruction on the condition that there should be careful planning and implementation. This study has highlighted the importance of further research to better understand the phenomenon of test impact and make testing beneficial for language teaching and learning.

Bibliography

- Alderson, J. C. (2004). Foreword. In L. Cheng, Y. J. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods*. Mahwah, N.J.: Lawrence Erlbaum.
- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13(3), 280–297. doi:10.1177/026553229601300304
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115–129. doi:10.1093/applin/14.2.115
- Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, 10, 18.
- Andrews, S. (1994). Washback or washout? The relationship between examination reform and curriculum innovation. In D. Nunan, R. Berry, & V. Berry (Eds.), *Bringing about change in language education*. Department of Curriculum Studies, University of Hong Kong.
- Andrews, S. (2004). Washback and curriculum innovation. In L. Cheng, Y. J. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods*. Mahwah, N.J.: Lawrence Erlbaum.
- Andrews, S., Fullilove, J., & Wong, Y. (2002). Targeting washback - a case study. *System*, 30(2), 207–223. doi:10.1016/S0346-251X(02)00005-2
- Anzul, M., Ely, M., Freidman, T., Garner, D., & McCormack-Steinmetz, A. (1991). *Doing qualitative research: Circles within circles*. London: Routledge.
- Apichatrojanakul, P. (2011). The washback effects of the TOEIC examination on the teachers and students of a Thai business school. *Language Testing in Asia*, 1(1), 62.
- Argyris, C., & Schön, D. A. (1978). *Organizational learning: A theory of action perspective*. Addison-Wesley.

- Ary, D., Jacobs, L. C., & Razavieh, A. (2009). *Introduction to research in education*. Belmont, CA: Wadsworth.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258–267.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257–279.
- Bailey, K. M., & Masuhara, H. (2013). Language testing washback: The role of materials. In B. Tomlinson (Ed.), *Applied Linguistics and Materials Development* (1st edition.). London; New York: Bloomsbury Academic.
- Bekis, W. (2008). *The impact of the No Child Left Behind Act on New Mexico public schools on the Navajo Reservation*. Arizona State University, United States - Arizona.
- Bell, J. (2005). *Doing your research project a guide for first-time researchers in education, health and social science*. Maidenhead, England; New York: Open University Press.
- Berwick, R., & Ross, S. (1989). Motivation after matriculation: Are Japanese learners of English still alive after exam hell? *JALT Publications*, 11(2), 193–2010.
- Bickham, T., Burns, P., & Monahan, D. (2001, June). *Emphasis placed on current assessment procedures affects the way teachers teach*. Retrieved from <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED458632>

- Booth, D. K. (2012). *Exploring the Washback of the TOEIC in South Korea: A sociocultural perspective on student test activity*. ResearchSpace@ Auckland. Retrieved from <https://researchspace.auckland.ac.nz/handle/2292/19379>
- Braun, H. (2004). Reconsidering the impact of high-stakes testing. *Education Policy Analysis Archives*, 12, 1.
- Brindley, G. (2001). Outcomes-based assessment in practice: Some examples and emerging insights. *Language Testing*, 18(4), 393–407. doi:10.1177/026553220101800405
- Carnoy, M. (2005). Have state accountability and high-stakes tests influenced student progression rates in high school? *Educational Measurement: Issues and Practice*, 24(4), 19–31. doi:10.1111/j.1745-3992.2005.00020.x
- Chapman, D. W., & Snyder Jr, C. W. (2000). Can high stakes national testing improve instruction: Reexamining conventional wisdom. *International Journal of Educational Development*, 20(6), 457–474. doi:10.1016/S0738-0593(00)00020-1
- Chapman, M. (2004). Voices in the Field: An Interview with Kazuhiko Saito. *Shiken: JALT Testing and Evaluation SIG Newsletter*, 8(2), 8-11.
- Chapman, M. (2005). TOEIC & TOEFL: A Partnership of Equals? In K. Bradford-Watts, C. Ikeguchi, & M. Swanson (Eds.), *JALT2004 Conference Proceedings* (pp. 1190–1195). Tokyo.
- Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, 11(1), 38–54. doi:10.1080/09500789708666717
- Cheng, L. (1998). Impact of a public English examination change on students' perceptions and attitudes toward their English learning. *Studies In Educational Evaluation*, 24(3), 279–301. doi:10.1016/S0191-491X(98)00018-2

- Cheng, L. (1999). Changing assessment: Washback on teacher perceptions and actions. *Teaching and Teacher Education*, 15(3), 253–271. doi:10.1016/S0742-051X(98)00046-8
- Cheng, L. (2003). *Looking at the impact of a public examination change on secondary classroom teaching: A Hong Kong case study*. Retrieved February 1, 2011, from http://www.eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_ERICExtSearch_SearchValue_0=EJ672338&ERICExtSearch_SearchType_0=no&accno=EJ672338
- Cheng, L. (2004). The washback effect of a public examination change on teachers' perceptions toward their classroom teaching. In L. Cheng, Y. J. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods*. Mahwah, N.J.: Lawrence Erlbaum.
- Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng, Y. J. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods*. Mahwah, N.J.: Lawrence Erlbaum.
- Cheng, L., Watanabe, Y. J., & Curtis, A. (Eds.). (2004). *Washback in language testing: Research contexts and methods*. Mahwah, N.J.: Lawrence Erlbaum.
- Chen, L.-M. (2002a). *Taiwanese junior high school English teachers' perceptions of the washback effect of the Basic Competence Test in English*. The Ohio State University, United States -- Ohio.
- Chen, L.-M. (2002b). Washback of a public exam on English teaching. Retrieved from <http://eric.ed.gov/?id=ED472167>
- Choi, I.-C. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing*, 25(1), 39–62. doi:10.1177/0265532207083744

- Cohen, J. (1969). *Statistical Power Analysis for the Behavioral Sciences*. New York and London: Academic Press.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education*. London; New York: Routledge.
- Collins, S., Reiss, M., & Stobart, G. (2010). What happens when high-stakes testing stops? Teachers' perceptions of the impact of compulsory national testing in science of 11-year-olds in England and its abolition in Wales. *Assessment in Education: Principles, Policy & Practice*, 17(3), 273 – 286.
- Cotton, D., Falvey, D., & Kent, S. (2012). *Market Leader Pre-intermediate* (3rd Revised edition.). Pearson Academic.
- Creswell, J. W. (2013). *Qualitative inquiry and research design: choosing among five approaches* (3rd ed.). Los Angeles: Sage Publications.
- Creswell, J. W., Hanson, W. E., Plano, V. L. C., & Morales, A. (2007). Qualitative research designs: Selection and implementation. *The Counseling Psychologist*, 35(2), 236–264. doi:10.1177/0011000006287390
- Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice*, 39(3), 124–130. doi:10.2307/1477543
- Cruz, A., & Brown, M. S. (2010). Impact of the accountability system on perceptions and practices of South Texas elementary school teachers. *Research in the Schools*, 17(1), 53–63.
- Cunningham, C. R. (2002). *The TOEIC test and communicative competence*. University of Birmingham. Retrieved from <http://www.bhamlive.bham.ac.uk/Documents/college-artslaw/cels/essays/matefltesldissertations/Cunndiss.pdf>

- Diamond, J. B. (2007). Where the rubber meets the road: Rethinking the connection between high-stakes testing policy and classroom instruction. *Sociology of Education*, 80(4), 285–313.
- Duff, P. (2008). *Case study research in applied linguistics*. Lawrence Erlbaum Associates.
- Dunne, M., Pryor, J., & Yates, P. (2005). *Becoming a researcher: a companion to the research process*. Open University Press.
- ETS. (2012a). Examinee Handbook Speaking & Writing. Retrieved June 1, 2015, from http://www.ets.org/Media/Tests/TOEIC/pdf/TOEIC_Speaking_and_Writing_Examinee_Handbook.pdf
- ETS. (2012b). TOEIC Examinee Handbook - Listening & Reading. Retrieved March 29, 2013, from http://www.ets.org/toEIC/listening_reading/about?WT.ac=toeichome_listeningreading_121127
- ETS. (n.d.-a). For workplace. Retrieved December 27, 2014, from <https://www.ets.org/toEIC>
- ETS. (n.d.-b). The TOEIC® Tests — the global standard for assessing English proficiency for business. Retrieved March 29, 2013, from http://www.ets.org/toEIC/succeed?WT.ac=toeichome_succeed_121127
- ETS. (n.d.-c). TOEIC Listening and Reading Test: About the Test. Retrieved December 27, 2014, from https://www.ets.org/toEIC/listening_reading/about?WT.ac=toeichome_listeningreading_121127
- Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational Research*, 38(1), 47–65.

- Faulkner, S. A., & Cook, C. M. (2006). Testing vs. teaching: The perceived impact of assessment demands on middle grades instructional practices. *RMLE Online: Research in Middle Level Education*, 29(7), 1–13.
- Ferman, I. (2004). The washback of an EFL national oral matriculation test to teaching and learning. In L. Cheng, Y. J. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods*. Mahwah, N.J.: Lawrence Erlbaum.
- Flyvbjerg, B. (2011). *The Sage handbook of qualitative research*. (N. K. Denzin & Y. S. Lincoln, Eds.) (4th ed.). Thousand Oaks: Sage.
- Fournier-Kowaleski, L. A. (2005). *Depicting washback in the intermediate Spanish language classroom: A descriptive study of teacher's instructional behaviors as they relate to tests*. The Ohio State University, United States -- Ohio.
- Fraenkel, J. R., & Wallen, N. E. (2009). *How to design and evaluate research in education*. New York, NY: McGraw-Hill.
- Gan, Z. (2009). IELTS preparation course and student IELTS performance: A case study in Hong Kong. *RELC Journal*, 40(1), 23–41. doi:10.1177/0033688208101449
- Gan, Z., Humphreys, G., & Hamp-Lyons, L. (2004). Understanding successful and unsuccessful EFL students in Chinese universities. *The Modern Language Journal*, 88(2), 229–244. doi:10.1111/j.0026-7902.2004.00227.x
- Gerwin, D., & Visone, F. (2006). The freedom to teach: Contrasting history teaching in elective and state-tested courses. *Theory and Research in Social Education*, 34(2), 259–282.
- Government. (2008, September 30). Quyết định 1400/QĐ-TTg phê duyệt đề án “dạy và học ngoại ngữ trong hệ thống giáo dục quốc dân giai đoạn 2008 - 2020.” Retrieved from <http://thuvienphapluat.vn/van-ban/Giao-duc/Quyet-dinh-1400-QD-TTg-phe->

duyet-de-an-day-va-hoc-ngoai-ngu-trong-he-thong-giao-duc-quoc-dan-giai-doan-2008-2020-71152.aspx

- Gray, D. E. (2004). *Doing research in the real world* (First Edition.). Sage Publications Ltd.
- Green, A. (2006). Watching for washback: Observing the influence of the International English Language Testing System Academic Writing Test in the classroom. *Language Assessment Quarterly*, 3(4), 333–368. doi:Article
- Green, A. (2007a). *IELTS washback in context: Preparation for academic writing in higher education*. Cambridge: Cambridge University Press.
- Green, A. (2007b). Washback to learning outcomes: A comparative study of IELTS preparation and university pre-sessional language courses. *Assessment in Education*, 14(1), 75–97.
- Griffiths, C. (2007). Language learning strategies: Students' and teachers' perceptions. *ELT Journal*, 61(2), 91–99. doi:10.1093/elt/ccm001
- Grodsky, E., Warren, J. R., & Kalogrides, D. (2009). State high school exit examinations and NAEP long-term trends in reading and mathematics, 1971-2004. *Educational Policy*, 23(4), 589–614.
- Gu, Y. (2014). To code or not to code: Dilemmas in analysing think-aloud protocols in learning strategies research. *System*, 43, 74–81. doi:10.1016/j.system.2013.12.011
- Hamp-Lyons, L. (1997). Washback, impact and validity: Ethical concerns. *Language Testing*, 14(3), 295–303. doi:10.1177/026553229701400306
- Hamp-Lyons, L. (1998). Ethical test preparation practice: The case of the TOEFL. *TESOL Quarterly*, 32(2), 329–337. doi:10.2307/3587587
- Hirai, M. (2002). Correlations between active skill and passive skill test scores. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 6(3), 2–8.

- Holme, J. J., Richards, M. P., Jimerson, J. B., & Cohen, R. W. (2010). Assessing the effects of high school exit examinations. *Review of Educational Research*, 80(4), 476–526. doi:10.3102/0034654310383147
- Hồng Ninh. (2010, November 3). Nhiều trường dùng chuẩn TOEIC môn Tiếng Anh. *Tiền Phong Online*. Retrieved October 21, 2011, from <http://www.tienphong.vn/Khoa-Giao/Giao-Duc/188545/Nhieu-truong-dung-chuan-TOEIC-mon-Tieng-Anh.html>
- Hwang, H.-J. (2003). *The impact of high-stakes exams on teachers and students: A washback study of the university entrance exam at the secondary school level in South Korea* (M.A.). McGill University (Canada), Canada.
- International Language Testing Association. (n.d.). ILTA Guidelines for Practice. Retrieved November 4, 2014, from http://www.iltaonline.com/images/pdfs/ILTA_Guidelines.pdf
- Kohn, A. (2000). Burnt at the high stakes. *Journal of Teacher Education*, 51(4), 315–327. doi:10.1177/0022487100051004007
- Liao, C., Qu, Y., & Morgan, R. (2010). *The relationships of test scores measured by the TOEIC Listening and Reading test and TOEIC Speaking*. Retrieved June 10, 2014, from http://www.ets.org/research/policy_research_reports/publications/report/2010/itkd
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: expectations and validation criteria. *Educational Researcher*, 20(8), 15–21. doi:10.3102/0013189X020008015
- Liu, J., & Costanzo, K. (2013). *The relationship among TOEIC listening, reading, speaking, and writing skills*. Retrieved June 10, 2014, from http://www.ets.org/research/policy_research_reports/publications/chapter/2013/jrob

- Li, X. (1990). How powerful can a language test be? The MET in China. *Journal of Multilingual and Multicultural Development*, 11(5), 393–404. doi:10.1080/01434632.1990.9994425
- Lodico, M. G., Spaulding, D. T., & Voegtle, K. H. (2006). *Methods in educational research: From theory to practice*. San Francisco, CA: Jossey-Bass.
- Lougheed, L. (2007). *Longman preparation series for the new TOEIC test* (4th ed.). Pearson ESL.
- Madaus, G. F. (1985). Public policy and the testing profession—you've never had it so good? *Educational Measurement: Issues and Practice*, 4(4), 5–11. doi:10.1111/j.1745-3992.1985.tb00294.x
- Manjarrés, N. B. (2005). Washback of the foreign language test of the state examinations in Colombia: A case study. *Arizona Working Papers in SLAT*, 12, 1–19.
- Marriott, P. (2009). Students' evaluation of the use of online summative assessment on an undergraduate financial accounting module. *British Journal of Educational Technology*, 40(2), 237–254.
- McCarthy, S. J. (2008). The impact of No Child Left Behind on teachers' writing instruction. *Written Communication*, 25(4), 462–505.
- McNamara, T. (1996). *Measuring second language performance*. Longman.
- Mehrens, W. A. (1984). National tests and local curriculum: Match or mismatch? *Educational Measurement: Issues and Practice*, 3(3), 9–15. doi:10.1111/j.1745-3992.1984.tb00757.x
- Mehrens, W. A. (1998). Consequences of assessment: What is the evidence? *Education Policy Analysis Archives*, 6, 13.
- Menken, K. (2009). No Child Left Behind and its effects on language policy. *Annual Review of Applied Linguistics*, 29, 103.

- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13–23. doi:10.3102/0013189X023002013
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241–256. doi:10.1177/026553229601300302
- Minh Giảng. (2012, April 28). TOEFL, TOEIC “nhái” - Giáo dục. *Tuổi Trẻ Online*. Retrieved December 27, 2014, from <http://tuoitre.vn/tin/giao-duc/20120428/toefl-toeic-nhai/489285.html>
- Muñoz, A. P., & Álvarez, M. E. (2010). Washback of an oral assessment system in the EFL classroom. *Language Testing*, 27(1), 33–49.
- Newfields, T. (2005). TOEIC® Washback effects on teachers: A pilot study at one university faculty. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 31(1), 83–106.
- Nhan, T. (2013). The TOEIC test as an exit requirement in universities and colleges in Danang City, Vietnam: Challenges and Impacts. Retrieved from <http://www.auamii.com/jiir/Vol-02/issue-01/4Nhan.pdf>
- Như Thuần. (2008, March 23). Áp dụng đại trà chuẩn TOEIC vào đại học: Không phù hợp với Việt Nam. *Sài Gòn Tiếp Thị Online*. Retrieved February 22, 2011, from <http://sgtt.vn/Khoa-giao/79905/Ap-dung-dai-tra-chuan-TOEIC-vao-dai-hoc-Khong-phu-hop-voi-Viet-Nam.html>
- Nichols, S. L., Glass, G. V., & Berliner, D. C. (2006). High-Stakes testing and student achievement: Does accountability pressure increase student learning? *Education Policy Analysis Archives*, 14(1), 1–172.
- Nunan, D., Berry, R., & Berry, V. (Eds.). (1994). *Bringing about change in language education*. Department of Curriculum Studies, University of Hong Kong.

- O, K.-M. (2001). *Washback on high school classrooms of the English tests within Korean university entrance exams*. Columbia University Teachers College, United States -- New York.
- Pallant, J. (2011). *SPSS survival manual a step by step guide to data analysis using the SPSS program, 4th edition*. Crows Nest, N.S.W., Australia: Allen & Unwin.
- Pan, Y.-C., & Newfields, T. (2011). Teacher and student washback on test preparation evidenced from Taiwan's English certification exit requirements. *International Journal of Pedagogies and Learning*, 6(3).
- Pan, Y.-C., & Newfields, T. (2012). Tertiary EFL proficiency graduation requirements in Taiwan: A study of washback on learning. *Electronic Journal of Foreign Language Teaching*, 9(1), 108–122.
- Parke, C. S., Lane, S., & Stone, C. A. (2006). Impact of a state performance assessment program in reading and writing. *Educational Research and Evaluation*, 12(3), 239–269.
- Popham, W. J. (1987). The merits of measurement-driven instruction. *The Phi Delta Kappan*, 68(9), 679–682.
- Powers, D. E., Kim, H.-J., & Weng, V. Z. (2008). *The Redesigned TOEIC (Listening and Reading) Test: Relations to Test-Taker Perceptions of Proficiency*. Retrieved June 10, 2014, from http://www.ets.org/research/policy_research_reports/publications/report/2008/htzv
- 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. doi:10.1080/09695940701272856

- Quesada Inces, R. (2001). *Washback overrides the curriculum: An exploratory study on the washback effect of a high-stakes standardized test in the Costa Rican EFL high school context*. The Pennsylvania State University, United States -- Pennsylvania.
- Read, J., & Hayes, B. (2003). *The Impact of IELTS on preparation for Academic Study in New Zealand*. (No. 4) (pp. 153–206).
- Robb, T. N., & Ercanbrack, J. (1999). *A study of the effect of direct test preparation on the TOEIC scores of Japanese university students*. Retrieved February 1, 2011, from http://www.eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_ERICExtSearch_SearchValue_0=EJ579815&ERICExtSearch_SearchType_0=no&accno=EJ579815
- Roderick, M., Jacob, B. A., & Bryk, A. S. (2002). The impact of high-stakes testing in Chicago on student achievement in promotional gate grades. *Educational Evaluation and Policy Analysis*, 24(4), 333–357.
- Ross, S. J. (2008). The social and political tensions of language assessment. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2, pp. 786–798). London: Routledge.
- Saif, S. (2006). Aiming for positive washback: A case study of international teaching assistants. *Language Testing*, 23(1), 1–34. doi:10.1191/0265532206lt322oa
- Scott, C. (2007). Stakeholder perceptions of test impact. *Assessment in Education: Principles, Policy & Practice*, 14(1), 27–49. doi:10.1080/09695940701272807
- Shih, C.-M. (2007). A new washback model of students' learning. *The Canadian Modern Language Review*, 64(1).
- Shohamy, E. (1993). *The power of tests: The impact of language tests on teaching and learning*. *NFLC occasional papers*. National Foreign Language Center, 1619 Massachusetts Ave., Washington, DC 20036. Retrieved from

<http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED362040>

- Shohamy, E. (1997). Testing methods, testing consequences: Are they ethical? Are they fair? *Language Testing*, 14(3), 340–349. doi:10.1177/026553229701400310
- Shohamy, E. (1998). Critical language testing and beyond. *Studies in Educational Evaluation*, 24(4), 331–45.
- Shohamy, E. (2001a). Democratic assessment as an alternative. *Language Testing*, 18(4), 373–391. doi:10.1177/026553220101800404
- Shohamy, E. (2001b). *The power of tests: A critical perspective on the uses of language tests*. Harlow, England; New York: Longman.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13(3), 298–317.
- Smith, J., & Kovacs, P. (2011). The impact of standards-based reform on teachers: The case of “No Child Left Behind.” *Teachers and Teaching*, 17(2), 201–225. doi:10.1080/13540602.2011.539802
- Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20(5), 8–11. doi: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.
- Stecher, B., Chun, T., & Barron, S. (2004). The effects of assessment-driven reform on the teaching of writing in Washington State. In L. Cheng, Y. J. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods*. Mahwah, N.J.: Lawrence Erlbaum.

- Stoneman, B. (2006). *The impact of an exit English test on Hong Kong undergraduates: A study investigating the effects of test status on students' test preparation behaviours*. Hong Kong Polytechnic University (Hong Kong), Hong Kong.
- Tanabe, H. (2006). The effect of a three-month TOEIC training program. 東京工芸大学工学部紀要. 人文・社会編, 29(2), 1–7.
- Taylor, L. (2005). Washback and impact. *ELT Journal*, 59(2), 154–155.
doi:10.1093/eltj/ccj030
- Thanh Hà. (2005, November 25). Sử dụng TOEIC đánh giá trình độ tiếng Anh của SV. *Tuổi Trẻ Online*. Retrieved February 22, 2011, from <http://tuoitre.vn/Giao-duc/110223/Su-dung-TOEIC-danh-gia-trinh-do-tieng-Anh-cua-SV.html>
- Thomas, P. (2004). The negative impact of testing writing skills. *Educational Leadership*, 62(2), 76.
- Trung Toàn. (2008, October 12). Đào tạo tiếng Anh trong các trường không chuyên ngữ. *Báo Giáo dục Online*. Retrieved May 18, 2015, from <http://www.giaoduc.edu.vn/news/giao-duc-phat-trien-728/dao-cao-tieng-anh-trong-cac-truong-khong-chuyen-ngu-109860.aspx>
- Tsagari, D. (2007, July 25). Review of washback in language testing: How has been done? What more needs doing? Retrieved February 1, 2011, from http://newfirstsearch.oclc.org/helicon.vuw.ac.nz/WebZ/FSPage?pagetype=return_frameset:sessionid=fsapp8-56673-gjnd7zkk-ltfdkk:entitypagenum=4:0:entityframedurl=http%3A%2F%2Fwww.eric.ed.gov%2Fcontentdelivery%2F servlet%2FERICServlet%3Faccno%3DED497709:entityframedtitle=ERIC:entityframedtimeout=15:entityopenTitle=:entityopenAuthor=:entityopenNumber=:

- Tsagari, D. (2011). *Washback of a high-stakes English exam on teachers' perceptions and practices*. Selected. Retrieved from http://www.enl.auth.gr/symposium19/19thpapers/042_Tsagari.pdf
- Vogler, K. E., Lintner, T., Lipscomb, G. B., Knopf, H., Heafner, T. L., & Rock, T. C. (2007). Getting off the back burner: Impact of testing elementary social studies as part of a state-mandated accountability program. *Journal of Social Studies Research, 31*(2), 20–34.
- Wall, D. (2000). The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled? *System, 28*(4), 499–509. doi:10.1016/S0346-251X(00)00035-X
- Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing, 10*(1), 41–69. doi:10.1177/026553229301000103
- Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing, 13*(3), 318–333. doi:10.1177/026553229601300306
- Watanabe, Y. (2000). Washback Effects of the English section of Japanese university entrance examinations on instruction in pre-college level EFL. *Language Testing Update, 27*(summer), 42–47.
- Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods*. Mahwah, N.J.: Lawrence Erlbaum.
- William, D. (2010). Standardized testing and school accountability. *Educational Psychologist, 45*(2), 107–122.
- Wilson, K. M. (1989). *Enhancing the interpretation of a norm-referenced second-language test through criterion referencing: A research assessment of experience*

in the TOEIC testing context (TOEIC Research Report No. 1). Retrieved June 10, 2014, from http://www.ets.org/research/policy_research_reports/publications/report/1989/hwwb

Woodford, P. E. (1982). *An introduction to TOEIC: The initial validity study* (TOEIC Research Summary Number 0). Retrieved April 19, 2015, from http://www.ets.org/research/policy_research_reports/publications/publication/1982/cglx

Wright, W. E., & Choi, D. (2006). The impact of language and high-stakes testing policies on elementary school English language learners in Arizona. *Education Policy Analysis Archives*, 14(13), 1–75.

Xie, Q., & Andrews, S. (2013). Do test design and uses influence test preparation? Testing a model of washback with Structural Equation Modeling. *Language Testing*, 30(1), 49–70. doi:<http://dx.doi.org/helicon.vuw.ac.nz/10.1177/0265532212442634>

Appendix A: Teacher Questionnaire

TOEIC-AS-EXIT-TEST QUESTIONNAIRE

Dear Colleague,

This questionnaire surveys the impact of the TOEIC as an exit test in Vietnamese universities. The results of the study will help improve English teaching, learning, and assessment effectiveness, so your contribution is very important. The information you provide is anonymous and will be used for research purposes only.

Thank you for your valuable time and opinions.

PART I: PERSONAL INFORMATION

Please check the appropriate box(es) or write your answers.

1. Gender: Male ☐ Female ☐
2. Age: years old
3. Highest qualification: BA ☐ MA ☐ PhD ☐
4. Experience teaching English: years..... months
5. Experience teaching TOEIC: years..... months

PART II: IMPACT OF THE TOEIC TEST

A. IMPACT OF THE TOEIC TEST ON TEACHING

Please indicate the extent to which you agree with the following statements by **circling** the appropriate number.

- 1 = Strongly disagree
2 = Disagree
3 = Somewhat disagree
4 = Neither agree nor disagree
5 = Somewhat agree
6 = Agree
7 = Strongly agree

- (1) To what extent has the TOEIC test affected your feelings and attitudes to teaching TOEIC or preparing students for the TOEIC test?

	Strongly disagree							Strongly agree						
1. I feel positive about my students' TOEIC test results.	1	2	3	4	5	6	7							
2. I feel anxious about improving my students' test scores.	1	2	3	4	5	6	7							
3. The success or failure of my students on the TOEIC test reflects on me.	1	2	3	4	5	6	7							
4. Preparing students for the TOEIC is enjoyable.	1	2	3	4	5	6	7							
5. Preparing students for the TOEIC test is difficult.	1	2	3	4	5	6	7							
6. Preparing students for the TOEIC test narrows the necessary content of English teaching.	1	2	3	4	5	6	7							

(2) To what extent has the TOEIC test influenced **WHAT** you teach?

- 1 = Strongly disagree
2 = Disagree
3 = Somewhat disagree
4 = Neither agree nor disagree
5 = Somewhat agree
6 = Agree
7 = Strongly agree

Since the introduction of the TOEIC test,...	Strongly disagree				Strongly agree		
7. ...the TOEIC test has been influencing my decisions about which language skills are more important to be taught.	1	2	3	4	5	6	7
8. ...the TOEIC test has been influencing my supplementary material selection.	1	2	3	4	5	6	7
9. ...I have been teaching students more test-taking skills than before.	1	2	3	4	5	6	7
10. ...I have been teaching students multiple-choice test-taking skills more than before.	1	2	3	4	5	6	7
11. ...I have been skipping over certain sections of the textbooks which are <u>less likely</u> to be tested on the TOEIC test.	1	2	3	4	5	6	7
12. ...I have been using past exam papers to prepare my students for the test.	1	2	3	4	5	6	7
13. ...I have been using mock tests to prepare my students for the test.	1	2	3	4	5	6	7
14. ...I have been using exam preparation materials to prepare my students for the test.	1	2	3	4	5	6	7
15. ...the TOEIC test has had <u>little</u> impact on what I teach.	1	2	3	4	5	6	7

(3) To what extent has the TOEIC test influenced **HOW** you teach?

Since the introduction of the TOEIC test,...	Strongly disagree				Strongly agree		
16. ...I have been using teaching methods that I did not use before.	1	2	3	4	5	6	7
17. ...I have changed my time allotment to teaching and learning activities in class.	1	2	3	4	5	6	7
18. ...the TOEIC test has had <u>little</u> impact on how I teach.	1	2	3	4	5	6	7

Please answer the following questions by **circling** the appropriate number.

- 1 = Decreased greatly
 2 = Decreased moderately
 3 = Decreased slightly
 4 = Stayed the same
 5 = Increased slightly
 6 = Increased moderately
 7 = Increased greatly

(4) To what extent has your total instructional time devoted to the following skills during the whole semester changed as a result of the introduction of the TOEIC test?

	Decreased greatly			Stay the same			Increased greatly	
19. Listening	1	2	3	4	5	6	7	
20. Speaking	1	2	3	4	5	6	7	
21. Reading	1	2	3	4	5	6	7	
22. Writing	1	2	3	4	5	6	7	
23. Grammar	1	2	3	4	5	6	7	
24. Vocabulary	1	2	3	4	5	6	7	
25. Pronunciation	1	2	3	4	5	6	7	

(5) To what extent has your total instructional time devoted to the following teaching and learning activities during the whole semester changed as a result of the introduction of the TOEIC test?

	Decreased greatly			Stay the same			Increased greatly	
26. Detailed grammatical analyses of texts	1	2	3	4	5	6	7	
27. Mechanical drills, e.g. repetition, substitution, addition, transformation	1	2	3	4	5	6	7	
28. Your use of metalanguage (grammatical terms)	1	2	3	4	5	6	7	
29. Your use of Vietnamese	1	2	3	4	5	6	7	
30. Role-plays	1	2	3	4	5	6	7	
31. Information gap activities	1	2	3	4	5	6	7	
32. Information gathering activities, e.g. survey, interviews	1	2	3	4	5	6	7	
33. Group discussions	1	2	3	4	5	6	7	
34. Your students' talking time	1	2	3	4	5	6	7	

(6) What are other impacts of the TOEIC test on your teaching that have not been mentioned?

.....

.....

.....

.....

.....

.....

.....

B. FACTORS ASSOCIATED WITH TEACHING

Please indicate the extent to which you agree with the following statements by **circling** the appropriate number.

- 1 = Strongly disagree
- 2 = Disagree
- 3 = Somewhat disagree
- 4 = Neither agree nor disagree
- 5 = Somewhat agree
- 6 = Agree
- 7 = Strongly agree

(7) How do you perceive the TOEIC test that your university is using?

	Strongly disagree				Strongly agree			
35. The purpose of this TOEIC is to assess test-takers' communicative ability.	1	2	3	4	5	6	7	
36. The purpose of this TOEIC is to assess test-takers' grammatical knowledge.	1	2	3	4	5	6	7	
37. This TOEIC test measures test-takers' listening skills.	1	2	3	4	5	6	7	
38. This TOEIC test measures test-takers' speaking skills.	1	2	3	4	5	6	7	
39. This TOEIC test measures test-takers' reading skills.	1	2	3	4	5	6	7	
40. This TOEIC test measures test-takers' writing skills.	1	2	3	4	5	6	7	
41. This TOEIC test measures the most important English knowledge and skills that university students should have.	1	2	3	4	5	6	7	
42. This TOEIC has a multiple-choice format.	1	2	3	4	5	6	7	
43. This TOEIC test is easy for the majority of my students.	1	2	3	4	5	6	7	
44. The listening section of this TOEIC test is difficult for the majority of my students.	1	2	3	4	5	6	7	
45. The reading section of this TOEIC test is difficult for the majority of my students.	1	2	3	4	5	6	7	
46. In general, this TOEIC test scores accurately reflect my students' English proficiency.	1	2	3	4	5	6	7	
47. Provincial students have more difficulty with this TOEIC test than students from large cities.	1	2	3	4	5	6	7	
48. This TOEIC test provides useful information as feedback to my teaching.	1	2	3	4	5	6	7	
49. This TOEIC test provides useful information as feedback to my students' learning.	1	2	3	4	5	6	7	
50. This TOEIC test helps my students to set goals for their English learning activities.	1	2	3	4	5	6	7	
51. This TOEIC test is useful for my students' future career.	1	2	3	4	5	6	7	
52. Failing to meet the TOEIC benchmarks has bad consequences for my students' graduation.	1	2	3	4	5	6	7	
53. Students' low TOEIC test scores result in bad consequences for teachers.	1	2	3	4	5	6	7	
54. Students' high TOEIC scores result in rewards to teachers.	1	2	3	4	5	6	7	
55. This TOEIC test is suitable for evaluating English learning outcomes of university students.	1	2	3	4	5	6	7	

(8) What are your beliefs about language teaching and learning?

- 1 = Strongly disagree
2 = Disagree
3 = Somewhat disagree
4 = Neither agree nor disagree
5 = Somewhat agree
6 = Agree
7 = Strongly agree

	Strongly disagree				Strongly agree		
56. Language learning is for communication.	1	2	3	4	5	6	7
57. The most important part of teaching a foreign language is teaching vocabulary words.	1	2	3	4	5	6	7
58. The most important part of teaching a foreign language is teaching grammar rules.	1	2	3	4	5	6	7
59. The most important part of teaching a foreign language is teaching how to translate things.	1	2	3	4	5	6	7
60. It's necessary to teach university students all the four skills: listening, speaking, reading and writing.	1	2	3	4	5	6	7
61. The <u>communicative approach</u> is an effective way of teaching language.	1	2	3	4	5	6	7
62. The <u>grammar translation method</u> is an effective way of teaching language.	1	2	3	4	5	6	7
63. The best way to learn a language is to use it for communication.	1	2	3	4	5	6	7
64. Coaching is necessary for students to get high test scores.	1	2	3	4	5	6	7
65. Learning test-taking strategies is necessary to improve test scores.	1	2	3	4	5	6	7
66. Overall improvement in language learning will result in improved test scores.	1	2	3	4	5	6	7

(9) To what extent do you feel the pressure to prepare your students for the TOEIC test?

	Strongly disagree				Strongly agree		
67. My students expect me to teach what will be tested.	1	2	3	4	5	6	7
68. My colleagues expect me to prepare students for the test.	1	2	3	4	5	6	7
69. The administrators expect me to prepare my students for the TOEIC test.	1	2	3	4	5	6	7
70. Employers prefer university graduates with high TOEIC scores.	1	2	3	4	5	6	7
71. My students test scores are often compared with those of my colleagues'.	1	2	3	4	5	6	7
72. My teaching is assessed by the performance of my students in the TOEIC test.	1	2	3	4	5	6	7

(10) To what extent do the following context factors influence your teaching?

- 1 = Strongly disagree
2 = Disagree
3 = Somewhat disagree
4 = Neither agree nor disagree
5 = Somewhat agree
6 = Agree
7 = Strongly agree

	Strongly disagree			Strongly agree			
73. I can use the textbook selectively.	1	2	3	4	5	6	7
74. I have to stick closely to the university syllabus.	1	2	3	4	5	6	7
75. I have access to all the materials I want to use.	1	2	3	4	5	6	7
76. I have access to all the equipment I need.	1	2	3	4	5	6	7
77. My students don't have enough learning resources.	1	2	3	4	5	6	7
78. The number of students in my class prevents me from organizing certain teaching and learning activities.	1	2	3	4	5	6	7
79. The number of students in my class prevents me from teaching certain contents.	1	2	3	4	5	6	7
80. I often consult my colleagues about what to teach.	1	2	3	4	5	6	7
81. I often consult my colleagues about how to teach.	1	2	3	4	5	6	7
82. I don't have enough time to teach what I think is necessary.	1	2	3	4	5	6	7
83. I don't have enough time to organize the activities that I think are useful.	1	2	3	4	5	6	7
84. My students are highly motivated to study English.	1	2	3	4	5	6	7
85. My students can get the required TOEIC scores without my help.	1	2	3	4	5	6	7

Thank you for your cooperation!

Appendix B: Student Questionnaire

TOEIC-AS-EXIT-TEST QUESTIONNAIRE

Dear Student,

This questionnaire surveys the impact of the TOEIC as an exit test in Vietnamese universities. The results of the study will help improve English teaching, learning, and assessment effectiveness, so your contribution is very important. The information you provide is anonymous and will be used for research purposes only.

Thank you for your valuable time and opinions.

PART 1: PERSONAL INFORMATION

Please check the appropriate box(es) or write your answers.

1. Gender: Male ☐ Female ☐
2. Experience learning English (since you first started learning English to present):
..... years..... months (example: 9 years 4 months)
3. Experience learning TOEIC: years..... months (example: 0 year 8 months)
4. Year at university: First year ☐ Second year ☐ Third year ☐ Fourth year ☐
5. Major:
6. Province/City where you go to secondary school:

PART 2: IMPACT OF THE TOEIC TEST

A. IMPACT OF THE TOEIC TEST ON LANGUAGE LEARNING

Please indicate the extent to which you agree with the following statements on a scale from 1-7 by **circling** the appropriate number.

- 1 = Strongly disagree
2 = Disagree
3 = Somewhat disagree
4 = Neither agree nor disagree
5 = Somewhat agree
6 = Agree
7 = Strongly agree

(1) To what extent has the TOEIC test affected your feelings and attitudes to learning English?

	Strongly disagree						Strongly agree
1) I find learning TOEIC enjoyable.	1	2	3	4	5	6	7
2) I feel stressed about having to pass the TOEIC test.	1	2	3	4	5	6	7
3) The TOEIC test motivates me to learn English.	1	2	3	4	5	6	7
4) The TOEIC test motivates me to learn the tested skills.	1	2	3	4	5	6	7
5) I have become more engaged in classroom activities.	1	2	3	4	5	6	7
6) The TOEIC benchmark makes me more aware of the importance of English learning.	1	2	3	4	5	6	7
7) The TOEIC benchmark makes me take English learning more seriously.	1	2	3	4	5	6	7

(2) To what extent has the TOEIC test affected WHAT you learn?

Since I knew I had to take the TOEIC test,...

	Strongly disagree						Strongly agree
8) ...I have been skipping over certain sections of the textbooks which are less likely to be tested on the TOEIC test.	1	2	3	4	5	6	7
9) ...I have only been learning what is tested on the TOEIC.	1	2	3	4	5	6	7
10) ...I have been learning more test-taking skills than before.	1	2	3	4	5	6	7
11) ...I have been learning multiple-choice test-taking strategies more than before.	1	2	3	4	5	6	7
12) ...the TOEIC test has been influencing my choice of materials I use for self-study.	1	2	3	4	5	6	7
13) ...I have been doing TOEIC past exam papers.	1	2	3	4	5	6	7
14) ...I have been doing mock tests.	1	2	3	4	5	6	7
15) ...I have been using TOEIC preparation materials to learn for the test.	1	2	3	4	5	6	7
16) ...the TOEIC test has had little impact on what I learn.	1	2	3	4	5	6	7

(3) To what extent has the TOEIC test affected HOW you learn English?

- 1 = Strongly disagree
 2 = Disagree
 3 = Somewhat disagree
 4 = Neither agree nor disagree
 5 = Somewhat agree
 6 = Agree
 7 = Strongly agree

Since I knew I had to take the TOEIC test,...		Strongly disagree					Strongly agree				
17)	...I have been using learning strategies that I did not use before.	1	2	3	4	5	6	7			
18)	...I have been spending more time studying English.	1	2	3	4	5	6	7			
19)	...I have changed my time allotment to my learning activities.	1	2	3	4	5	6	7			
20)	...I have been taking TOEIC cram courses outside the university.	1	2	3	4	5	6	7			
21)	...I have been finding information about the test, e.g., timing, format, question types.	1	2	3	4	5	6	7			
22)	...I have been seeking help from English language teachers about the test.	1	2	3	4	5	6	7			

Please indicate to what extent the introduction of the TOEIC test has changed your learning time allocation by **circling** the appropriate number.

- 1 = Decreased greatly
 2 = Decreased moderately
 3 = Decreased slightly
 4 = Stayed the same
 5 = Increased slightly
 6 = Increased moderately
 7 = Increased greatly

(4) To what extent has the TOEIC test changed your total self-study time devoted to the following skills during the course?

		Decreased greatly			Stay the same			Increased greatly		
23)	Listening	1	2	3	4	5	6	7		
24)	Speaking	1	2	3	4	5	6	7		
25)	Reading	1	2	3	4	5	6	7		
26)	Writing	1	2	3	4	5	6	7		
27)	Grammar	1	2	3	4	5	6	7		
28)	Vocabulary	1	2	3	4	5	6	7		
29)	Pronunciation	1	2	3	4	5	6	7		

(5) To what extent has the TOEIC test changed your total self-study time devoted to the following learning strategies during the course?

- 1 = Decreased greatly
 2 = Decreased moderately
 3 = Decreased slightly
 4 = Stayed the same
 5 = Increased slightly
 6 = Increased moderately
 7 = Increased greatly

		Decreased greatly			Stay the same			Increased greatly		
		1	2	3	4	5	6	7		
30)	Watching/listening to English media, e.g., TV, movies, videos, radio	1	2	3	4	5	6	7		
31)	Talking to my friends or family members in English	1	2	3	4	5	6	7		
32)	Talking to native speakers of English in person	1	2	3	4	5	6	7		
33)	Communicating in English on the Internet, e.g., exchange emails or chat with foreigners	1	2	3	4	5	6	7		
34)	Reading English media, e.g., newspapers, magazines, books	1	2	3	4	5	6	7		
35)	Learning grammar rules	1	2	3	4	5	6	7		
36)	Doing grammar exercises	1	2	3	4	5	6	7		
37)	Doing translation exercises	1	2	3	4	5	6	7		
38)	Learning by repetition	1	2	3	4	5	6	7		

(6) To what extent has the TOEIC test affected your achievement?

- 1 = Strongly disagree
 2 = Disagree
 3 = Somewhat disagree
 4 = Neither agree nor disagree
 5 = Somewhat agree
 6 = Agree
 7 = Strongly agree

Preparing for the TOEIC test has improved ...	Strongly disagree			Strongly agree			
39) ...my speaking skills.	1	2	3	4	5	6	7
40) ...my listening skills.	1	2	3	4	5	6	7
41) ...my reading skills.	1	2	3	4	5	6	7
42) ...my writing skills.	1	2	3	4	5	6	7
43) ...my achievement test scores, e.g., quizz, midterm, end-of-tem test scores.	1	2	3	4	5	6	7
44) ...my English proficiency in general.	1	2	3	4	5	6	7

(7) What are other impacts of the TOEIC test on your learning that have not been mentioned?

.....

B. FACTORS ASSOCIATED WITH LEARNING

Please indicate the extent to which you agree with the following statements by **circling** the appropriate number.

- 1 = Strongly disagree
- 2 = Disagree
- 3 = Somewhat disagree
- 4 = Neither agree nor disagree
- 5 = Somewhat agree
- 6 = Agree
- 7 = Strongly agree

(8) How do you perceive the TOEIC test that YOUR UNIVERSITY is using?

		Strongly disagree				Strongly agree			
		1	2	3	4	5	6	7	
45)	The purpose of this TOEIC is to assess test-takers' communicative ability.	1	2	3	4	5	6	7	
46)	The purpose of this TOEIC is to assess test-takers' grammatical knowledge.	1	2	3	4	5	6	7	
47)	This TOEIC test measures test-takers' listening skills.	1	2	3	4	5	6	7	
48)	This TOEIC test measures test-takers' speaking skills.	1	2	3	4	5	6	7	
49)	This TOEIC test measures test-takers' reading skills.	1	2	3	4	5	6	7	
50)	This TOEIC test measures test-takers' writing skills.	1	2	3	4	5	6	7	
51)	The TOEIC test measures the most important English knowledge and skills that university students should have.	1	2	3	4	5	6	7	
52)	The TOEIC test is easy for me.	1	2	3	4	5	6	7	
53)	The listening section of the TOEIC test is difficult for me.	1	2	3	4	5	6	7	
54)	The reading section of the TOEIC test is difficult for me.	1	2	3	4	5	6	7	
55)	Provincial students have more difficulty with the TOEIC test than students from large cities.	1	2	3	4	5	6	7	
56)	In general, this TOEIC test scores accurately reflect students' English proficiency.	1	2	3	4	5	6	7	
57)	This TOEIC test is useful for my future career.	1	2	3	4	5	6	7	
58)	This TOEIC test provides useful information as feedback to my learning.	1	2	3	4	5	6	7	
59)	This TOEIC test helps me set goals for my English learning activities.	1	2	3	4	5	6	7	
59a)	This TOEIC test has a multiple-choice format.	1	2	3	4	5	6	7	

(9) What are your beliefs about language teaching and learning?

- 1 = Strongly disagree
2 = Disagree
3 = Somewhat disagree
4 = Neither agree nor disagree
5 = Somewhat agree
6 = Agree
7 = Strongly agree

	Strongly disagree						Strongly agree
60) Language learning is for communication.	1	2	3	4	5	6	7
61) The most important part of learning a foreign language is learning vocabulary words.	1	2	3	4	5	6	7
62) The most important part of learning a foreign language is learning grammar rules.	1	2	3	4	5	6	7
63) The most important part of learning a foreign language is learning how to translate things.	1	2	3	4	5	6	7
64) It's necessary for university students to learn all the four skills: speaking, listening, reading and writing skills.	1	2	3	4	5	6	7
65) Students' overall improvement in language learning will result in improved test scores.	1	2	3	4	5	6	7
66) Learning test-taking strategies is necessary to improve test scores.	1	2	3	4	5	6	7

(10) To what extent do you feel the pressure to prepare for the TOEIC test?

	Strongly disagree						Strongly agree
67) My teachers expect me to learn what will be tested.	1	2	3	4	5	6	7
68) My parents urge me to study intensively for the test.	1	2	3	4	5	6	7
69) My peers expect me to prepare for the test.	1	2	3	4	5	6	7
70) Employers prefer university graduates with high TOEIC scores.	1	2	3	4	5	6	7
71) I feel embarrassed if I perform less well on the TOEIC test than my classmates.	1	2	3	4	5	6	7
72) I must get the required TOEIC score before I graduate.	1	2	3	4	5	6	7

(11) To what extent do the following context factors influence your learning?

	Strongly disagree						Strongly agree
73) I have access to the learning materials I want to learn.	1	2	3	4	5	6	7
74) I have access to the learning equipment I want to use, e.g., CD player, computer, USB, MP3, MP4...	1	2	3	4	5	6	7
75) I often consult my classmates about what to learn.	1	2	3	4	5	6	7
76) I often consult my classmates about how to learn.	1	2	3	4	5	6	7
77) My teachers influence my decision about what to learn.	1	2	3	4	5	6	7
78) My teachers influence my decision about how to learn.	1	2	3	4	5	6	7
79) I don't have enough time to learn what I think is necessary.	1	2	3	4	5	6	7
80) I don't have enough time to do the activities that I think are useful.	1	2	3	4	5	6	7

Thank you for your cooperation!

Appendix C: MOET Official Interview Schedule

MOET OFFICIAL INTERVIEW SCHEDULE

Interview purpose

The purpose of this interview is to understand your and MOET's position on universities' English exit test policy in general and the use of the TOEIC-as-exit-test in particular.

Interview questions

General information

1. What is MOET's involvement in universities' foreign language teaching policies?
2. What do you think are the purposes of English teaching at university?
3. What level of English do you think university graduates should reach? Why?
4. How best, in your opinion, can these targets be achieved?

MOET's rationales and intentions behind its support for the English exit test policy and the use of TOEIC as a language exit test

5. Many universities and junior colleges now use an external English test as their language exit test. Does MOET support this policy? If yes, what are MOET's rationales and intentions behind its support for the policy? If no, why not?
6. According to a newspaper report, about one third of universities and junior colleges use TOEIC as a language exit test. In 2005, MOET representatives signed memorandums with ETS on using TOEIC scores as English benchmarks for graduating students. Why does MOET support this use of the TOEIC test?
7. Do you think the TOEIC test is suitable for this gate-keeping purpose? Why? (format, content, skills tested, test purposes)
8. Do you think the TOEIC test scores accurately reflect university students' language ability?
9. Do you think the use of the TOEIC-as-exit-test can ensure students' ability to use English effectively in their future jobs?
10. What else does MOET hope to achieve with its support for the TOEIC-as-exit-test policy? (effects on teaching content, teaching methods, learning content, learning activities, achievement, possibility of students' getting employment after graduation, teacher and student motivation, status of the English subject, etc.)

Wrap up

11. Is there anything else you would like to share with me so that I can fully understand your and MOET's position about the use of TOEIC as a language exit test in universities?

Appreciation

Appendix D: Head of Department Interview Schedule

HOD INTERVIEW SCHEDULE

Interview purpose

The purpose of this interview is to get information about the use of TOEIC as a language exit test at your university and to understand your opinions and perceptions regarding this use of the TOEIC test.

Interview questions

General information

1. How long have you been Dean of the English Department?
2. What is your involvement in the university's foreign language teaching policies? (decisions regarding syllabus, exams, and textbooks)
3. What do you think are the purposes of English teaching at university?
4. What level of English do you think university graduates should reach? Why?
5. How best, in your opinion, can these targets be achieved?

Rationales and Intentions behind the TOEIC-as-exit-test policy

6. When did the university start the English exit test policy?
7. Why did the university decide to adopt the English exit test policy?
8. Why did the university choose TOEIC to be the language exit test?
9. What do you hope to achieve with the TOEIC-as-exit-test policy? (effects on teaching content, teaching methods, learning content, learning activities, achievements, possibility of students' getting jobs after graduation, teacher and student motivation, status of the English subject, etc.)
10. What level of English competency should your students have when they achieve the specified TOEIC score?

Perception of the TOEIC test

11. What do you think is the purpose of the TOEIC test?
12. Which language skills does the test measure?
13. Overall, how difficult is the TOEIC test for your students?
14. How difficult is each section of the test for your students?
15. What do you think of the test format?
16. How do you interpret TOEIC test scores?
17. Do you think improved test scores means improved language proficiency?
18. Do you think TOEIC test scores accurately reflect your students' language ability?
19. Overall, do you think the TOEIC test is suitable for this gate-keeping purpose?

Wrap up

20. Is there anything else you would like to share with me so that I can fully understand the issues related to the use of TOEIC as a language exit test at your university?

Appreciation

Appendix E: Teacher Interview Schedule

TEACHER INTERVIEW SCHEDULE

Interview purpose

The purpose of this interview is to understand the impact of the TOEIC-as-exit-test on your teaching and your students' learning. It also seeks to understand the factors associated with your teaching and the test impact.

Interview questions

Personal information

1. What is the highest qualification that you have?
2. Why did you choose to become an English teacher? Do you like teaching?
3. How long have you been teaching English?
4. How long have you been teaching TOEIC?
5. What types of English courses have you taught?

Beliefs about language teaching and learning

6. What are the goals of language teaching at university?
7. What is the most important part of teaching a foreign language (grammar, vocabulary, translation or language use, communication)
8. How do you understand "communication ability"?
9. Do you think it's important for university students to learn all four language skills? Why?
10. Which skills are most important to teach to university students? Why?
11. What is the best way to learn a language? Why?
12. What is the best way to improve test scores? Why?
13. Is learning test-taking skills important to improve test scores? Why?

Perception of the TOEIC test

14. What is the purpose of the TOEIC test? Why do you think so?
15. Which skills does the test measure?

16. What format does it have?
17. Is the test difficult for your students? Which parts of the test are difficult for them? Why?
18. Do you think the test accurately reflects students' English proficiency? Why?
19. Do you think the test is useful for university students? In what ways? (for their learning, their careers)
20. Do you think the test is suitable for graduation purposes? Why?

Pressure and context factors

21. What will be the possible consequences for you if your students don't get the required score?
22. Is preparing your students for the TOEIC test stressful for you? Why? In what ways?
23. Do your students/ colleagues/administrators expect/urge you to prepare for the TOEIC test?
24. Do you have enough time to teach what you think is necessary?
25. Do you have enough resources to teach English?
26. Are there any other factors that influence your teaching? What are they? (students' demands, students' language proficiency, students' learning styles, your colleagues, administrators, your personality, your teaching experience, your educational background, teaching and learning culture)
27. What do you think is the most important factor that motivates you to prepare your students for the TOEIC test?

Test impact on teaching and learning

28. Has the university made any changes regarding curriculums, textbooks, other teaching materials, regulations, staff's professional development, etc. due to the adoption of the TOEIC-as-exit-test policy?
29. Has the TOEIC test affected your teaching? In what ways?

30. What do you think about preparing your students for the TOEIC test? Is it difficult? Is it interesting? Why?
31. Since the inset of the TOEIC test, have you been teaching your students more test-taking strategies than before? Why?
32. Have you been using mock tests, past exam papers, and/or test preparation materials?
33. Have you been using any different methods than before?
34. Has the TOEIC test affected your instructional time allotment? Why?
35. Which skills have you been spending more time on than before? Why?
36. Which classroom activities have you been spending more time on than before? Why?
37. Has the TOEIC test affected your students' motivation? In what ways?
38. Has the TOEIC test affected your students' content of learning? In what ways?
39. Has the TOEIC test affected your students' learning activities? In what ways?
40. Has TOEIC preparation improved students' English? In what ways?

Wrap up

41. Is there anything else you would like to share with me so that I can fully understand the issues related to the use of TOEIC as a language exit test in universities?

Appreciation

Appendix F: Student Interview Schedule

STUDENT INTERVIEW SCHEDULE

Interview purpose

The purpose of this interview is to understand the impact of the TOEIC-as-exit-test on your learning. It also seeks to understand the factors associated with your learning and the test impact.

Interview questions

Personal information

1. How old are you?
2. Where do you come from?
3. How many people are there in your family?
4. What do your parents do? How about your siblings?
5. Why did you choose this major?
6. What do you want to do after graduation?
7. What do you want to achieve in your career?
8. How long have you been learning English?
9. What types of English courses have you taken? What English course are you taking?
10. How would you rate your own English ability on a scale of 1 to 10, with 1 being “very poor” and 10 being “excellent”?
11. Have you ever taken any English certificate tests? If yes, what were your scores?

Beliefs about language learning

12. What are the goals of your English learning? Why?
13. What is the most important part of learning a foreign language (grammar, vocabulary, translation or language use, communication)? Why?
14. How do you understand “communication ability”?
15. Do you think it’s important for university students to learn all four skills? Why?

16. Which skills are most important to learn? Why?
17. What is the best way to learn a language? Why?
18. What is the best way to improve test scores? Why?
19. Is learning test-taking skills important to improve test scores? Why?

Perception of the TOEIC test

20. What is the purpose of the TOEIC test? Why do you think so?
21. Which skills does the test assess?
22. What format does it have?
23. Is the test difficult for you? Which parts of the test are difficult for you and why?
24. Do you think the test accurately reflect students' English proficiency? Why?
25. Do you think the test is useful for university students (for their learning, their career)? In what ways?
26. Do you think the test is suitable for graduation purposes? Why?

Pressure and context factors

27. What will be the consequences for you if you don't get the required score?
28. Is the TOEIC test stressful for you? How? Why?
29. Do your parents/ teachers/ friends expect/urge you to prepare for the TOEIC test?
30. Do you have enough time to study English?
31. Do you have enough resources to study English?
32. What do you think is the most important factor that motivates you to prepare for the TOEIC test?

Test impacts on learning

33. Has the TOEIC test motivated you to learning English (aware of the importance of learning English? more engaged in class room activities? taking English learning more seriously? Spending more time learning English?) In what way?
34. Has the TOEIC test affected your learning? How?

35. Have you been doing test preparation? What have you been doing to prepare for the test? (learning test-taking skills? doing past exam papers/ mock tests/ TOEIC preparation materials? finding information about the test? seeking help from English teachers?)
36. Since the inset of the TOEIC test, have you been spending more time on learning the tested skills than before/ than the non-tested skills? Why?
37. Since the inset of the TOEIC test, have you been spending more time on the non-tested skills than before/ than the tested skills? Why?
38. What have you been doing outside class to prepare for the TOEIC test?
39. Have you been spending more time on language use activities such as reading newspapers, watching TV/ videos, communicating with native speakers or people around you in English than before? Why? How helpful are they?
40. Have you been spending more time on form-focused activities such as learning grammar rules, doing grammar exercises, doing translations, learning by repetition than before? Why? How helpful are they?
41. Has your achievement been improved as a result of your preparation for the TOEIC test? In what ways? (mid-term, final term test score, skills...?)
42. To what extent has the TOEIC test impacted on your learning? Why do you think it has such an effect on your learning?
43. Do you have any difficulties when you prepare for the TOEIC test?
44. Would you study English differently, if the TOEIC test were abolished?

Wrap up

45. Is there anything else you would like to share with me so that I can fully understand the issues related to the use of TOEIC as a language exit test in universities?

Appreciation

Appendix G: Observation Scheme

OBSERVATION SCHEME

Date: _____ Time: _____ Location: _____

University: _____ Class: _____ Teacher: _____

[illegible]

Note. WC = Whole class; PW = Pair work; GW = Group work; IW = Individual work

Appendix H: Weekly Journal Sheet

JOURNAL SHEET

for the week of to....., 2012

Student's name:

When you fill in this journal sheet, please keep your out-of-class English learning in mind.

1) Which of these skills have you been working on in the last week? Why have you been working on them? How much time have you spent on them? Please complete the table with your answers.

Skills	Yes	Time (in hours)	Reasons
Listening	<input type="checkbox"/>		
Speaking	<input type="checkbox"/>		
Reading	<input type="checkbox"/>		
Writing	<input type="checkbox"/>		
Grammar	<input type="checkbox"/>		
Vocabulary	<input type="checkbox"/>		
Pronunciation	<input type="checkbox"/>		

2) What materials have you been using to learn English in the last week? Why have you been using them? Please complete the table with your answers.

	Materials	Reasons
1		
2		
3		
4		
5		

3) What activities have you been doing to learn English in the last week? Why have you been doing them? How much time have you spent on them? Please complete the table with your answers. The first item is an example.

	Activities	Time (in hours)	Reasons
1	<i>I've been reading English newspapers.</i>	<i>2</i>	<i>This helps me learn new words and ...</i>
2			
3			
4			
5			

4) *What are the three most important things that you have learned in the last week?*

.....

.....

.....

.....

.....

.....

5) *Are you happy with what you have achieved in the last week? Why? Why not?*

.....

.....

.....

.....

.....

.....

6) *What are the three most difficult things that your have had to deal with in the last week?*

.....

.....

.....

.....

.....

.....

7) *How have you been feeling about learning English in the last week? Why have you been feeling that way?*

.....

.....

.....

.....

.....

.....

Appendix I: Ethics Approval



Phone 0-4-463 5676
Fax 0-4-463 5209
Email Allison.kirkman@vuw.ac.nz

MEMORANDUM

TO	Thi Thanh Ha Nguyen
COPY TO	Peter Gu Angela Joe
FROM	Dr Allison Kirkman, Convener, Human Ethics Committee
DATE	28 February 2012
PAGES	1
SUBJECT	Ethics Approval: 19118 Impact of TOEIC as a University Exit Exam: a Vietnamese Perspective

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 April 2014. 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

Appendix J: Presentation Topics of TOEIC 5

TOPIC 5.6: JUDGING PEOPLE BY PAPER QUALIFICATIONS

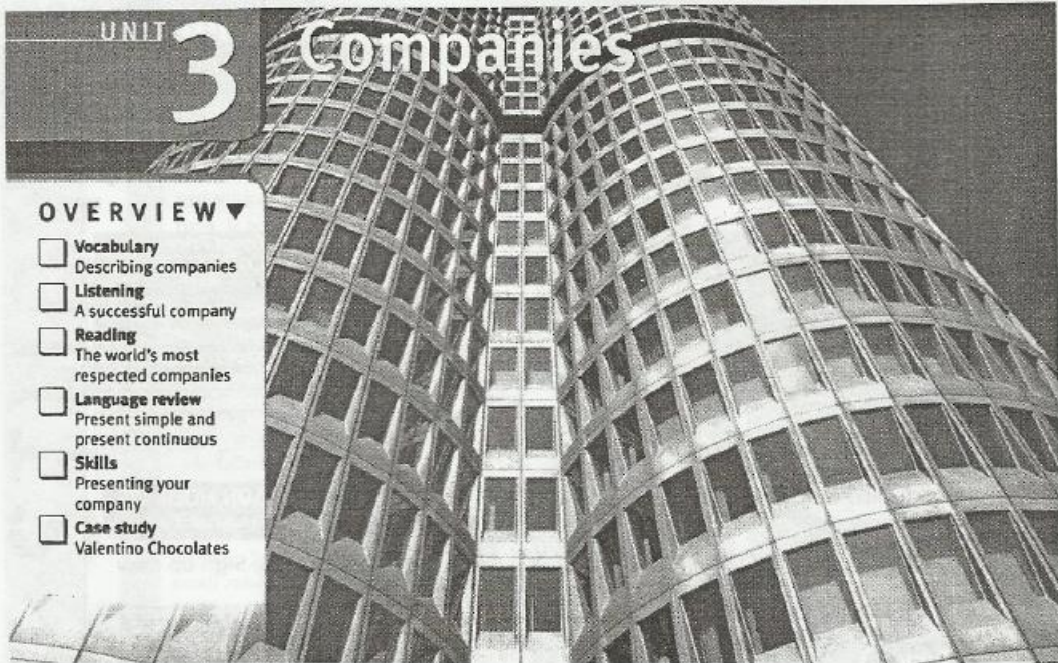
Suggested questions:

1. Do you think one's paper qualification is important to his or her success? Why or why not?
.....
2. Is it always right to judge someone's ability by his or her paper qualification? Why or why not?
.....
3. Why are you attending this university course?
.....
4. Can you name some very successful people without well-qualified papers? In your opinion, what made them successful?
.....
5. Do you think that paper qualifications can be bought? If yes, what would you do to stop people from buying and selling paper qualifications if you were the Prime Minister?
.....
6. Why do some people with high qualifications cannot work well in reality?
.....
7. Can people who have abilities but don't have paper qualifications get good jobs? Why or why not?
.....
8. Why do many young people quit school early nowadays?
.....
9. How do you think about the education and training system in Vietnam nowadays?
.....
10. Can you suggest some solutions to improve the quality of teaching and learning?
.....

Appendix K: Market Leader Pre-intermediate, p. 22

UNIT 3

Companies



OVERVIEW ▼

- ☐ **Vocabulary**
Describing companies
- ☐ **Listening**
A successful company
- ☐ **Reading**
The world's most respected companies
- ☐ **Language review**
Present simple and present continuous
- ☐ **Skills**
Presenting your company
- ☐ **Case study**
Valentino Chocolates

“Be everywhere, do everything, and never fail to astonish the customer.”

Anonymous

Starting up

A Which of these companies would you like to work for? Why?

- 1 a family owned company
- 2 a multinational company
- 3 your own company (be self-employed)

➔ Vocabulary file page 157

B Which of these business sectors do you work in (or would you like to work in)? Can you name a company in each sector?

• Telecommunications / Media	• Transport
• Engineering	• Vehicle manufacturing
• Retailing	• IT (Information Technology) / Electronics
• Construction	• Food and drink
• Tourism	• Pharmaceuticals / Chemicals
• Banking and finance	• Other

Vocabulary
Describing companies

A Complete the sentences below with words and phrases from the box.

share price workforce profit turnover
subsidiary market share head office

- 1 The amount of money a company receives from sales in a particular period is called its *turnover*.
- 2 The money a company makes after taking away its costs is its
- 3 A company which is more than 50% owned by a parent company is called a

multinational company /mʌlti'næʃənəl 'kʌmpəni/: (n) công ty đa quốc gia

family owned company: (n) công ty gia đình

pharmaceuticals /fɑːmə'sjuː tɪkəlz/: (n) dược phẩm

subsidiary /səb'sɪdiəri/: (n) chi nhánh, công ty con

workforce /'wɜːk'fɔːs/: (n) lực lượng lao động, nguồn nhân lực

head office /hed 'ɒfɪs/: (n) trụ sở chính

**Appendix L: Longman Preparation Series for the New TOEIC test:
Introductory Course, p. 270**

Questions 149–152 refer to the following memo.

To: All personnel
From: Marina Petrowski, Director
Re: Travel expenses

We are all aware that the procedures for charging and reporting expenses for business trips taken on behalf of the company have long been out of hand. As a result _____ recommendations from the Budget Office, the following procedures

149. (A) of
(B) as
(C) for
(D) if

will be adopted.

Company personnel will no longer be given company credit cards to cover expenses while on out of town trips. Instead, all travel expenses, with the _____ of

150. (A) excepts
(B) excepted
(C) exception
(D) excepting

airline tickets which will continue to be charged directly to the company, will be paid for out of pocket. In order to receive _____, an expense report must be

151. (A) bonuses
(B) assistance
(C) supervisors
(D) reimbursement

submitted to your department head within ten days of returning from a trip. All charges must be itemized on the report and accompanied by receipts. Approval of each item will be made at the discretion of each department head, following, of course, the company expense guidelines (see attached). Generally, charges for hotels, meals, and transportation will be _____. Non-work related items such as

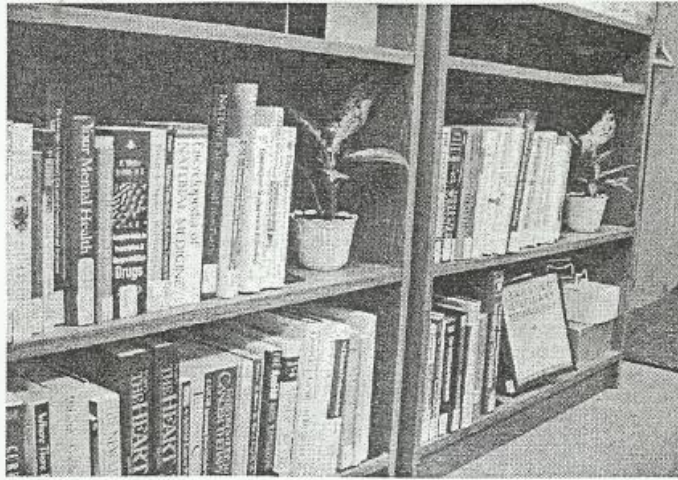
152. (A) reported
(B) expensive
(C) authorized
(D) unallowable

entertainment, excessive taxi rides, and bar bills will not. I am sure you will all understand the necessity of this strict attitude toward expense reporting. I am counting on everyone's cooperation.

Appendix M: TOEIC 5, p. 71

122. When Ms. Myer ----- Beijing for an international conference, she will meet with the regional manager as well.
(A) visits
(B) to visit
(C) was visiting
(D) visiting
123. InfoWorks Technology has been dedicated to offering - products at reasonable prices.
(A) skilled
(B) earnest
(C) tactful
(D) reliable
124. ----- in the entertainment and art industry has dramatically increased in the last few years.
(A) Competition
(B) Competitor
(C) Competitive
(D) Compete
125. ----- a news report, the agricultural industry has recently generated more profits than ever before.
(A) According to
(B) Nevertheless
(C) Even though
(D) As if
126. The city governor is taking ----- measures to shift passenger traffic from road to rail.
(A) signify
(B) significant
(C) significantly
(D) significance
127. The heating system malfunctioned due to a maintenance problem that ----- last night.
(A) emerged
(B) engaged
(C) released
(D) reacted
128. Last Friday, marketing staff from all of the branches ----- at Canaan Valley Resort for a quarterly meeting.
(A) convene
(B) convened
(C) convenes
(D) convening
129. Many ----- scientists around the world were invited to the 20th annual Science Expo held in Bangkok.
(A) definite
(B) customary
(C) notable
(D) elaborate

Appendix N: TOEIC 5, p. 55



8. (A) (B) (C) (D)

