

**The impact of implementing English proficiency tests
as a graduation requirement at Taiwanese
universities of technology**

Hui-Fen Hsu

Presented for the degree of Doctor of Philosophy

University of York

Department of Educational Studies

November 2009

Abstract

Both the Taiwanese government and the community at large recognise that a high level of English language ability among the workforce is essential for success in virtually all walks of life, as Taiwan is heavily dependent on international business for its economic growth. A growing concern for national standards of educational achievement in a competitive global economy, together with a heightened demand for accountability in government expenditure has been reflected in tertiary institutions, where the problem of graduates' English proficiency has been addressed. An innovation, the Graduation Threshold (GT), has been proposed by some universities of technology in order to ensure higher standards of English proficiency among graduating students.

GT is a university policy whereby students are expected to fulfil graduation requirements by obtaining a certain level in, or getting at least a certain score on, an English proficiency test. Tests have long been advocated by policymakers in diverse national contexts as one form of effective leverage of control. The phenomenon of how tests influence teaching and learning is commonly described as 'washback'. Yet while the connection between testing and learning is commonly made, it is not known whether it really exists and, if it does, it is seldom crystal clear whether tests have favourable or deleterious washback. Following a critical review of selected literature on the washback of tests, particularly English language tests, on teaching and learning, the focus of the current study is to determine whether the effort of the concerned Taiwanese universities of technology to implement the GT has changed institutional policies, teaching and learning. If so, are English proficiency tests playing any role in this? If there has not been any marked change, are English proficiency tests in any way responsible for blocking or impeding changes?

The research sites were non-English departments of Taiwanese universities of technology, which were divided into two groups. One of the groups (Group 1) required non-English major students to pass one of a set of English proficiency tests at a specified level as a graduation requirement, whereas the other group (Group 2) did not prescribe any English

graduation requirement. In each group, 27 to 28 teachers and 300 to 321 students completed questionnaires. Two teachers from each group, along with three departmental directors and three advisory committee members within the Taiwanese Ministry of Education, were interviewed. Two lessons taught by each interviewed teacher were also observed.

Results indicated that the policy of implementing English proficiency tests as a graduation requirement had a superficial or at times no impact on teaching for both groups, with a slightly greater impact on Group 1, who complied with their university's policy of English graduation requirement. Although the majority of Group 1 teachers, departmental directors and advisory committee members had generally positive attitudes towards the policy, teachers' fundamental beliefs about English language teaching and learning were not changed. The new policy influenced what the teachers taught, but not how they taught. In addition, the teachers, departmental directors and advisory committee members pointed out several issues and problems with the diffusion and implementation of the educational innovation. The teachers and educational administrators nevertheless were aware of the problems they currently faced and appeared determined to resolve them.

The results seem to argue against using English proficiency tests as a degree requirement or for other gatekeeping purposes. Guidelines are proposed for those universities which want to adopt the English proficiency tests for these high-stakes purposes.

Keywords: washback; attitudes and perspectives of English instructors in Taiwan; English proficiency tests; graduation requirement; attitudes and perspectives of students of English in Taiwanese universities of technology

Table of Contents

Abstract	2
Table of Contents.....	4
List of Tables.....	8
List of Figures.....	10
Acknowledgement.....	12
Chapter One: Introduction.....	13
1.1 The problem	13
1.2 Context and purposes of the study	15
1.3 Research questions.....	17
1.4 Overview of research methodology	18
1.5 Definition of key terms.....	19
1.6 Significance of the study	21
1.7 Organisation of the thesis	22
Chapter Two: The Taiwanese Research Context.....	24
2.1 Introduction	24
2.2 Context of the study.....	24
2.2.1 Education in Taiwan.....	24
2.2.2 English language education at tertiary level in Taiwan	27
2.2.3 English instructors in universities of technology	29
2.3 Background of the study.....	30
2.3.1 National Development Plan.....	30
2.3.2 The project on the Enhancement of Students' Foreign Language Proficiency	34
2.4 Implications for the present study	41
Chapter Three: Literature Review.....	44
3.1 Introduction	44
3.2 The definition of washback	44
3.3 Washback: positive, negative, neither or both.....	46
3.4 The mechanism of washback.....	50
3.5 Empirical studies exploring washback on teaching and learning.....	55
3.6 Tests as a means of control	61
3.7 Educational innovation.....	68
3.8 Exploring washback within research models of teaching and learning	78
3.8.1 Classroom research.....	78
3.8.2 Classroom research models.....	82
3.9 Summary.....	85

Chapter Four: Research Methodology	87
4.1 Overall research design.....	87
4.1.1 Research questions.....	87
4.1.2 Research framework.....	89
4.1.3 Methodological considerations from other washback studies.....	91
4.1.4 Research methodology for this study.....	92
4.1.5 Research phases.....	98
4.2 Research design for the pilot study and main study – the two questionnaires.....	100
4.2.1 Teacher questionnaire.....	100
4.2.2 Student questionnaire.....	105
4.2.3 Sampling for both teacher and student surveys in the main study.....	110
4.2.4 Data collection procedures for the main study.....	111
4.2.5 Data analysis procedures for the main study.....	112
4.3 Research design for the pilot study and main study – classroom observation.....	113
4.3.1 Rationale for the classroom observation study.....	113
4.3.2 Research focus on classroom teaching.....	115
4.3.3 The observation scheme.....	116
4.3.4 Data collection and analysis procedures.....	120
4.3.5 Reliability of classroom observations.....	121
4.4 Research design for the pilot study and main study – interviewing.....	122
4.5 The ethical issues in the study.....	124
4.6 Summary.....	127
Chapter Five: Teachers’ Perceptions of the New Policy	128
5.1 Introduction.....	128
5.2 Demographic information on the teachers.....	129
5.3 Teachers’ reactions to and perceptions of the new policy.....	131
5.3.1 Teachers’ reactions to the new policy.....	131
5.3.2 Teachers’ perceptions of the reasons behind imposing the new policy.....	132
5.3.3 Alterations to teaching methods due to the new policy.....	134
5.4 Washback on teaching materials.....	136
5.4.1 Textbook arrangements related to teaching materials.....	136
5.4.2 Teaching and learning resources.....	137
5.5 Washback on classroom teaching behaviour.....	139
5.5.1 Teaching planning and medium of instruction.....	139
5.5.2 Teacher talk and teaching activities in General English lessons.....	141
5.6 Washback on assessment and evaluation.....	144
5.6.1 Factors that influence teaching.....	144
5.6.2 The use of mock exams.....	145
5.7 Washback on teachers’ attitudes towards aspects of learning.....	146
5.7.1 Learning aims.....	147
5.7.2 Learning strategies.....	148

5.8 Summary.....	149
Chapter Six: Students' Perceptions of the New Regulation.....	152
6.1 Introduction.....	152
6.2 Students' characteristics and their learning contexts.....	153
6.3 Students' attitudes towards teaching activities inside and outside class.....	157
6.3.1 Teachers' talk in class.....	158
6.3.2 Students' perceptions of their teachers' teaching activities in class.....	160
6.3.3 Students' classroom learning activities.....	161
6.3.4 Students' use of English in relation to their classroom activities.....	165
6.3.5 Students' use of English in relation to their learning activities outside class.....	167
6.4 Students' attitudes towards aspects of learning.....	169
6.4.1 Students' motivation to learn English.....	169
6.4.2 Students' preferred learning strategies.....	172
6.4.3 Students' attitudes towards whether and in which aspects they are influenced by the English proficiency tests.....	173
6.5 Students' attitudes towards aspects of examinations.....	177
6.6 Summary.....	181
Chapter Seven: Teachers' and Students' Actions and Reactions.....	184
7.1 Introduction.....	184
7.2 Choosing the sample.....	185
7.3 Classroom observation scheme.....	188
7.3.1 Rationale behind the observation scheme.....	188
7.3.2 The observation scheme.....	189
7.3.3 Observation scale.....	191
7.4 Classroom observation outcomes.....	192
7.4.1 Participant organisation.....	192
7.4.2 Activity types and content as percentage of class time.....	195
7.4.3 Teaching materials.....	197
7.4.4 Further analysis.....	198
7.5 Classroom observations of individual teachers' lessons and interviews.....	198
7.6 Summary.....	206
Chapter Eight: Advisory Committee Members' and Departmental Directors' Reactions.....	209
8.1 Introduction.....	209
8.2 The Taiwanese educational context at the macro level.....	209
8.2.1 Interviews with advisory committee members.....	211
8.3 The Taiwanese educational context at the micro level.....	218
8.3.1 Tertiary institutions.....	218
8.3.2 Department administration.....	219
8.4 The major issues emerging from the interviews.....	228

Chapter Nine: Discussion	232
9.1 Introduction	232
9.2 Synthesis of the findings.....	233
9.2.1 Research question 1: What strategies did the Taiwanese Ministry of Education and university departments use to implement the Enhancement of Students' Foreign Language Proficiency Project?.....	233
9.2.2 Research question 2: What was the nature and scope of the washback effect on teachers' and students' perceptions of aspects of teaching in relation to English proficiency tests?	235
9.2.3 Research question 3: What was the nature and scope of the washback effect on teachers' behaviours as a result of implementing the English proficiency tests?.....	242
9.3 Review of the washback model in the Taiwanese educational context.....	245
 Chapter Ten: Conclusion	 250
10.1 General aims of the study	250
10.2 Implications of the study	252
10.2.1 Implications for making an ideal graduation test.....	254
10.2.2 Implications for teaching and learning.....	256
10.3 Limitations of the study.....	262
10.4 Future research directions.....	265
 Appendices	 269
Appendix 1 Comparative graduation English language proficiency benchmarks from various institutions.....	269
Appendix 2 Common Reference Levels.....	270
Appendix 3 Questionnaires and interview schedules for the pilot study	271
Appendix 4 Questionnaires and interview schedules for the main study.....	289
 References	 306

List of Tables

Table 4. 1 Structure and themes of the teacher questionnaire in the pilot study.....	102
Table 4. 2 Structure and themes of the student questionnaire in the pilot study.....	107
Table 4. 3 Classroom observation scheme.....	117
Table 4. 4 Modified Brown Interaction Analysis System.....	120
Table 5. 1 Characteristics of the respondents (TQ1-5).....	129
Table 5. 2 Teachers' reactions to the new policy (TQ14).....	131
Table 5. 3 Group 1 teachers' perceptions of the reasons behind imposing the new policy (TQ8)	133
Table 5. 4 Alterations to teaching methods (TQ10).....	134
Table 5. 5 Textbook arrangements in Taiwanese universities of technology (TQ16).....	136
Table 5. 6 Teaching and learning resources (TQ20).....	137
Table 5. 7 Chi-square tests of differences in aspects of daily teaching (TQ15 and 17).....	140
Table 5. 8 Teacher talk in General English lessons (TQ18).....	141
Table 5. 9 Teachers' teaching activities (TQ19).....	142
Table 5. 10 Factors that influence teaching (TQ12).....	144
Table 5. 11 The use of mock exams (TQ13).....	146
Table 5. 12 Teachers' perceptions of students' major reasons for learning English (TQ6).....	147
Table 5. 13 Learning strategies that teachers had recommended to students (TQ11).....	149
Table 6. 1 Students' demographic information and their learning contexts (SQ1-7).....	153
Table 6. 2 Differences in students' perceptions of their teachers' talk (SQ8).....	159
Table 6. 3 Students' perceptions of their teachers' teaching activities (SQ9).....	161
Table 6. 4 Students' perceptions of their classroom activities (SQ10).....	162
Table 6. 5 Differences in the students' classroom activities between Year1 and Year2 (SQ10)	163
Table 6. 6 Students' use of English in relation to their classroom activities (SQ11).....	166
Table 6. 7 Students' use of English in relation to their learning activities outside class (SQ12)	

.....	167
Table 6. 8 Students' perceptions of their reasons for learning English (SQ13)	170
Table 6. 9 Students' preferred learning strategies (SQ14).....	173
Table 6. 10 Students who had and had not been affected by English proficiency tests (SQ15)	174
Table 6. 11 Aspects in which students are affected by the English proficiency tests (SQ16)..	175
Table 6. 12 Students' attitudes towards English examinations (SQ17).....	178
Table 7. 1 Classroom observation scheme for the main study (sample sheet).....	190
Table 7. 2 Interaction analysis system for the main study	191
Table 7. 3 Participant organisation of lessons	193
Table 7. 4 Classroom activities of all four teachers' lessons as a percentage of class time	195
Table 7. 5 Summary of the observations of the four teachers.....	200

List of Figures

Figure 1. 1 Overview of the research methodology.....	19
Figure 2. 1 Taiwan's education system.....	26
Figure 3. 1 A basic model of washback.....	51
Figure 3. 2 Variables for classroom research.....	82
Figure 3. 3 A language teaching analysis scheme.....	83
Figure 3. 4 Explanatory model of researching washback in the present study.....	85
Figure 4. 1 The Taiwanese education system relating to the English proficiency tests.....	93
Figure 4. 2 Data and method triangulation for the study.....	97
Figure 4. 3 The three phases of the research study.....	99
Figure 4. 4 Stages of the teacher questionnaire design.....	105
Figure 4. 5 Stages of the student questionnaire design.....	110
Figure 5. 1 Teaching and learning resources (TQ20).....	138
Figure 5. 2 Reported medium of instruction in Taiwanese universities of technology (TQ17)	140
Figure 5. 3 Teacher talk in General English lessons (TQ18).....	142
Figure 6. 1 Students' year of study (SQ2).....	154
Figure 6. 2 Medium of instruction that teachers use in the classrooms (SQ4).....	155
Figure 6. 3 Students' study hours per week (SQ5).....	156
Figure 6. 4 Students' acknowledgement of the regulation (SQ6).....	157
Figure 6. 5 Students' perceptions of their teachers' talk (SQ8).....	158
Figure 6. 6 Students' perceptions of their teachers' teaching activities (SQ9).....	160
Figure 6. 7 Students' perceptions of their classroom activities (SQ10).....	163
Figure 6. 8 Differences in the students' classroom activities between Year1 and Year2 (SQ10).....	164
Figure 6. 9 Students' use of English in relation to their classroom activities (SQ11).....	165
Figure 6. 10 Year1 and Year2 students' use of English in relation to their classroom activities (SQ11).....	166
Figure 6. 11 Learning reasons – To satisfy university requirements (SQ13).....	170

Figure 6. 12 Learning reasons– To meet society's expectations (SQ13)	171
Figure 6. 13 Learning reasons– To be able to communicate with foreigners (SQ13).....	171
Figure 6. 14 Students who were and were not affected by the English proficiency tests (SQ15).....	174
Figure 6. 15 Aspect affected– Anxiety and emotional tension (SQ16).....	175
Figure 6. 16 Aspect affected– Less interested in learning English (SQ16).....	176
Figure 6. 17 Attitudes towards English examinations– Taking examinations is a valuable learning experience (SQ17).....	179
Figure 6. 18 Attitudes towards English examinations– Every university graduating student should pass at least one English proficiency test (SQ17).....	180
Figure 6. 19 Attitudes towards English examinations– Students' learning is improved by practising mock exams (SQ17).....	180
Figure 6. 20 Attitudes towards English examinations– Examinations increase students' motivation to learn English (SQ17).....	181
Figure 9. 1 Model of the effect of washback within the Taiwanese educational context.....	248

Acknowledgement

I would like to take this opportunity to extend my deepest gratitude and appreciation to all of those people – over 650 students, teachers, departmental directors and advisory committee members – who have contributed to this research project by completing questionnaires, and taking part in interviews and classroom observations. Without their help, the completion of this research project would not have been possible.

I am deeply indebted to my supervisor and members of my Thesis Advisory Group for providing me with enlightenment, support and encouragement. I would like to thank my supervisor, Dr. Graham Low for his remarkable supervision, brilliant guidance, enthusiasm and invaluable support throughout the research, which I am sure was above and beyond the call of duty. I regard myself as extremely fortunate to have had the chance to learn from him. I would also like to thank Dr. Emma Smith for her great inspiration and encouragement. My sincere thanks too to Dr. Vanita Sundaram for her helpful advice, detailed comments, suggestions and ongoing encouragement. I could not have wished for better supervision and support than these three people provided.

Very special thanks to Ms Zenda Nash for proofreading and suggestions. I wish to thank all of my fellow PhD students at the University of York, particularly my colleagues from Taiwan for their friendship, their understanding and their support.

Finally, I would like to extend my heartfelt appreciation and thanks to my parents, my sister and my husband for standing by me and for giving me their unconditional love, care, good will assistance and patience.

Chapter One: Introduction

1.1 The problem

Enchanted by the power of standardised testing, politicians, policymakers or educators time after time come out strongly in favour of adopting tests as a means to improve teaching and learning, or to contribute to intended top-down educational changes (Alderson & Wall, 1993; Spolsky, 1997; Wall, 1997). Using tests as a lever of educational change is commonly believed to have an impact on teaching, learning, education systems and society, a phenomenon described as ‘backwash’ in general education or as ‘washback’ in language instruction, and the impact is often taken for granted. Yet while the connection between testing and learning is commonly made, it is not known whether it really exists and, if it does, it is seldom crystal clear whether tests have favourable or deleterious washback. A great body of studies related to the effects of language tests, as well as the relationships between tests and language instruction, has been published in recent decades (e.g. Alderson & Hamp-Lyons, 1996; Cheng, 2005; Shohamy, 1993), but no definitive influence has been reported (e.g. Alderson & Hamp-Lyons, 1996; Watanabe, 1996). Indeed, the literature has repeatedly indicated that testing washback is a complex concept. Some studies conclude that no simple washback effect occurs (Alderson & Hamp-Lyons, 1996; Watanabe, 1996), whereas others find language testing to be a powerful determiner of classroom teaching, influencing topics such as curricular scope and sequence, and choices of teaching materials (Herman & Golan, 1991; Cheng, 2005).

As Taiwan is heavily dependent on international business for its economic growth, English is deemed to be an indispensable tool for success in virtually all walks of life. Both the government and the community at large recognise that a high level of English language ability among the workforce is essential. In the last few decades, there has been a growing

concern for national standards of educational achievement in a competitive global economy, together with a heightened demand for accountability in government expenditure. This concern has been reflected in tertiary institutions, where the problem of graduates' English proficiency has been addressed. The concern emerges as a consequence of mounting dissatisfaction with graduates' linguistic ability, both internally from teaching staff of educational institutions, and externally from the public, especially graduates' employers. An innovation has been proposed in order to ensure higher standards of English proficiency among students graduating from Taiwanese tertiary institutions.

In 2002, the Taiwanese Ministry of Education introduced in universities of technology 'The Grant Project on the Enhancement of Students' Foreign Language Proficiency'. The goal of the project was to strengthen students' English proficiency throughout the educational system. Accordingly, every university of technology conducted an organisational English project to reflect the grant project. One common feature of those organisational English projects was to impose graduation standards control; that is, to require graduates to achieve certain minimum standards before they could graduate from their study programmes. It appeared obvious to many tertiary institutions that integrating graduation standard control into study programmes was essential for assuring the quality of institutions' qualifications. There were and are several means to achieve the goal. Most universities of technology adopted standardized English proficiency tests, either self-developed tests or external public tests, as the means of standards control. This solution is not surprising in the Taiwanese context, where human resources are cultivated and screened through fierce competition. It is not going too far to say that university graduation determines the future social status of each student, which is why performing well on such tests is considered crucial to future success.

With Taiwan's measurement-led system, the innovation – passing English proficiency tests as a graduation requirement – was expected to have an impact on what and how teachers taught and what and how students learned, thereby promoting university of technology students' English skills. It was a deliberate attempt to influence the teaching of English in a positive direction, by moving away from the traditional, teacher-dominant classroom towards

a more active learning context, where students had sufficient opportunity to practise the skills the English proficiency tests hoped to encourage. Petrie (1987) stated that “it would not be too much of an exaggeration to say that evaluation and testing have become the engine for implementing educational policy” (p. 175). The following questions, however, need to be examined: Is introducing changes through tests effective? Can the introduction of tests *per se* cause real improvement in learning and teaching? How are test results used by teachers, students and administrators? The answers remain unclear. The focus of the current study is thus to determine whether the effort of the Taiwanese universities of technology to implement an English graduation requirement has changed the English teaching in the institutions concerned. If so, are English proficiency tests playing any role in this? If there has not been any marked change, are English proficiency tests in any way responsible for blocking or impeding changes?

1.2 Context and purposes of the study

As mentioned in the previous section, the graduation requirement involves using several English proficiency tests as a gatekeeping device. The primary purpose of the regulation is to measure the attainment in English of students who have completed a full-time university course of four years' duration. The English proficiency tests are taken by the majority of students by the final year of university. Students either proceed to further studies at postgraduate programmes or leave university and seek employment.

The recognised English proficiency tests are the General English Proficiency Test (GEPT), the College Student English Proficiency Test (CSEPT), the Test of English as a Foreign Language (TOEFL), the International English Language Testing System (IELTS), the Test of English for International Communication (TOEIC), the Cambridge ESOL Exams, and tests developed in-house by particular universities. The students are free to take whichever of the tests they choose, but they have to achieve above the A2 level defined in the Common

Reference Levels outlined in the CEFR – *Common European Framework of Reference for Languages* (Council of Europe, 2001).

Due to the decentralisation of university governance and academic freedom, universities of technology have the right to decide on whether or not to make the English graduation requirement compulsory. I, therefore, conducted a comparative study in non-English departments in Taiwan. The participants were divided into two groups. The most salient difference between the two groups was the English graduation requirement; Group 1 students were required to pass one of the English proficiency tests as one of their degree requirements, whereas there was no such requirement for Group 2 students. The study records the aspirations and concerns of some of those who have been most closely involved in their university's evolution in relation to the teaching and learning of English. These voices echo throughout the study, as it becomes clear that many of the issues they highlight are also issues, in one form or another, that are of concern to others – both educationalists and the general public – throughout Taiwan.

The general purpose of this study is to investigate how those involved, directly and indirectly, in teaching and learning in Taiwanese universities of technology are affected by, and perceive themselves to be affected by, the implementation of the English graduation requirement. It is an attempt to shed light on different stakeholders', particularly teachers' and students', conceptualisations of the implementation of English proficiency tests as a graduation requirement. The specific purposes of this study are as follows.

The first purpose is to investigate the phenomena of the washback effect in the light of using English proficiency tests as a gatekeeping device. Given that more and more universities of technology have imposed the English graduation regulation and there has been a paucity of relevant studies, this subject deserves further investigation. I have focused on the washback on non-English major students, because a large number of them have been asked to pass a certain level of English proficiency tests as one of their degree requirements, even though they do not major in English and have less English input compared with English majors. The impact of the English graduation regulation on the non-English major students'

English language acquisition will also be examined.

The second purpose is to understand how the main participants within the Taiwanese educational context reacted to the implementation of English proficiency tests as a graduation requirement in universities of technology. More specifically, the study aims at offering teachers and students an opportunity to voice their concerns and opinions about the graduation regulation and to inform the government authorities and university authorities of teachers' and students' opinions.

The third purpose is to explore the nature and scope of the washback effect on aspects of institutional policies, teachers' and students' perceptions, and teachers' behaviours, within the context of the English graduation regulation. I argue that educational policies should be taken into account because they may have an impact on teachers' teaching and students' learning. In addition, Bailey's (1996) basic model of washback shows that teaching has a direct impact on learning. Teaching or university policies may be integral to the graduation requirement's washback effects.

1.3 Research questions

Based on the research purposes, the washback effect of the English graduation requirement was observed both at the macro level with respect to major parties within the Taiwanese educational context, and at the micro level in universities of technology with regard to different facets of classroom teaching and learning. It is important to emphasise that both teaching and learning were studied in this project, as both of these constructs occur interactively in the classroom. Therefore, both teachers and students were included in the study. However, aspects of learning and learners were studied only when they related to classroom teaching.

The study explored the following research questions:

1. What strategies did the Taiwanese Ministry of Education (MOE) and university

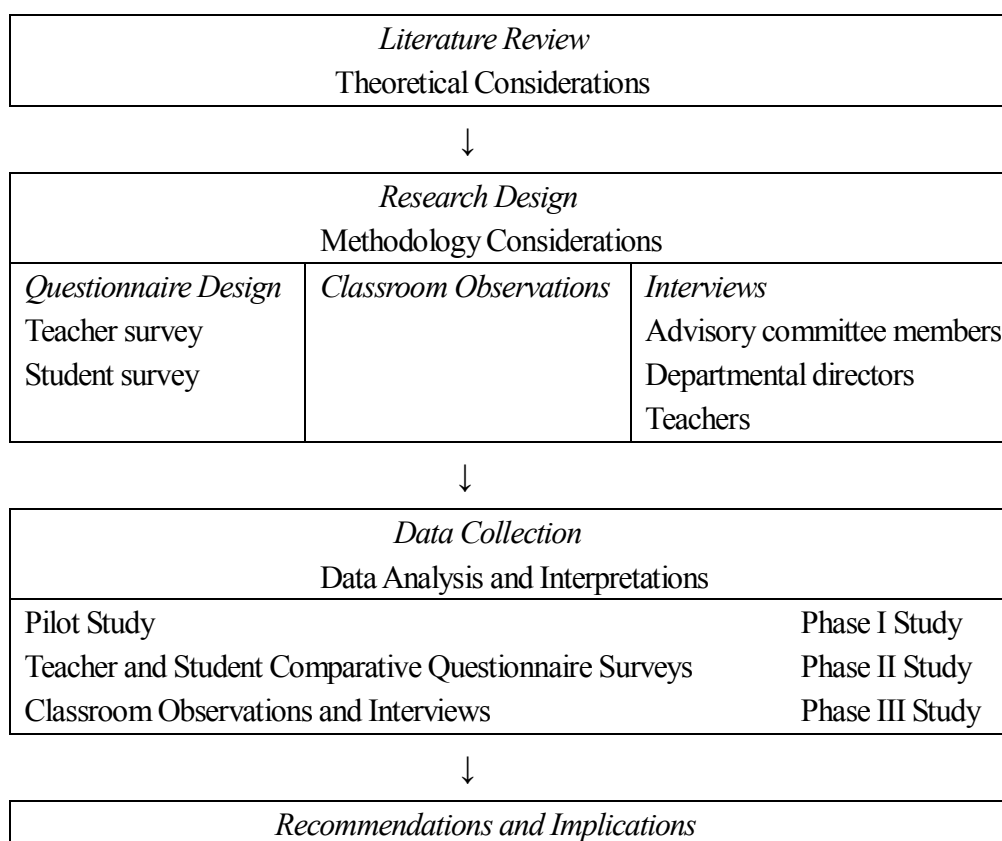
departments use to implement the Enhancement of Students' Foreign Language Proficiency Project?

2. What was the nature and scope of the washback effect on teachers' and students' perceptions of aspects of teaching in relation to English proficiency tests?
3. What was the nature and scope of the washback effect on teachers' behaviours as a result of implementing the English proficiency tests?

1.4 Overview of research methodology

To answer the research questions, a combined framework was employed. First, the study stressed the importance of context, setting, and subjects' frames of reference (Dunkin & Biddle, 1974; Marshall & Rossman, 2006). Second, it emphasised the importance of multiple perspectives on the research problem, as well as the characteristics of policymakers, departmental directors, teachers and students (in line with Stern, 1989). Third, the study used a mixed-methods strategy for obtaining both quantitative and qualitative data (Denscombe, 2007). Figure 1.1 presents an overview of the research methodology in relation to this study.

Figure 1. 1 Overview of the research methodology



1.5 Definition of key terms

For the purposes of this study, the following research terms are operationally defined.

Washback in applied linguistics (Alderson & Wall, 1993) or *backwash* in general education (Biggs, 1996) is a common notion in educational and applied linguistics literature. *Washback* is defined as “the impact of a test on teaching, and ... tests can be powerful determiners, both positively and negatively, of what happens in classrooms” (Wall & Alderson, 1993: 41). The term *washback* is preferred and used throughout this study to indicate both intended and unintended directions and functions of curriculum change by means of English proficiency tests. I have, however, retained the use of the term *washback* or *backwash* in its original form when quoting directly from authors.

Proficiency tests refer to tests that are designed to measure people’s ability in a language

(Hughes, 2003: 44). Major proficiency tests to fulfil the assessment purpose in this study are the General English Proficiency Test (GEPT), the College Student English Proficiency Test (CSEPT), the Test of English for International Communication (TOEIC), the Test of English as a Foreign Language (TOEFL), the International English Language Testing System (IELTS), and the Cambridge ESOL Exams. All these proficiency tests, except for TOEIC and CSEPT, have a mixture of receptive and productive components: listening, reading, speaking and writing. Students at a university of technology have to take one of these English proficiency tests before graduation and demonstrate that their English ability equals the Common European Framework (CEF) Waystage (A2) level or above.

Curriculum change usually refers to pedagogical and policy-based changes, often initiated by an educational body. Curriculum changes also occur at the fundamental level when teachers feel that the curriculum they are following requires reform. In this study, the term *curriculum change* refers to those changes to the curriculum that are driven by assessment; that is, as a result of the implementation of English proficiency tests and a trend towards adopting English proficiency tests as a means of encouraging learning.

Classroom teaching is where “the classroom can be defined as a place where more than two people gather together for the purpose of learning, with one having the role of teacher” (Tsui, 1995: 1). Teaching and learning are studied together as they occur interactively in the classroom. In this study, classroom teaching is defined and studied at the following levels:

- basic theoretical or philosophical level
- policy level
- behavioural level

The study accordingly investigates the washback effect of English proficiency tests upon the above three levels of classroom teaching.

An *integrated approach* is where a task is referred to as “some kind of activity designed to engage the learner in using the language communicatively or reflectively in order to arrive at an outcome other than that of learning a specific feature of the language” (Ellis, 1994: 595). An integrated approach can refer to a real life activity or a pedagogical activity (Nunan, 1989).

In this study, an integrated approach to language teaching and learning involves learners in using English communicatively and reflectively, and involves the integration of the four major language skills to carry out real-life type tasks in the classroom. Such an approach is the driving force behind the policy of implementing English proficiency tests as a graduation requirement.

1.6 Significance of the study

This was a medium-scale empirical study of washback in Taiwan. It was one of the few washback studies that has employed both quantitative and qualitative data. Based on both quantitative and qualitative data, this study provides solid research evidence to explain and predict the washback effect of the English graduation requirement on various aspects of teaching and learning, and on the Taiwanese educational system as a whole. Although this investigation provides data on and evidence of the washback effect in a specific educational context, it should also contribute to the general understanding of washback in education. The study is potentially significant in that it offers educators and policymakers in Taiwan insights into English language teaching and learning in Taiwanese universities of technology. Most importantly, the study highlights the voices of teachers and students, the very people at the centre of the teaching and learning process.

Haladyna, Nolen and Haas (1991) noted that “to judge the value of an outcome or end, one should understand the nature of the processes or means that led to that end. It is not just that means are appraised in terms of the ends they lead to, but ends are appraised in terms of the means that produce them” (p. 6). Concurring with Haladyna et al., Gee, Hull and Lankshear (1996) appealed to researchers and educators to “confront directly, at a fundamental level [...] the nature of language, learning, and literacy in and out of school” (p. 158). The findings of this study should provide important information to help the educational parties involved in Taiwanese English language education improve the policy, practice and

implementation of the English graduation requirement.

Since more and more Taiwanese universities are starting to implement the English graduation requirement, further washback studies can be expected to follow. The present study has employed both quantitative and qualitative data. It is hoped that the study can provide a starting point for future researchers to find the most appropriate method for their own contexts. It would facilitate other research on washback and allow easier comparison of the results between the studies. The questions asked in the various instruments, questionnaires, interviews and the classroom observation scheme, drew on theoretical considerations in the areas of language teaching and learning along with interviews with relevant stakeholders in Taiwan. The instruments are, therefore, easily applicable to future studies conducted in other universities of technology in Taiwan. Moreover, with just a slight adaptation, they could easily be modified to collect data on language learning policies in Taiwanese universities other than universities of technology. The questions are particularly reusable in that they were formulated in English. This means that the set of instruments is readily available for use much more widely than would otherwise be the case since English is the most widely studied, read and spoken foreign language in the world (Kachru & Nelson, 1996; Richards & Rodgers, 2001). The techniques involved could easily be employed with small modifications by many universities in other countries.

1.7 Organisation of the thesis

The thesis is divided into ten chapters. In chapter one, I begin by stating the general problem, giving the rationale behind the study, setting out research questions, describing the significance of the study and an outline of the thesis. Chapter two then outlines some of the background factors relevant to the current situation. Chapter three reviews theories and empirical studies germane to the subject of washback and demonstrates the need to conduct research to investigate the washback of the English graduation requirement. Chapter four

explains the basis for the three research questions and delineates the sampling procedure and the development of the research instruments, including the procedures for validating them and increasing their reliability, the procedures for collecting relevant data and the methods of data analysis. Chapters five, six, seven and eight present the results of the various analyses. Chapter nine compares the results across the analyses in order to answer the research questions and discusses the findings. The thesis concludes in chapter ten with a discussion of the educational implications of the study, an acknowledgement of its limitations, and finally a number of suggestions for future research directions.

Chapter Two: The Taiwanese Research Context

2.1 Introduction

Improving national proficiency in English forms a main focus of educational policy in many countries (Graddol, 2006: 70). For example, “by the end of 2005, Thailand, the Philippines, Japan, and Taiwan were all expressing grave anxiety about their national proficiency in English and had announced new educational initiatives” (Graddol, 2006: 95). This chapter begins by providing background information on the Taiwanese educational context and on the present study. It addresses the main features of the Taiwanese education system and the English language teaching context, and describes the situation of English language instructors in Taiwanese universities of technology. The rationale behind the present study is then described and discussed.

2.2 Context of the study

2.1.1 Education in Taiwan

Taiwan’s education system has moved from being a highly selective, elite system to one providing nine years of compulsory education. In 1968, at the end of the 38 years of martial law, the compulsory education system was extended from six to nine years. The nine-year compulsory education system provides six years of primary education and three years of junior high school education. Once compulsory education has been completed, a distinction is made between academic education and vocational education systems (see Figure 2.1). Junior high school graduates, mostly at the age of 15, sit for a set of competitive national examinations, together with applications, interviews and further institution-specific examinations, in order to attend three years of senior high school¹, followed by four-year

college or university programmes or seven years at medical school. This track of learning is called the academic education system. The educational goal is to nurture high quality specialists with expertise and international vision (Ministry of Education, 2008).

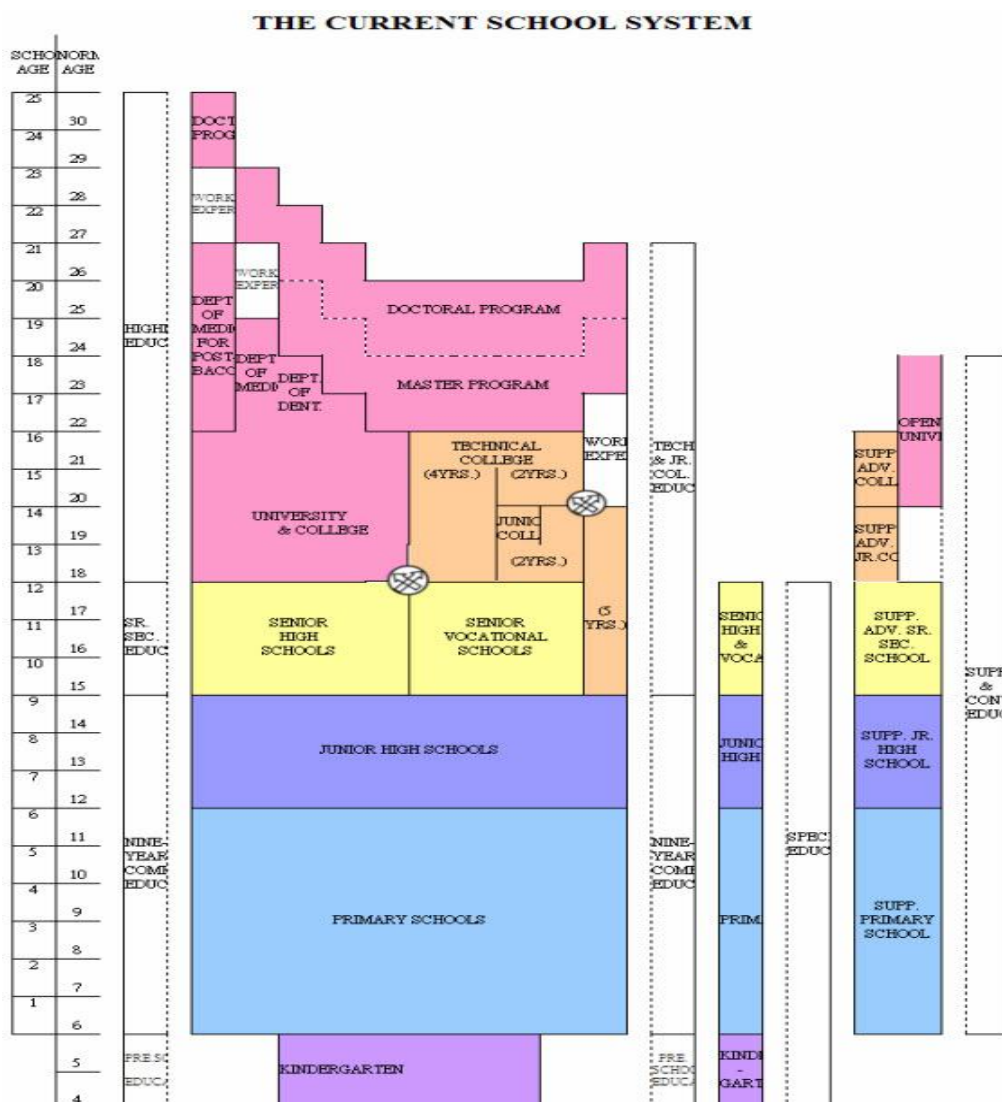
The other option for junior high school graduates is entering the vocational education system, which is the focus of the present study. The purpose of vocational education is to develop technical manpower for economic development, which is considered more practical and is directly linked to the needs of the job market. As shown in Figure 2.1, there are three paths for students to follow to acquire their Bachelor's degrees in the vocational education system. Junior high school graduates can choose to study for three years at a senior vocational high school², followed by a four-year college or university programme. Another path is that senior vocational high school graduates can pursue their studies in a two-year junior college programme, followed by a two-year college programme³. Finally, junior high school graduates can enrol in a five-year junior college programme⁴, and an optional further two-year course.

In the vocational education system, five-year junior colleges are at the lower level of higher education and provide two-year and five-year junior college programmes. Institutes of technology and universities of technology are at the higher level of higher education, and offer four-year college programmes and two-year college programmes. On completion of degree-level study in both tracks, students may enter postgraduate programmes leading to Masters and then doctoral programmes (Ministry of Education, 2008).

Because the Private School Act in 1974 provided the legal basis for establishing private schools and due to the demand created by economic and social development, the private sector commitment to education has expanded rapidly. The current number of private colleges and universities is more than 60% of the total (Ministry of Education, 2008). Indeed, the proportion of colleges and universities of technology has increased and the upgrade of five-year junior colleges to universities of technology has accounted in great part for the increase. These universities of technology primarily accept senior vocational high and junior college graduates and they have been established to train high-level technical personnel. By

2008, there were 162 colleges and universities with a student population of 1.3 million (Ministry of Education, 2008). Of these colleges and universities, there were 38 universities of technology, 9 public and 29 private (Technological and Vocational Education, 2008). The specialist areas involved are generally classified into eight categories: engineering, design, agriculture, management, marine technology, home economics, medicine and nursing, and commerce.

Figure 2.1 Taiwan's education system
(Source: Ministry of Education, 2008)



Education in Taiwan is centrally administered by the Ministry of Education, which is responsible for formulating education policy, as well as for overseeing the operations of all national schools and colleges, national social education organisations, and private colleges and universities. By the 1990s, Taiwanese universities were governed by the central government, who monopolised the administration, development and funding of higher education (Law, 2004). With the easing of martial law in the early 1990s, the Taiwanese government initiated a series of educational reforms. The major concern for higher education reform included: (1) the liberalisation of university governance, (2) the mobilisation of non-government sectors in higher education provision, and (3) the establishment of a quality education system for higher education (Council on Education Reform, 1996). The educational reforms have provided a greater measure of institutional and personal choice than before.

2.2.2 English language education at tertiary level in Taiwan

Taiwan is located near the southeastern coast of China, midway between Japan and Korea to the north and Hong Kong and the Philippines to the south. Its population is approximately 23 million with Mandarin as the official language and Taiwanese and Hakka as widely spoken dialects⁵. Taiwan's economy developed rapidly following World War II. In the 1950s, Taiwan's economy was labour-intensive, with 90% of exports relating to agriculture or the food industry; in the 1960s and 1970s the emphasis moved to manufacturing; transforming in line with the global market of the 1980s and 1990s, major export earnings now related to high technology and chemical products. In 2002, Taiwan joined the World Trade Organisation (WTO), providing official links to international markets, but also leading to an increase in business competition. By 2007, Taiwan had become the 16th largest exporter and the 17th largest importer in the world, and it had the fifth largest foreign exchange reserves (Government Information Office, 2008). As economic globalisation has increased, English has served a unique and important role, since it is the main medium for international relations

and use of the internet. The primacy of English has influenced English-language education policies in Taiwan, and conversely the promotion of English in Taiwan has been driven largely by economic needs (Tsao, 2001).

For decades, English as a foreign language has been a compulsory subject for students in junior and senior high schools, and in tests to allow students to enter a senior high school or a university. In secondary education, English is allotted from three to six hours per week in a semester⁶ depending on students' years of study and their aptitude, with more hours allotted to seniors who are preparing to go into academic, rather than vocational, tracks. Compared with students in academic tracks of education, senior vocational high school students are required to take English instruction for only two hours per week in each semester for the first two years and take conversational English for two hours per week in the final two semesters (Ministry of Education, 2008). In addition to attending English classes in school, many students, particularly in the metropolitan areas, take English courses at cram schools, which are independent after-school learning establishments. At the tertiary level, English is an obligatory subject for all majors. Non-English majors are required to take the college English course for at least two years. The major change in English language education in Taiwan was that English was officially taught as a subject at Grade 5 of primary schooling from 2001, and was introduced at Grade 3 in 2005 (Ministry of Education, 2004; Oladejo, 2005).

With regard to the English curricula at the tertiary level, prior to 1993 the focus was placed on reading and grammar translation; reading and translating works from literary anthologies was the teaching method mainly adopted (Chern, 2002). In 1993 the Taiwanese Ministry of Education mandated that the eight-credit Freshman English course be replaced by a six-credit foreign language course, allowing students to choose any foreign language offered by their universities to fulfil the requirement (Huang, 1997; Shih, 2000). However, some universities still kept English as the only option due to the lack of teachers for other foreign languages (Huang, 1997). Others, believing in the importance of English, only allowed students with a certain level of English proficiency to choose a foreign language other than English. At the time, reading was still the core in many programmes, with only a few

universities specifying an English programme that focused on skill integration (Huang, 1997). In 1997 the Ministry of Education lifted the 1993 mandate and allowed universities to decide on the foreign language requirements, including credit hours and course contents. Upon gaining autonomy over their curricula, most Taiwanese universities made general English instruction compulsory for first- and second-year students and offered them courses focusing either on different language skills or on specific topics, rather than the traditional unified programmes focusing on grammar and translation. Other than the general English courses, many universities now also offer optional courses to third or fourth year students to continue building their English proficiency (Huang, 1997; Shih, 2000).

It can be seen that there has been a progressive movement away from centralised and standardised English curricula. The decentralisation and deregulation in policy have gradually allowed more autonomy to universities and teachers at tertiary level.

2.2.3 English instructors in universities of technology

English as a Foreign Language (EFL) programmes for non-English majors in universities of technology are usually taught by local instructors. Most local EFL instructors in Taiwan have obtained their masters or doctorates from American or British universities, but they have not all majored in English-related areas. A common phenomenon is that returning graduates from overseas universities are hired to teach English, even though they have had no formal study of the English language, linguistics, or literature at their respective universities (Internet check conducted by myself in 2008). They are hired for the sole reason that they possess some skills in speaking English, the assumption being that if they can speak English, they can teach it.

Although EFL instructors at universities of technology generally have a high degree of professional commitment, there is still a need for institutions to build a system to check teaching quality, because this will influence the students' learning attitudes and abilities. The shortage of qualified English teachers has contributed to and has impacted negatively on the

current teaching situation in Taiwan. As a result, it seems reasonable to assume that the majority of EFL instructors in Taiwan are unprepared either for any upcoming changes to educational policy or for pedagogical changes.

2.3 Background of the study

2.3.1 National Development Plan

The widespread fear of losing ground in international economic competition has generally led to a focus on education as a means of improving human resources, enhancing productivity and increasing capital (Crump, 1999; Ratnavadivel, 1999). In common with many other countries and following the Asian financial crises of the late 1990s, Taiwan has been fearful of losing ground to other countries in the competition for global capital and has faced a number of technological challenges and a loss of investment and skilled managers to Mainland China. Mr. Yu Shyi-Kun, the Prime Minister in 2002, and his government predicted that Taiwan would need more highly qualified manpower, for example, in the areas of languages, information and communication technology, and research and development (Ministry of Education, 2005a). Accordingly, they formulated a comprehensive six-year national development plan – Challenge 2008 – where the importance of English to the economic and technological advancement of Taiwan was formally endorsed. The objective of the national development plan was to enhance Taiwan’s internationalisation and thus improve the nation’s competitive edge both economically and technologically (Executive Yuan, 2003). This could be achieved through various implementation strategies. One of these was the development of “an internationalised living environment and enhancement of people’s English proficiency”, which included the development of an English-friendly environment, a balance of urban and rural English education resources, internationalisation of higher education, enhancement of government employees’ English proficiency, and promotion of international cultural

exchanges (Council for Economic Planning and Development, 2002). In addition, the Taiwanese government also planned to make English a semi-official language by the year 2008, recognising it as a major tool for improving the country's status on the global stage (Executive Yuan, 2003).

The focal point of the national development plan, which is also particularly relevant to the present study, was the ability to use foreign languages, especially English, and internet communication. This could be achieved through the subprojects as follows. The first subproject involved creating an English-friendly environment, as part of the effort to come into line with global trends and attract more foreign visitors to Taiwan. The action plan started with the bilingualisation of general signs, websites, multimedia and government publications. Furthermore, in order to attract more international professionals to dedicate themselves to Taiwan's future development, the Ministry of Education used nine strategies to answer the needs of foreigners living in Taiwan. These were as follows: establishing an international service platform; setting up a service website for foreigners; establishing a bilingual environment; designing government and civilian websites bilingually; producing English broadcast programmes; promoting English translations of central government laws and regulations; developing professional human resources for applied English and translation; establishing a counselling service; promoting the counselling service to the people.

The second subproject involved establishing first-class universities. As Graddol (2006) noted,

One of the most significant educational trends world-wide is the teaching of a growing number of courses in universities through the medium of English. The need to teach some subjects in English, rather than the national language, is well understood: in the sciences [where], for example, up-to-date textbooks and research articles are obtainable much more easily in one of the world languages and most readily of all in English. (p. 45)

So far as Taiwan is concerned, promoting international trends in universities is important for developing the country. The Taiwanese government is currently setting plans in motion for

creating first-class universities and research centres, with the aim of boosting the nation's higher education to international standards. In addition, to enhance the international competitive advantage of higher education institutions nationwide, active academic and cultural exchange with foreign countries is considered essential. In order to attract students from abroad, the government has begun to establish world-class colleges, provide scholarships for international students to attend Taiwanese universities, establish education centres overseas, raise awareness of studying in Taiwan, encourage and support nationwide institutions to host education fairs or attend them overseas, encourage nationwide institutions to form academic cooperation agreements and exchanges with foreign universities, encourage dual degrees with foreign universities, and encourage the development of courses that are taught in English. Coleman (2006) observed that "the recruitment of international students and international staff, which English facilitates, leads to enhanced institutional prestige, greater success in attracting research and development funding, and enhanced employability for domestic graduates" so that "[institutional] and individual self-interest ... coincide both for academic staff, whose international careers depend on a demonstrated ability to teach and publish in English, and for students whose access to a good employment track on graduation also depends heavily on their proficiency in English" (p. 5-6).

The third subproject involved promoting the General English Proficiency Test (GEPT), which is sponsored by the Ministry of Education and has been developed, trialled and validated by the Language Training and Testing Centre (LTTC), and serves as a standard to verify the citizen's English ability. English as an international language has received more attention than ever from the government and the people in Taiwan because national boundaries are disappearing under the influence of globalisation (Yung & Hsu, 2005). Nowadays, people in Taiwan increasingly treat passing GEPT as a survival skill as more and more enterprises, both in the private and the governmental sectors, list 'GEPT score' in recruiting advertisements as a qualification for employment. According to the LTTC, the number of people taking GEPT in 2007 reached 2,700,000, which reflects this growing concern for English as a survival skill. With regard to tertiary education, for the purposes of

creating an international and educational vision and introducing a trend towards learning more English, it was considered crucial to encourage colleges and universities to strengthen English courses if student abilities were to improve. The GEPT was accordingly introduced into the educational system and the Ministry of Education set an increasing goal for the percentage of university students passing GEPT: 30% by 2005, 40% by 2006, and 50% by 2007.

At the same time, no figure was more important in this vision than the teacher on the front line. The teacher's ability was recognised as the determining factor in producing achieving or failing students. Therefore, the fourth subproject was to enhance the quality of English educators. Primary and secondary English teachers were encouraged to participate in in-service training in order to improve their teaching methodology. Teachers were also encouraged to take the GEPT themselves.

The last subproject centred on fostering study abroad. In recent years, both economic factors and expanded opportunities in Taiwan for domestic higher education degrees have evidently lowered expectations of studying abroad (Ministry of Education, 2005a). However, overseas study is still considered as an important factor which can enhance the quality of the nation's highly educated workforce. To this end, the government has improved websites offering related information on complete study plans and overseas safety in order to encourage more people to study abroad.

It can be seen from the national development plan that developing and promoting the nation's proficiency in English is seen as vital to the development of the country. This attempt is reflected in the major changes to the General English programmes at universities, whereby standardised English proficiency tests were introduced into the educational system and important decisions were made for students based on the scores of the tests.

2.3.2 The project on the Enhancement of Students' Foreign Language Proficiency

The need to move towards globalisation and internationalisation is considered to be of great urgency in Taiwan and one of the main areas involved is languages; foreign languages are felt to be the key to expanding knowledge and academic research. Since English is the most widely studied, read and spoken foreign language in the world, having become the dominant language of education, commerce, communication, science, technology, and entertainment (Kachru & Nelson, 1996; Richards & Rodgers, 2001), learning English well is recognised as a gateway to commanding much of the world's information and communication. In a public opinion survey published by King Car Education Foundation in January 2006, 80% of 2,059 respondents reported that they hoped that the Taiwanese government would designate English the second official language, which showed that the English language is regarded as critical to Taiwan's future (Graddol, 2006: 89). At the same time, the Taiwanese government has expressed "grave anxiety about its national proficiency in English" (Graddol, 2006: 95).

Apart from being an academic requirement, English language proficiency has recently, as noted in the previous section, gained an important role for new university graduates in securing employment in Taiwan. When applying for a job, new graduates are more often required to show some evidence of their English language ability based on scores from a standardised English proficiency test than a couple of years ago. Some companies even consider English ability an essential prerequisite for employment. In addition, English skills are tested for all those seeking promotion in governmental, educational, medical, financial, business and other government-supported institutions. Concerns, however, have been expressed, especially by the business community, that the general level of English language proficiency for students and the workforce in Taiwan has not been high enough and may even be on the decline (Graddol, 2006: 95). In an effort to address such concerns, the Ministry of Education launched 'The Grant Project on the Enhancement of Students' Foreign Language Proficiency' for universities of technology in 2002, along with implementing the national development plan and improving the quality of higher education. The six main features of the

project were (1) implementing foreign language examinations on and off campus, (2) providing remedial instruction, (3) encouraging new teaching methodologies for and experiments in teaching English and other foreign languages, (4) running a range of English camps and related activities, (5) cooperating with foreign universities and inviting foreign instructors to Taiwan, and (6) developing courses and other related supporting programmes delivered in English in order to attract international students (Ministry of Education, 2005b).

The testing and examination history of Taiwan can be traced back to the imperial period nearly two thousand years ago since the Han Dynasty (206 BC to AD 220). The imperial examinations were used to select the highest officials in the civil service of the country, based on merit and education. This tradition of using examinations for selection is still evidenced in the current education system in Taiwan. Given that it has been the test that has provided equal opportunities for the entire population to climb up the ladder of social status, the Taiwanese in general have tended to have an implicit faith in the potential usefulness of testing. The Taiwanese government places students according to the Basic Achievement Test (BAT) after junior high school (Grade 7-9), the Joint University Entrance Examination (JUEE) after senior high school (Grade 10-12), the master's degree programme entrance examinations after four-year university education programmes, and doctorate entrance examinations after a master's programme. Over the years, students take numerous examinations at school, municipal and national levels. It is therefore not very surprising that, when it comes to implementing educational changes, English proficiency tests have been introduced into the existing assessment system. This approach takes the view most commonly held in Taiwan that the best way to force students to improve their English skills is to test them. According to information gained from the governmental grant project website and preliminary interviews with two departmental directors and four teachers, conducted by myself in November and December 2006, the five main reasons given for introducing English proficiency tests were to raise the prestige of the English language, to equalise levels of teaching in universities of technology, to encourage teachers to increase the rate of teaching, to increase the emphasis on teaching oral language in the classroom in order to upgrade students' oral proficiency, and

lastly to increase the motivation of both teachers and students to teach and learn English.

In the face of the call from enterprise and the government to improve English proficiency, many university authorities have come up with the idea of a ‘Graduation Threshold’ (GT) since 2002. GT is a university policy whereby students have to fulfil graduation requirements by obtaining a certain level in, or getting at least a certain score on, an English proficiency test. It is important to note that taking English proficiency tests is not yet compulsory in all universities of technology; nevertheless a number of major employers, including the civil service, have stated that they require applicants to have attained a specific level in an English proficiency test, which leads university authorities to feel obligated to support the call for the introduction of the English graduation requirement. A variety of standardised English proficiency tests, developed by foreign or domestic testing institutes, are being utilised to measure overall English proficiency required in both academic and workplace contexts. The tests recognised by the universities which have implemented GT include the General English Proficiency Test (GEPT), the College Student English Proficiency Test (CSEPT), the Test of English as a Foreign Language (TOEFL), the Cambridge ESOL Exams, the Test of English for International Communication (TOEIC), the International English Language Testing System (IELTS), and in-house tests developed by the university itself (see Appendix 1). In order to appreciate the overall impact of testing in Taiwanese universities of technology, it is worth exploring briefly how the various English proficiency tests differ. An overview of the purpose, test content and methods, and scoring system of those proficiency tests follows.

The General English Proficiency Test (GEPT) is a criterion-referenced test and was developed in 1999 by the Language Training and Testing Center (LTTC) with a commission from the Ministry of Education in Taiwan. This test is open to all English learners with no restriction on profession or academic background. It aims to “promote the idea of life-long learning and to further encourage the study of English” (LTTC, 2008). Since its introduction in 2000, the GEPT has provided individuals with a gauge of their English language proficiency and has assisted employers and educational institutions in selection and placement. It provides assessment and certification of English language skills in reading, writing,

listening and speaking at five levels of competence: elementary, intermediate, high-intermediate, advanced, and superior.

The College Students English Proficiency Test (CSEPT) is a norm-referenced test developed by LTTC for college and university students. It allows tertiary institutions to place students in appropriate class levels, to evaluate students' learning during their degree and to serve as a graduation requirement. CSEPT measures students' ability in listening and reading as well as their English usage in daily life and academic contexts. It is offered at two levels: Level One is for students with elementary to intermediate levels of English proficiency, whereas Level Two is for students with intermediate to advanced levels of English proficiency. The total scores for Level One and Level Two are 240 and 360 respectively. There is no specific pass score.

The Test of English as a Foreign Language (TOEFL) was developed by the Educational Testing Service (ETS) in the USA and has been administered by that body since 1964 and has been taken by more than 750,000 people each year to pursue further study in English-speaking countries. It measures the ability of non-native English speakers to communicate in English in an academic setting. The TOEFL has gone through several major changes in format, including revision of sections and a shift from a paper-based test (PBT) to a computer-based version (CBT). In May 2006, the TOEFL CBT was changed into an internet-based test (iBT), which emphasises integrated skills, even though it is divided into four sections: reading, listening, speaking and writing. The total score for the iBT is 120.

The Test of English for International Communication (TOEIC) was developed by ETS in 1979 and was designed to measure non-native English speakers' overall communication skills required for international business. The TOEIC is a multiple-choice test that consists of 200 questions divided into two sections: a Listening Section and a Reading Section. The number of correct responses in each section is converted to a number on a scale of 5 to 495. Adding the two section scores together gives a total score on a scale ranging from 10 to 990.

The TOEIC has been widely accepted as the main English proficiency test for hiring employees at corporations in Taiwan and thus has forced many universities of technology to

accept it as a legitimate test for graduation. The market for the TOEIC test has expanded dramatically. The entire range of private language institutes throughout Taiwan offers preparation courses for the TOEIC, as do some universities of technology at their language institutes. Some universities even offer TOEIC preparation courses as regular credit-bearing courses.

There seems to be nothing wrong with selecting a standardised test like the TOEIC as a requirement for job acceptance, as competent employees are supposed to be equipped with the English communication skills required for international business. It is problematic, however, to employ the TOEIC for uses other than those intended by the developers. Utilising it as a multi-purpose test may compromise the validity of the test, if validity is defined as the extent to which test results are interpreted appropriately and meaningfully (Bachman, 1990). It is essential that the test be utilised to serve the purpose intended by the test developer, if it is to be considered a valid test.

The academic module of the International English Language Testing System (IELTS) is jointly managed by Cambridge ESOL, The British Council and IDP Education Australia, and has been used since the 1980s to measure the readiness of candidates to study through the medium of English at an undergraduate or postgraduate level. It is widely used as a selection tool by universities and other educational institutions for academic study. The IELTS is made up of four language skill-based tests: reading, listening, speaking and writing. The scores are reported on a Band Scale from 1 to 9. It is worth noting that the test batteries developed by Cambridge ESOL are far less familiar to Taiwanese test-takers than those developed by ETS, despite the fact that the tests are well known for their content validity.

The English version of the Business Language Testing Service (BULATS) was developed by Cambridge ESOL and assesses all four language skills that are needed in real business situations: reading, listening, speaking and writing. It provides four different types of test to meet different needs: the Standard Test, the Computer Test, the Speaking Test and the Writing Test. Each test can be used independently of the others, or they can be used in various combinations. BULATS results are given as a score on a scale of 1 to 100, and as levels on the

six-level system, which links to the Council of Europe's Common European Framework (CEFR) for modern languages. The results are also supported with 'can-do' statements, which indicate what candidates at each level are able to do in work contexts.

The Main Suite Exams were developed by Cambridge ESOL. They cover the four language skills – reading, listening, speaking and writing – and are designed for learners at all levels, ranging from the elementary level Key English Test to the very advanced Certificate of Proficiency in English, and in all contexts. They are widely recognised for work and study purposes.

Wu's (2004) study, however, indicated that the benchmarks are not consistent across universities depending on the university's resolution. For example, universities listed in Appendix 1 all adopt TOEFL CBT as one of the recognised tests. The benchmarks, however, are different in different universities. One university requires students to obtain 137 and above, another requires students to obtain 200 and above, and yet another requires students to obtain only 90 and above. At the same time, the Taiwanese Ministry of Education is considering implementing GT for all university students in Taiwan. It states in its mission plan that students' performance in English proficiency tests will become one of the items for the evaluation of that institution. The current proposal for the proficiency benchmarks is that they should be in line with the Common Reference Levels outlined in the CEFR – *Common European Framework of Reference for Languages* (Council of Europe, 2001). At a conference in Kaohsiung, it was suggested that the *Framework* offers a way in which tertiary institutions could consider creating a common core curriculum for their institution, something that "would ... not only encourage collaboration among staff, but also make their language teaching and learning objectives and outcomes more coherent, consistent and transparent" (Crombie, 2006: 1-11).

The development of proficiency scales and descriptors in the CEFR has taken communicative competence as the starting point. The CEFR treats proficiency as involving "a hierarchy of *global* characterisations of integrated performance" as in the American Council on the Teaching of Foreign Languages guidelines (SIL International, 1999)⁷. North (2000),

however, noted that any attempt to find essential links between communicative competence and proficiency is a very complex matter. Johnson (2004) found that proficiency benchmarking frameworks used in different countries vary widely: “Some ... have only one general (overarching) proficiency descriptor for each level; others have, at each level, both general proficiency descriptors and proficiency descriptors for different skills (e.g. reading, writing, listening and speaking)” (p. 2-3). In the case of CEFR, there are six bands in three categories (see Appendix 2). These are A1 (Breakthrough); A2 (Waystage); B1 (Threshold); B2 (Vantage); C1 (Effective-proficiency); and C2 (Mastery). The *Common Reference Levels* (CRL) were originally designed to apply to all of the languages spoken in the member countries of the Council of Europe and, by extension, to languages generally (Council of Europe, 2001: 21-42)⁸. The CRL have had impacts outside Europe including in Taiwan. The Taiwanese Ministry of Education has set a goal for the Graduation Threshold (GT) at B1 for normal universities and A2 for universities of technology, declaring the expected percentages of university students passing GT to be: 30% by 2005, 40% by 2006, and 50% by 2007 (Ministry of Education, 2005b).

The intention was, no doubt, gradually to raise the expected level. However, plans for implementing new measures were not drafted in sufficient detail at the outset and setting the Graduation Threshold based on the Common Reference Levels is problematic. There remains the critical issue of what test instrument to use. Simply transplanting systems formulated for use in other countries, such as Hong Kong, into Taiwan might well ignore important local factors and would seem to be thoughtless. In addition, the university in-house English examinations and university entrance examinations are unrelated to the same proficiency benchmark system, so university educators may be confused about what they ought to expect of entrants to their programmes and it will be difficult to determine how effectively tertiary institutions actually raise the proficiency levels of their students. As Chen and Johnson (2004) noted, “very little information is available about the current English language proficiency achievement of students following different programs in different institutions” (p. 136).

Elder and Wigglesworth (1996) stated that “the assessment of second language learners

raises complex issues about the nature of language proficiency, the validity of assessment instruments, the reliability of scores, and the manner in which the whole process may influence the curriculum” (p. 1). The use of tests as a device for creating impact on the educational process is often referred to as the ‘washback effect’. The introduction of GT, involving the use of standardised English proficiency tests, in higher education in Taiwan will inevitably have washback effects of some sort. This is probably desirable as well as inevitable. It is hoped that the introduction of English proficiency tests in universities of technology will upgrade the level of English learning and teaching. The rationale is that it will lead to the improvement of students’ English proficiency, as teachers will attempt to meet the new goals and criteria.

Although there is no shortage of ideas in response to worrying English standards, and there is plenty of courage on the part of the Taiwanese government and universities of technology in embracing bold solutions, a coherent policy for the development of English in Taiwan is harder to identify. The goal of being part of the recent trend of globalisation has encouraged an element of risk-taking in educational policy, but high-risk language education policy measures need to be implemented with great care. As Berry and McNeill (2005) suggested, “the *laissez-faire* approach, which is associated positively with economic growth does not always lead to success in educational terms” (p. 373).

2.4 Implications for the present study

Learning English is ‘a whole-nation movement’ in the 21st century in Taiwan. One of the tasks specified in the national development plan is the promotion of English language competences of the people in Taiwan (Ministry of Education, 2005a). One of the strategies to achieve this goal is to implement standardised English proficiency tests. At the tertiary level which is relevant to the present study, efforts have been made to set up a Graduation Threshold (GT) for university graduates. It is good to see government policies, as well as tertiary institutional

policies on language education, become decentralised and changes implemented in time to reflect the needs of society. The Graduation Threshold, however, will inevitably have washback effects on stakeholders. Given this situation, it is worth asking a number of fairly broad and basic questions in order to explore the Graduation Threshold's future development in universities of technology:

- RQ 1 What strategies do the Taiwanese Ministry of Education (MOE) and university departments use to implement The Enhancement of Students' Foreign Language Proficiency Project?
- RQ 2 What is the nature and scope of the washback effect on teachers' and students' perceptions of aspects of teaching in relation to English proficiency tests?
- RQ 3 What is the nature and scope of the washback effect on teachers' behaviours as a result of implementing the English proficiency tests?

In order to further understand these issues, it is important to explore in detail the concept of washback effects and develop a working definition for the term after reviewing the relevant literature. More detailed sub-questions will be allowed to emerge as the study proceeds.

Notes

¹ “Effective from 2001, the Joint Public Senior High School Entrance Examinations were eliminated, and a multi-route program to enter senior high school was implemented, allowing junior high graduates to enter high schools through assignment, application or selection by recommendation. However, junior high graduates must still pass the Basic Achievement Test for Junior High Students (BAT). After obtaining a BAT score, students can file applications, be selected by recommendation or get assigned based on their BAT score” (Clark, 2002: Recent Secondary School Reforms section). Senior high schools focus on general education subjects and are mainly intended to prepare students for admission to higher education institutions, which was by means of a highly competitive Universities and Colleges Joint Entrance Examination until 2002. Since 2002, entry to higher education has involved one of three processes: (1) an application process, (2) selection by recommendation, and (3) a Joint University Entrance Examination (Ministry of Education, 2008).

² Senior vocational high schools offer programmes combining general education subjects with a vocational component selected from one of seven main areas: industry or medicine, commerce, maritime studies, agriculture, nursing, home economics, drama and the arts (Ministry of Education, 2008).

³ “Students who have completed a two-year, junior-college-level program in certain technological disciplines may complete a bachelor’s degree in the same field at a college/university of technology. This requires an additional two years of study” (Clark, 2002: Programs and Degrees section).

⁴ “Five-year junior college programs, primarily technical and vocational in content, combine a student’s three remaining years of high school with two years of higher education. Successful students are awarded a *Certification/Diploma of Graduation*” (Clark, 2002: Secondary Education section).

⁵ The Taiwanese dialect is spoken by approximately 73% of the population, the Hakka dialect by approximately 12% of the population.

⁶ There are twenty weeks in each of the two school semesters.

⁷ Bachman (1990) and Bachman and Palmer (1996) treat proficiency as including language competence, strategic competence, and psychophysiological mechanisms, which includes organizational competence (grammatical and textual) and pragmatic competence (illocutionary and sociolinguistic competence).

⁸ Attempts have recently been made to create language-specific descriptors (Council of Europe, 2006).

Chapter Three: Literature Review

3.1 Introduction

This chapter focuses on the theoretical underpinnings that shaped and guided this study. It begins with an exploration of the concept of washback, by discussing various terms that have been used to describe this educational phenomenon. It will then illustrate the mechanism of washback and this is followed by a discussion of the washback phenomenon in different educational contexts. The following section explores how and why washback can work to influence other components within the educational system, traces the rationale behind the use of tests, and examines their power to change teaching and learning. Finally, there is a review of current models of teaching and learning research in the context of the theoretical and practical considerations of washback discussed in other sections in the chapter.

3.2 The definition of washback

It has long been widely recognised, and a phenomenon of considerable interest to language testers, that tests, particularly ones associated with important decisions such as university graduation, can have a major impact on educational systems and on the societies in which they operate. The term ‘washback’ (Alderson & Wall, 1993) in the field of applied linguistics and ‘backwash’ (Biggs, 1996) in general education has been used since the 1990s to refer to the impact which occurs in the form of teaching and learning directed towards a test, in terms of both intended positive effects or unintended effects and perhaps negative effects (Alderson & Wall, 1993; Bachman & Palmer, 1996; Cheng, Watanabe, & Curtis, 2004; Hughes, 2003). The definition made the assumption that “tests or examinations can or should drive teaching,

and hence learning” (Cheng & Curtis, 2004: 4); the result, if taken far enough, was often referred to as *measurement-driven instruction* (Popham, 1987). In order to achieve the goal of driving teaching, a match between the content and format of the test and of the curriculum was encouraged. This was referred to as *curriculum alignment* by Shepard (1993). This alignment, in which a new or revised test was introduced into the educational system with the aim of improving teaching and learning, was referred to as increasing *system validity* by Frederiksen and Collins (1989). As a result, the test became a part of a dynamic process in which changes in the educational system took place according to feedback obtained from the test, referred to as *consequential validity* by Messick (1996) and *test impact* by Bachman and Palmer (1996) and Baker (1991).

Although different terms were preferred by different researchers, they all referred to different facets of the same phenomenon. Washback was an educational phenomenon derived from research studies into (1) the relationship between teaching, learning and testing, and the relationship between different curriculum components, and into (2) curriculum change and innovation. The two areas were highly related to the present study.

In recent years, some researchers have differentiated between ‘washback’ and ‘impact’. Wall (1997) defined test impact as “any of the effects that a test may have on individuals, policies or practices, within the classroom, the school, the educational system or society as a whole” and defined washback as “the effects of tests on teaching and learning” (p. 291). Bachman and Palmer (1996) defined washback and test impact in a similar fashion and viewed washback as one dimension of test impact (see Green & Hawkey, 2004; McNamara, 2000; Taylor, 2005). They suggested that test impact could be seen as operating on two levels: the micro level, such as the effect of a test on individual students and teachers, and the macro level, such as the impact on society and its educational systems (see Weir & Milanovic, 2003). For Bachman and Palmer (1996), washback was a more complex phenomenon than simply the effect of a test on teaching and learning. They therefore suggested that the impact of a test should be evaluated with reference to contextual variables such as society’s goals and values, the educational system in which the test was used, as well as the potential outcomes of its use

(see Hamp-Lyons, 1997). Hughes (2003), however, explicitly equated washback with test impact when he defined washback as “a part of the impact a test may have on learners and teachers, on educational systems in general, and on society at large” (p. 53).

The term ‘washback’ was chosen in this study as it was the most commonly used in the field of applied linguistics. Concurring with Hughes’ (2003) definition, washback is interpreted broadly as encompassing test effects not only on teaching and learning, but also on the education system and society as a whole, in order to capture the full gamut of washback mechanisms. For the purposes of this study, the term ‘washback’ covers the impact of implementing English proficiency tests as a graduation requirement, including both accidental and intentional effects and the perceptions as regards about teaching and learning of a range of stakeholders. I shall use ‘washback’ as the default label, but the alternative ‘backwash’ will be retained where it is used in direct quotations. Where the term ‘impact’ occurs, it is used non-technically as a synonym for ‘effect’.

3.3 Washback: positive, negative, neither or both

Washback very often implies movement in a particular direction. Pearson (1988) pointed out that “public examinations influence the attitudes, behaviours, and motivation of teachers, learners and parents, and because examinations often come at the end of a course, this influence is seen working in a backward direction, hence the term ‘washback’” (p. 98). In addition, washback has been generally perceived as being bipolar – either negative (harmful) or positive (beneficial). Messick (1996) cited Alderson and Wall’s (1993: 117) definition of washback as the “extent to which a test influences language teachers and learners to do things they would not necessarily otherwise do that *promote* or *inhibit* [emphasis added] language learning” (p. 241). Alderson and Wall (1993) also emphasised the fact that evidence of washback was typically demonstrated in behavioural and attitudinal changes in teachers and learners that were associated with the introduction of tests bearing important educational

consequences. They noted that “tests can be powerful determiners, *both positively and negatively*, [emphasis added] of what happens in classrooms” (Alderson & Wall, 1993: 117; Wall & Alderson, 1993: 41). Buck (1988) gave a similar but slightly amplified description:

There is a natural tendency for both teachers and students to tailor their classroom activities to the demands of the tests, especially when the test is very important to the future of the students, and pass rates are used as a measure of teacher success. This influence of the test on the classroom (referred to as *washback* by language testers) is, of course, very important, this washback effect can either beneficial or harmful. (p. 17)

Bailey (1996) summarised the general research situation by noting that washback was defined as the influence of testing on teaching and learning, that it was widely held to exist and to be important, but that relatively little empirical research had been done to document its exact nature or the mechanisms by which it works. He also identified “concerns about what constitutes both positive and negative washback, as well as about how to promote the former and inhibit the latter” (Bailey, 1996: 259).

Some studies have shown that influences of tests on teaching and learning were almost always harmful – so-called ‘negative washback’ (see Alderson & Banerjee, 2001; Wall & Alderson, 1993). Alderson and Wall (1993) defined negative washback as the undesirable effect on teaching and learning of a particular test, such as “something that the teacher or learner does not wish to teach or learn” (p. 5). If a test failed to reflect the learning principles or the course objectives to which it was supposedly related, it would generate a harmful effect (Pearson, 1988). Teachers and learners ended up teaching and learning towards the test, regardless of whether or not they supported the test or fully understood its rationale or aims.

Tests have often been perceived as exerting a conservative force which impedes progress. The findings of Smith’s (1991a) extensive qualitative study of the role of external testing in US elementary schools further indicated that, due to the pressure to improve students’ test scores, “testing programs substantially reduce the time available for instruction, narrow curricular offerings and modes of instruction, and potentially reduce the capacities of teachers

to teach content and to use methods and materials that are incompatible with standardised testing formats” (p. 8). Noble and Smith (1994a: 3) pointed out that any form of high-stakes testing could affect teachers directly and negatively, and that “teaching test-taking skills and drilling on multiple-choice worksheets is likely to boost the scores but unlikely to promote general understanding” (Noble & Smith, 1994b: 6; see Alderson & Hamp-Lyons, 1996). Test-driven learning has proved to be a worldwide issue. In Japan, Korea and Taiwan, to name just a few countries, students approaching their final year of secondary school have been reported as focusing obsessively on passing the year-end university entrance examination, a major section of which was English (Kuba, 2002). Little attention, Kuba (2002) noted, was given to any topic or task that did not directly contribute to passing that one exam. In short, what was assessed was what got taught; the curriculum was driven by the test and thus narrowed.

Teachers’ tendencies to teach to a test have often been cited as an impediment to introducing new instructional practices (Chapman & Snyder, 2000: 460). When teachers believe they are being circumscribed and controlled by the tests and students’ focus is on what will be tested, teaching and learning have been found to be in danger of becoming limited and confined to those aspects of the subject and field of study that are testable (see Calder, 1997). The result of negative washback is “an atmosphere of high anxiety and fear of test results among teachers and students” (Shohamy, Donitsa-Schmidt, & Ferman, 1996: 309). As Shohamy et al. (1996) pointed out, teachers felt that the success or failure of their students reflected on them and they spoke of pressure to cover the materials for the exam. When students knew that one single measure of performance would determine their lives, they were less likely to take a positive attitude toward learning. The motives in such a context were almost exclusively extrinsic, with little likelihood of intrinsic interests being generated (Brown, 2004).

Some researchers, on the other hand, have seen washback in a more positive way (see Andrews, Fullilove, & Wong, 2002; Bailey, 1996; Davies, 1985; Wall, 2000). Heyneman (1987) claimed that many achievement testing proponents viewed “coachability” not as a

drawback, but rather as a virtue (p. 262), and Pearson (1988) argued that “good tests will be more or less directly usable as teaching-learning activities” (p. 107). Andrews et al. (2002) suggested deliberately introducing innovations in the language curriculum through modifications in language testing; for example, an oral proficiency test was introduced in the expectation that it would promote the teaching of speaking skills. The authors thus believed that it was both feasible and desirable to bring about beneficial changes in teaching by changing examinations, representing the ‘positive washback’ scenario. Davies (1985) maintained similarly that “creative and innovative testing ... can, quite successfully, attract to itself a syllabus change or a new syllabus which effectively makes it into an achievement test” (p. 8). To Davis, as to Andrews et al. (2002), the test no longer needed to be only an obedient servant; rather, it could be a leader.

Beyond this point in time, however, there appear to be conflicting reactions toward positive and negative washback on teaching and learning, and no obvious consensus in the research community as to whether certain washback effects were positive or negative. One reason for this was that the potentially positive or negative nature of a test can be influenced by many contextual factors. Pearson (1988) summarised the two positions in a rather deterministic way, by concluding that a test’s washback effect would be negative if it failed to reflect the learning principles and/or course objectives to which it supposedly related, but it would be positive if the effects were beneficial and “encourage(d) the whole range of desired changes” (p. 101). Furthermore, Alderson and Wall (1993: 117-118) emphasised that the quality of the washback effect might be independent of the quality of a test. Any test, good or bad, may result in beneficial or detrimental washback effects.

It was important that research into washback should turn its attention toward looking at the complex causes of the phenomenon in teaching and learning, rather than focusing on deciding whether or not the effects can be classified as positive or negative. Alderson and Wall (1993) pointed out that “it is surely conceivable that other forces exist within society, education, and schools that might prevent washback from appearing, or that might affect the nature of washback despite the ‘communicative’ quality of a test” (p. 116). Whether the

washback effect was positive or negative largely depended on how it worked and within which educational contexts it was situated. If the consequences of a particular test for teaching and learning were to be evaluated, the educational context in which the test took place needed to be investigated. Accordingly, it was important to investigate the nature of an examination with respect to teaching and learning, and then where and how washback existed and manifested itself within a particular educational context, keeping in mind the importance of the context in which it might take place in order to understand how washback functions in practice.

3.4 The mechanism of washback

Tests have often been used at the end of the teaching and learning process to provide a diagnosis of the effects of teaching and learning. However, testing may well be considered before the teaching and learning, in order to influence either or both processes. This view of testing is derived from the realisation of test power and its manifestations with regard to high-stakes decisions based on test results for individuals, educational systems and society as a whole. This section looks at the functions and mechanisms by which washback works in relation to other educational theories and practices.

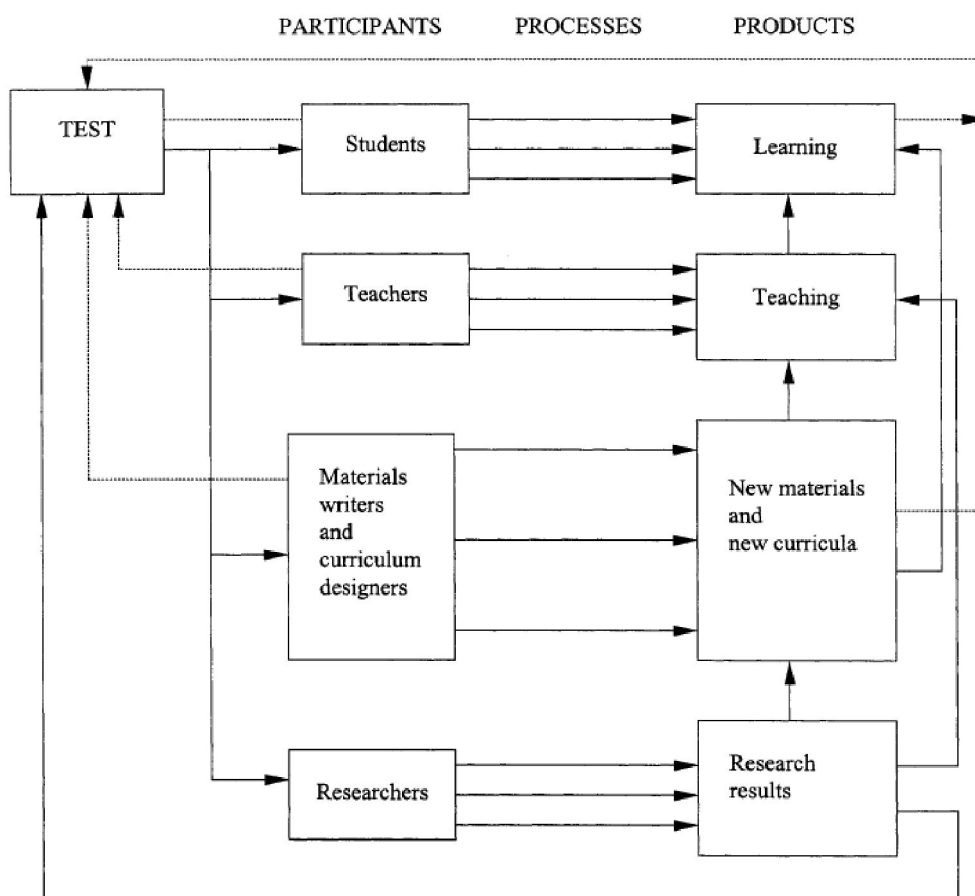
In discussing the complex mechanisms through which washback occurs in actual teaching and learning environments, Hughes (1993) introduced a concept of trichotomy and argued for distinguishing between participants, processes and products in both teaching and learning, recognising that all three may be affected by the nature of a test. In the Hughes' framework, 'participants' were people, such as students, teachers, administrators, materials developers and publishers, whose perceptions and attitudes towards their work might be affected by a test. The term 'processes' covered any actions taken by the participants which might contribute to the process of learning, such as the development of materials, syllabus design, and teaching methods. Finally, 'products' referred to what was learned (facts, skills,

etc.) and the quality of the learning (fluency, etc.). Hughes (1993) further noted:

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 the work. (p. 2)

Figure 3.1 A basic model of washback

(Source: Bailey, 1996, p. 264)



* → Both arrows and arrows with dotted lines mean 'influences'.

Based on Hughes' (1993) tripartite distinction between participants, processes and

products, Bailey (1996) developed and illustrated a model in which a test not only affected products through the participants and the processes they engaged in, but where the participants and processes also in turn provided feedback and thereby also had an impact on the test, as dotted lines in Figure 3.1 represent.

Alderson and Wall (1993), taking a different approach to understanding how washback worked in their Sri Lankan study, focused on aspects of teaching and learning that might be influenced by tests. They argued that there was little evidence provided by empirical research to support the assumption that tests impacted on teaching. They claimed that “the concept is not well defined, and we believe that it is important to be more precise about what washback might be before we can investigate its nature and whether it is a natural or inevitable consequence of testing” (Alderson & Wall, 1993: 117). As a result, they proposed the following 15 hypotheses in an attempt to indicate areas of teaching and learning that were commonly affected by washback and that might serve as a basis for further research:

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

- (15) Tests will have washback effects for some learners and some teachers, but not for others. (Alderson & Wall, 1993: 120-121)

The Washback Hypothesis has highlighted a correlation between the importance of tests and the extent of washback. Alderson and Wall (1993) concluded that further research on washback was needed and must entail “this increasing specification of the Washback Hypotheses” (p. 127). They suggested that in order to understand how washback works, it was important to take account of findings in the research literature in at least two areas: that of motivation and performance, and that of innovation and change in educational settings.

Wall (1996) followed up their study and stressed the difficulties in finding explanations of precisely how tests exerted influence on teaching. She turned to innovation theory to offer “insights into why attempts to introduce change in the classroom are often not as effective as their designers hoped they would be” (p. 334). Wall further suggested that in order to propose ways of exploring the complex topic of washback, research areas should include (a) the writing of detailed baseline studies to identify important characteristics in the target system and the environment, including an analysis of current testing practices (Shohamy et al., 1996), current teaching practices, resources (Bailey, 1996; Read & Hayes, 2003; Saville & Hawkey, 2004), and attitudes of key stakeholders (Bailey, 1996; Cheng, 2005; Hughes, 1993; Read & Hayes, 2003; Shih, 2007), and (b) the formation of management teams representing important interest groups, for example, teachers, teacher trainers, university specialists, ministry officials, parents and learners (Cheng, 2005; Shih, 2007).

As well as trying to illustrate the mechanism of washback as a phenomenon of change in teaching and learning, Smith (1991b) investigated what ‘success’ means in ELT project terms and attempted to construct his own model of the variables involved in attempts to introduce change. Smith’s model comprised five components of change: the target system, the management system, the innovation itself, the resources available, and the environment in which the change was supposed to take place. Similarly, Markee (1997) illustrated through his study of curricular innovation how change might be designed, implemented, and

maintained. His framework was based on the composite question posed by Cooper (1989): who (participants) adopts (process) what (the innovation), where (the context), when (the time duration), why (the rationale), and how (different approaches in managing innovation)?

There has been disagreement in the literature about whether washback should be seen as a deliberate and focused effect or not. Spolsky (1994) suggested that it “is better applied only to accidental side-effects of examinations, and not to those effects intended when the *first* purpose of the examination is control of the curriculum” (p. 55). However, Cheng (1997) used the term to mean “an active direction and function of intended curriculum change by means of the change of public examination” (p. 38). Having said that, Cheng (1997) did later acknowledge that “unintended and accidental side-effects could occur” (p.39). Messick (1989) implied that there was unintended as well as intended washback, writing that “judging validity in terms of whether a test does the job it is employed to do ... requires evaluation of the intended or unintended social consequences of test interpretation and use. The appropriateness of the intended testing purpose and the possible occurrence of unintended outcomes and side effects are the major issues” (p. 84). McNamara (1996) held a similar view, stating that “high priority needs to be given to the collection of evidence about the intended and unintended effects of assessments on the ways teachers and students spend their time and think about the goals of education” (p. 22). These various statements strongly suggest that researchers needed to investigate not only intended washback, but also unintended washback effects.

In summary, washback variables influencing various aspects of learning and teaching can be divided into “washback to the learner” and “washback to the programme” (Bailey, 1996, 1999a); the former refers to the impact of the test on test takers, while the latter is concerned with the impact of the test on teachers, administrators, and curriculum developers. The washback effect, however, is not solely confined to teaching and learning. Variables such as materials, curriculum and research are encompassed, making the mechanisms of washback more intricate and comprehensive. Unlike the Washback Hypothesis, which only proposes a linear relationship between tests and teaching or learning, Bailey’s (1996) model emphasises the importance of the interaction among the different components.

3.5 Empirical studies exploring washback on teaching and learning

The influence of washback has been observed on various aspects of learning and teaching and it has been found that the washback generated was mediated by numerous factors. Research projects looking at washback have been carried out in several different countries and contexts: new university oral exams in Sri Lanka (Wall, 1996; Wall & Alderson, 1993); Netherlands school language exams (Luijten, 1991); a needs-based exam at a Turkish university (Hughes, 1988); the national university entrance examination in China (Li, 1990; Qi, 2005); the Japanese university entrance exams (Watanabe, 2004); EFL and ASL exams in Israel (Shohamy, 1993; Shohamy et al., 1996); the various exams and exam revisions in Hong Kong (Andrews, 1994; Cheng, 2005; Fullilove, 1992; Lam, 1994); and the reformed high school entrance exam and a new test in Taiwan (Chen, 2002; Shih, 2007). It is worth examining several of these studies in a little more detail.

Wall and Alderson (1993) reported on the Sri Lankan project, which was intended to provide a ‘lever for change’ (Pearson, 1988) by means of new textbooks and a new English test. Wall (1996) summarised the results of their study by noting that “the examination had had considerable impact on the content of English lessons and on the way teachers designed their classroom tests (some of this was positive and some negative), but it had had little to no impact on the methodology they used in the classroom or on the way they marked their pupils’ test performance” (p. 348). This effect led to the narrowing of the curriculum to those areas most likely to be tested and as the exam got closer there was greater use of past papers and commercial exam-related publications (Alderson & Wall, 1993). Wall and Alderson (1993) found that the potential factors which impeded teachers from employing new teaching methods included insufficient teacher training, problematic management within schools, and teachers’ beliefs in the efficacy of various teaching methods. They therefore suggested that educational authorities had to consider all factors that led to the effectiveness of educational change and that affected teachers’ behaviour.

Shohamy (1993) and Shohamy et al. (1996), in the Israeli context, concluded that test

impact was multi-faceted and was a function of, *inter alia*, language status and test use. They conducted two research projects to examine the washback of an interactive English Foreign Language (EFL) oral matriculation test on teaching. Results indicated that there was a difference between experienced and inexperienced teachers' behaviour. The experienced teachers had devoted more time to teaching oral language and had adopted more mock test materials, whereas their counterparts, who had been trained in teaching of oral language on teacher-training programmes, had not. Moreover, in relation to the EFL test, "ample new material has been published and marketed since the announcement of the test changes became public" (Shohamy et al., 1996: 309), and the new materials were mostly clones of the new test format. Shohamy et al. (1996) concluded that the modified oral test had been successfully used to promote teaching oral language and it substantially changed what and how teachers taught. The results also showed that the experienced teachers, who did not obtain oral language teaching training, expressed their anxiety and fear toward the results of the test due to its high-stakes status, and they took the shortest possible route by teaching to the test to prepare their students for the test. Ferman (2004) later used questionnaires to assess students' views of the EFL oral matriculation test and found differential washback on learners, according to proficiency level. Low ability students engaged in more intensive preparation for the test, including private tutoring and memorising prompts. These students generally believed that they could boost their scores through cramming, and studied independently for areas of the test that were not covered in class. In addition, washback appeared to be most intense in students at the middle ability level, who reported the highest average levels of anxiety and felt that the test had led to an improvement in their language skills.

While investigating the impact of an Arabic Second Language (ASL) test, which was intended to raise the prestige of Arabic among Hebrew speakers, Shohamy (1993) found that teachers prepared students specifically for the ASL test prior to its test date. After the test, teachers returned to their previous routines. In a subsequent study, Shohamy et al. (1996) found that the ASL test did not generate a high degree of washback on teaching or on students' learning, and the effects decreased to become quite minor as time went by. The test

had influenced teaching, learning and the curriculum to the extent that teaching and testing essentially became synonymous. However, although the test had changed teaching and learning, it had not successfully raised the status of Arabic. The study suggested that the influence of the ASL test only lasted for a short period of time, because the stakes of the test were low and teachers thought that the test was not useful for facilitating higher-level learning. Shohamy et al. concluded that washback could change over time due to the importance of the test, language status, the purpose of the test, the format of the test, and the skills tested.

Alderson and Hamp-Lyons (1996) found, in their study of washback of the Test of English as a Foreign Language (TOEFL) on preparation courses, that the TOEFL was seen to have a more direct washback effect on teaching content than on teaching methodology. They also reported that many of the teachers they investigated made heavy use of exam materials. They suggested that the reason why teachers employed exam materials was that their negative attitudes towards the exam discouraged them from creating their own materials. In addition, the degree and types of washback differed between instructors. They suggested that the amount and type of washback which occurred depended on “the status of the test, the extent to which the test is counter to the current practice, the extent to which teachers and textbook writers think about appropriate methods for test preparation, and the extent to which teachers and textbook writers are willing to and able to innovate” (Alderson & Hamp-Lyons, 1996: 296).

Watanabe (1996, 2004) also found similar results while investigating the effect of the university entrance examination in English on the prevalent use of the grammar-translation method in Japan. He found that all the textbooks used by the teachers observed consisted of past exam papers and materials. In addition, the results showed that washback affected some teachers, but not others. Watanabe pointed out that teacher factors, including personal beliefs, educational background and past experience, seemed to be more important than the test in determining the teaching methodology a teacher employed. He suggested that individual teacher factors might outweigh the potential washback of the test and he thus emphasised the importance of teacher education. Watanabe also noted that school cultures might influence the

degree of washback; a positive school atmosphere which encouraged students to interact with authentic language might infiltrate into individual classrooms.

Cheng (1997, 1998, 1999, 2004, 2005), in her large-scale quantitative and qualitative empirical study, investigated whether the modified Hong Kong Certificate of Education Examination (HKCEE) taken by most secondary school graduates brought about the positive washback on teaching that was intended. She developed the notion of ‘washback intensity’ to refer to the degree of the washback effect in an area or a number of areas of teaching and learning affected by the HKCEE. The results of the study revealed that changes to the ‘what’ of teaching and learning, such as teaching content and materials used, occurred quickly. However, the intended changes to the ‘how’ of teaching and learning appeared to have been at a superficial level. Cheng (2005) concluded that the washback effects were negative because the high-stakes exams drove teaching and learning in the direction of drilling what was required for the exam, thereby narrowing the curriculum, and limiting learning activities to drilling or rote memorisation, which in turn reduced possibilities for authentic language use. With respect to washback from HKCEE on students’ learning, results indicated that the HKCEE was the most significant factor motivating students to learn English, surpassing other factors such as future career concerns, parent concerns and classmate competition (Cheng, 1997). Cheng, in her later 1998 study, found that although students shifted their attention after the test content was changed, they reported retaining their original learning processes, learning strategies, and individual motivations to learn English. In terms of feelings and attitudes, Cheng (1998: 296) found that students showed mixed feelings towards the test itself, recognising on the one hand that the exam made them work to achieve good scores, but at the same time thinking that exams were not an accurate reflection of all aspects of their study. To summarise, both teachers’ and students’ perceptions and beliefs with respect to teaching and learning remained the same. Cheng (1998) suggested that “to change the how, ... genuine changes in how teachers teach and how textbooks are designed must be involved. A change in the examination syllabus itself will not alone fulfil the intended goal” (p. 297).

Li (1990) studied the National Matriculation English Test (NMET) for entrance to all

universities in China. Apart from its major function of selection, the NMET aimed to produce a positive washback effect on school teaching and learning. Li suggested that because a large number of people took this test annually (three million), it could be seen as a high-stakes test. According to Li, washback from the NMET could be seen in teaching materials and lesson content, but the most changes were observed outside the classroom. These changes reflected “a new awareness of time and resources and a new enthusiasm for after-class learning of English” (Li, 1990: 397). Nevertheless, in the classroom, Li discovered a shift away from formal linguistic knowledge to practising and using the language, suggesting the intended washback had been attained. Interestingly, while teachers reported being uncomfortable with the NMET, students were much more adaptable and accepted the new testing methods quite readily, due to the significant influence of the test on their future opportunities. In contrast, Qi (2004, 2005) examined whether the NMET had attained the intended washback and found that the NMET was not an efficient lever to bring about pedagogical changes, as teaching of linguistic knowledge rather than communicative competence was still emphasised and the type of language taught was limited to the skills tested. Qi imputed this failure to the stakes of the test. Because the NMET was a high-stakes test, teachers tended to teach to the test and saw assisting students in achieving high scores as the immediate goal.

Hayes and Read (2004) demonstrated how changes to the English language testing system in New Zealand showed washback effects. They compared two International English Language Testing System (IELTS) preparation courses; one claimed to be more test-oriented, and the other leaned more towards English for academic purposes. Their results indicated that the IELTS had an impact on both courses, and teachers and students of the more test-oriented course focused narrowly on test tasks rather than on academic language proficiency in the broader sense, suggesting that teaching and learning objectives influenced teaching approaches.

Chen (2002) examined the nature and scope of the impact of a reformed high school entrance exam associated with educational reform. She found that the washback effect on teachers’ teaching attitudes was superficial. There were washback effects on teaching contents

because of the new teaching materials for junior high schools throughout Taiwan, but no washback effect was observed on teaching methodology, because of a lack of in-service teacher training, such that teachers did not know how to change their teaching methods to align with the new curriculum. Chen also reported that there was a mismatch of goals between the new curriculum and the reformed test. The main aim of the new curriculum was to promote a communicative syllabus for classroom teaching and learning. However, due to the new test's multiple choice format and to the fact that it excluded oral and aural tests, how students' communicative competence could be assessed and whether washback from the test made classroom teaching more interactive were unclear. Chen highlighted the importance of providing teachers with professional development opportunities and including oral and aural tests in the national public examination.

Shih (2007) investigated the washback of the General English Proficiency Test (GEPT) on English learning in Taiwan. He reported that the GEPT had induced various but limited degrees of washback on learning. It had no impact on some students, but motivated other students to study English for a short period of time. On the question of students' test preparation after school, commercially-written practice books were the most commonly employed materials. With regard to learning strategies, students took preparatory courses at cram schools. The impact of the GEPT on teachers' teaching was observed in test preparation courses, in which teachers coached students' test-taking strategies and coaxed students to prepare for the GEPT, but in general the impact on other courses was small. Shih concluded that the GEPT generated little washback, because of the status of the test, students' learning attitudes, and departmental policies.

Most research findings to date seem to have shown that the tests concerned had an impact on the teaching content (see Alderson & Hamp-Lyons, 1996; Chen, 2002; Cheng, 1997, 1999, 2004, 2005; Hayes & Read, 2004; Qi, 2004, 2005; Shohamy, 1993; Wall & Alderson, 1993). Some findings have shown that tests altered teaching methods and materials, but others have indicated that the tests had limited or no impact on either (see Alderson & Hamp-Lyons, 1996; Chen, 2002; Cheng, 1997, 1998, 1999, 2004, 2005; Wall & Alderson,

1993). Washback may also be differential; it occurs with some teachers, but not others (see Alderson & Hamp-Lyons, 1996; Watanabe, 2004). Tests have also been found to have an impact on how teachers administered them (Wall & Alderson, 1993). In addition, tests have imposed anxiety and fear on teachers due to their high-stakes status (Shohamy et al., 1996). The available evidence suggests that tests have a superficial impact on students' learning, and that individual learners, like teachers, have experienced this influence in different ways, with the potential for considerable impact in terms of affective factors and teacher behaviours (Cheng, 2005; Ferman, 2004; Shih, 2007; Shohamy et al., 1996). Such differences between the degree of washback on teachers and students have raised questions about the extent to which washback to teachers can be assumed to be generalisable to washback to learners (see Shohamy et al., 1996). To date, there has been a lack of evidence indicating how washback to teachers and programmes might interact with washback to learners.

In short, the factors that induce or constrain washback have been found to include the following (Alderson & Hamp-Lyons, 1996; Cheng, 2005; Shohamy et al., 1996; Wall & Alderson, 1993; Watanabe, 2004): *test factors*, such as test methods and contents, skills tested, and the purpose of the test; *prestige factors*, such as the stakes of the test, and the status of the test within the entire educational system; *personal factors*, such as teachers' attitudes towards the test, teachers' educational backgrounds and their beliefs about the best methods of teaching/learning, and teachers' personalities and their willingness to innovate; *micro-context factors*, such as 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.

3.6 Tests as a means of control

Tests have long been advocated by policymakers in diverse national contexts as one form of effective control. Tests often occur at gateways, controlling access to opportunities *within* societies, and/or controlling the flow of people *between* societies. In traditional China, tests

were used for the selection of individuals who would go on to be trained to become the governmental ruling elite. Although the goal of the examination was to select civil servants, its washback effect was to establish and control an educational programme, as career-minded Chinese set out to prepare themselves for the examination (Spolsky, 1994, 1995). Tests thus played a crucial role in constructing the fundamental character of Chinese cultural and political life over many centuries (Fairbank & Goldman, 1998).

In the modern world, tests are frequently used for accountability within the system, and in particular for certification of achievement in education. They form part of the procedure for decisions about the allocation of scarce resources at both a systemic and an individual level. For example, in many countries, tests control the transition between school and higher education, and they may lead to the awarding of a degree. Tests are also seen as ways to upgrade knowledge and to improve the performance of institutions (Eckstein & Noah, 1992; Shohamy, 2000). Through testing, education policy can be rapidly diffused and implemented at relatively low cost (Linn, 2000). Test results are visible and ideally measurable, can be reported by the media in terms the public can understand and can be used to show that change has or has not taken place. Tests are often subject to much criticism, but in spite of all the criticism leveled at them, tests continue to occupy a leading place in the educational arrangements of many countries.

The driving force behind these developments has essentially been economic. Due to a concern for national standards of educational achievement in a competitive global economy, together with a heightened demand for accountability of government expenditure, a number of countries have used tests to gear up their education systems (McNamara, 2004). As pointed out by Shohamy (1993), the use of tests for power and control has often led to top-down educational change strategies, though this can be characterised as unethical, undemocratic and unbeneficial to test takers (Shohamy, 1998, 2001). This is an especially common practice in countries with centralised educational systems, where the curriculum is controlled by central agencies (see Heyneman, 1987; Heyneman & Ransom, 1990; Li, 1990; Shohamy et al., 1996). Policymakers in centralised systems, aware of the power of tests, have introduced tests

as ways of controlling and imposing specific knowledge on students, teachers, principals and educational systems, alongside particular agendas and educational ideologies (Broadfoot, 1996; Shohamy, 2001). They have used tests to manipulate educational or political systems, to control curricula, and to impose new textbooks and new teaching methods. In such settings, tests have been viewed as a primary and a cost-effective tool through which changes in the educational system can be introduced without having to change other educational components such as teacher training or curricula. Furthermore, Shohamy et al. (1996) suggested that “the power and authority of tests enable policymakers to use them as effective tools for controlling educational systems and prescribing the behaviour of those who are affected by their results – administrators, teachers and students” (p. 299). For example, administrators have used institution-wide exams to direct teachers’ teaching, hence to enforce learning, while in classrooms, teachers have used tests and quizzes to impose discipline and to motivate students to learn (Linn, 2000; Stiggins & Faires-Conklin, 1992).

It is precisely the power of high-stakes tests that make them potentially so influential with respect to educational changes. High-stakes tests are those whose results are seen by students, teachers, administrators, parents, or the general public as the basis upon which important decisions are made that immediately and directly affect the test-takers and other stakeholders (Langerfeld, Thurlow, & Scott, 1997; Madaus, 1985). They create a context in which educators and others are likely to pay attention to scores and to act in ways intended to maximise test performance (Corbet & Wilson, 1991). High-stakes tests have frequently reflected the belief that “the promise of rewards or the threat of sanctions is needed to ensure change” (National Research Council, 1999: 35). The primary use of high-stakes tests has been “to ration future opportunity as the basis for determining admission to the next layer of education or to employment opportunities” (Chapman & Snyder, 2000: 458), which illustrates well the power and control aspects of testing, as well as its role in cultural and social reproduction. As such, high-stakes tests are usually public examinations or large-scale standardised tests.

Recognition of the power of high-stakes tests has led educators to use them as a force for

exerting a positive influence in support of curricular innovation, which was described by Markee (1997) as “a managed process of development whose principal products are teaching (and/or testing) materials, methodological skills, and pedagogical values that are perceived as new by potential adopters” (p. 46). The assumption has been that because important consequences are involved, high-stakes tests can shape teaching and learning in the way intended by policymakers and test designers (see Popham, 1987; Pearson, 1988). High-stakes tests are thus believed to have the potential to achieve certain positive effects on teaching and learning. Teachers may, for example, respond positively to high-stakes testing by working harder (Bishop & Mane, 1999). Elton and Laurillard (1979) suggested succinctly: “The quickest way to change student learning is to change the assessment system” (cited in Tang & Biggs, 1996: 159). This suggestion is similar to the thinking involved in measurement-directed instruction, which “occurs when a high-stakes test of educational achievement ... influences the instructional program that prepares students for the test” (Popham, 1993, cited in Chapman & Snyder, 2000: 46). Popham (1987) claimed that if tests were properly conceived, focused on appropriately selected content and skills, and were sensibly implemented with sufficient support for teachers, then aligning teaching with what such tests assess was likely to have positive educational outcomes.

On the other hand, numerous studies have shown that high-stakes testing has influenced teaching and learning in a negative manner, by narrowing teachers’ focus excessively. In his study of effects of external high-stakes tests on teachers in USA primary schools, Smith (1991a) observed an unfortunate alignment of testing and teaching. He found that pressure to improve students’ test scores caused some teachers to neglect untested materials. On the basis of this and other evidence, he came to the general conclusion that teachers often had negative feelings regarding the publication of test scores and would do whatever was necessary to avoid low scores, such as teaching to the test. Moreover, when a teacher’s professional worth was estimated in terms of examination success, teachers would ‘corrupt’ the skills measured by reducing them to the level of strategies used to drill students for the examination. Qi (2005) too, in his study of effects of the National Matriculation English Test in China, discovered that

teachers altered their instructional materials to resemble the format of the imposed test. The situation also caused a reduction in teachers' ability to adapt, create, or diverge, so that the methods which they had in their arsenal were reduced and the teaching work became a deskilled activity. As Qi (2005) explained, the notion of using high-stakes testing to positively change teaching was innately problematic, due to the pressure it created to teach to the test.

Similarly, Herman and Golan (1991) found in the USA that teachers adjusted their instructional planning according to what was included in a test. Teachers tended to look at prior tests to make sure that their instruction included all or most of the test content, and planned to ensure that they covered all the test objectives. They also based their instructional plans and the sequence of the curriculum on the test performance of the class they had had the previous year. Testing also affected the teaching and learning process through instructional items devoted to direct test preparation activities. Herman and Golan (1991) concluded that testing substantially influenced teachers' classroom planning. The greater the consequences attached to the test, the more likely it would be to have an impact on teaching. Madaus (1985), taking a different view, pointed out that if the examination was perceived as important enough, a commercial industry would develop to prepare students for it, hence the rise of most testing agencies.

Other studies pointed to standardised tests' narrowness of content, their lack of congruency with curricula and instruction, their neglect of higher order thinking skills, and the limited relevance and meaningfulness of their multiple choice formats (Hayes & Read, 2004; Nobel & Smith, 1994b; Qi, 2005; Shepard, 1990). Rather than exerting a positive influence on student learning, testing may trivialise the learning and instructional process, distort curricula, and usurp valuable instructional time (Cheng, 2005; Ferman, 2004; Smith, 1991b). There were also a number of studies that raised questions about whether improvements in test score performance actually signalled learning (Cannell, 1987; Linn, Grave, & Sanders, 1989; Shepard, 1990). The research studies showed that very often the use of tests did not lead to improved learning or to higher achievement. The use of tests simply created parallel systems in which tests became the *de facto* knowledge; the result was an artificial way of granting

power.

Still other scholars, however, found that high-stakes testing had little or no influence on teaching and learning at all. Dorr-Bremme and Herman (1983) found that the teachers in their study, conducted in the USA, paid little attention to school-based standardised tests, as they believed such tests were of little relevance to their students, and the teachers viewed the results of such tests as relatively unreliable sources of information. While referring to testing impacts on teacher planning and instruction, Dorr-Bremme and Herman also found relatively little influence of standardised tests on teacher decision-making activities, such as placing students, planning instruction, or grading. After analysing two major curriculum changes in the previous 25 years in the context of his study, Hogan (1986) also found that standardised tests exercised little influence upon major curriculum change. It seemed that what was important was “not whether a test is high-stakes, but whether participants believe it to be high-stakes” (Gipps, 1994: 34).

In summary, it can be seen that studies into the influence of high-stakes testing have led to a variety of results:

- influences at different levels on different stakeholders (*institutions*: accountability pressure; *teachers*: attitudes and responses to standardised tests, and attitudes about the fairness and utility of testing; *learners*: reactions, self-concepts, and self-assessment; and *parents*: their knowledge about and attitudes towards standardised testing, and familiarity with the changes in the evaluation of their children’s school progress), and
- influences on different aspects of teaching and learning (teaching content, instructional planning, approaches to teaching, time spent on test preparation, the teacher’s professional worth, and student learning outcomes).

From those studies, it seemed that the possibilities for positive washback effects from high-stakes testing were limited by the nature of the exams themselves. It is worth pointing out that testing was likely to be just one influence among many, and the effects of testing would in most cases interact with other factors such as teachers’ beliefs and knowledge about

pedagogy, their professional development experiences, their access to necessary resources, and the responses of their colleagues and supervisors (Cimbricz, 2002; Haertel, 1999). Numerous studies have shown that high-stakes tests rarely have the consequences intended by policymakers and teacher factors seem to be crucial, as Swain (1985) pointed out: “it has frequently been noted that teachers will teach to the test; that is, if they know the content of a test and/or the format of a test, they will teach their students accordingly” (p. 43). When test scores are associated with consequences which are important or meaningful to teachers, it is likely that instruction will be affected accordingly, and vice versa. Ultimately, teachers are the final arbiters of policy implementation.

Nevertheless, testing has become “the darling of policymakers” (Madaus, 1985: 5; Shohamy, 2007). Petrie (1987) emphasised that “it would not be too much of an exaggeration to say that evaluation and testing have become *the* engine for implementing educational policy” (p. 175). Given the important decisions attached to tests, it is unsurprising that they have so regularly been used as instruments and targets of control in educational systems (Eckstein & Noah, 1993a, 1993b). Their relationship with the curriculum, with what a teacher teaches and what a student learns, and to an individual’s life chances are of vital importance in most societies. Resnick and Resnick (1992), however, argued that “there is rarely sustained or widespread consideration of the possibility that the very idea of using test technology as it has developed over the past century may be inimical to the real goals of educational reform” (p. 1). Using tests to solve educational problems is a simplistic approach to a complex problem. Policymakers need to be aware that washback effects may undergo changes over time and may not last indefinitely within the system. Therefore, studies of particular changes are needed to examine the role and function of tests in teaching and learning in detail and, if possible, longitudinally.

It is therefore of interest to investigate whether or not the situation found in the studies cited above has arisen in the context of general English teaching in Taiwanese universities of technology. Examinations have a long history of dominating and influencing actual teaching and learning in Taiwan (see Chapter 2). Taiwanese school classrooms have been portrayed as

teacher-centred, textbook-centred and test-centred (Shannon, 2006). Very often teachers and textbook writers have followed the direction of the test, although the degree of attention to the test has varied according to the teacher, textbook writer, and publisher (Shannon, 2006). As has been true of previous attempts at educational change, tests were central to the Taiwan educational change debate for at least three reasons. First, test results were relied upon by governmental and educational organisations to document the need for change. Second, tests were seen as critical agents for change (Clark, 2002; Oladejo, 2005). Third, test results were used to demonstrate whether or not change had occurred.

The English proficiency tests used in Taiwanese universities of technology are good examples of high-stakes tests. They are taken by the majority of university students prior to their graduation. Students either proceed to further studies (e.g. postgraduate study) or leave university and seek employment. In both situations, passing an English proficiency test is vital for them as it is necessary if they are to obtain a first degree and a specific job. Some researchers have argued that tests can induce only superficial changes (Andrews, 1994, 2004; Cheng, 2005). The reasons for high-stakes tests' failing to produce fundamental changes in teaching and learning are not well documented in the literature. Accordingly, the current study draws on previous research findings to explore the areas in teaching and learning that are likely to have been influenced by English proficiency tests, with a view to informing the testing profession and policymakers about the reasons why high-stakes tests are often blunt instruments for causing significant change in teaching or learning.

3.7 Educational innovation

Wall (1996: 338), referring back to Alderson and Wall (1993), suggested that in order to understand how washback worked, it was important to take account of findings in the research literature in the area of innovation and change in educational settings. There has been a well-established tradition of disseminating innovation in educational literature (Fullan, 2007;

Rogers, 2003) and an increasing body of literature focusing on the English language teaching context (Henrichsen, 1989; Kennedy, 1990; Li, 2001; Markee, 1993, 1997, 2001; Stoller, 1994; White, 1988, 1991, 1993). These studies have helped to clarify the complexity of the innovation process, and the various factors which inhibit or facilitate successful diffusion and implementation.

Following Nicholls (1983), innovation can usefully be defined as a planned and deliberate effort, perceived as new by an individual and individuals, which is intended to bring about improvement in relation to desired objectives. Educational innovation originates from the identification of aspects of an educational system that are viewed as problematic, such as falling standards of student achievement, poor performance by students in international comparisons, or lack of transparent accountability reporting. It also involves systematic attempts by governments or other educational authorities to change educational policies and practices with a view to improving learner outcomes (Brindley, 2008: 365). What is crucial in innovation is that it signifies change involving human intervention (White, 1993: 244).

People have innovated for all sorts of reasons, sometimes with encouragement, recognition or reward. Wagner (1988) argued that innovations were more often introduced for practical reasons, rather than because the innovation had any intrinsic pedagogical value (cited in White, 1993: 245). A successful innovation, whatever the motivation, was difficult to achieve. The key difficulty identified by White (1993: 246) was that gaps occurred in the innovation process: between advocacy and adoption, and between diffusion and implementation. For example, fundamental tensions can often exist between policymakers and local practitioners about the purposes of an innovation. Whereas government officials may see assessment as a tool for implementing and managing policy, teachers may be primarily concerned with ways in which it can be used for the improvement of learning (Shohamy, 2001). In addition, the same innovation may be desirable for one adopter in one situation, but undesirable for another potential adopter whose situation differs.

Change is often a lengthy and painful process (Fullan, 2007; Pinar, 1999). In the

innovation literature, a number of characteristics have been suggested as correlating with the rate and success of implementing and spreading an innovation. One of the most cited set of perceived attributes is perhaps that proposed by Rogers (2003). Rogers' framework consisted of five attributes which were relative advantage, compatibility, observability, trialability, and complexity. Rogers suggested that the first question likely to be raised by those whom the innovation most affected would relate to its relevance. Relative advantage represented the extent to which the innovation in question was perceived as being better than the idea it replaced. The greater an individual perceived the relative advantage of an innovation to be, the more rapid its rate of adoption would be. Rogers made one specific point concerning relative advantage that may be of particular importance to practitioners of educational change, and that was particularly relevant to the present study. The point related to the use of incentives. Incentives were direct or indirect payments of cash that were given to an individual or a system in order to encourage behavioural change. Rogers' point was that if incentives contributed strongly to a decision to adopt a change, there might be little relative advantage to its continued use after the incentive was reduced or removed.

The second attribute, compatibility, described the congruence of the innovation with the existing values, past experiences and perceived needs of potential adopters. An innovation, Rogers (2003) claimed, had a far better chance of being accepted and adopted rapidly if it could be seen to be compatible with the existing values and norms of a social system. Complexity was Rogers' third attribute which was the extent to which an innovation was perceived as difficult to understand and use. Innovations that were seen as difficult to understand and adopt would diffuse more slowly, as few people would voluntarily embrace change that made their lives more difficult.

The fourth attribute, trialability, referred to the extent to which a prospective adopter could try out an innovation with a limited basis before committing to full adoption. An innovation that was trialable represented less uncertainty to the individual who was considering it for adoption, as it was possible to adopt the innovation a little at a time rather than all at once and to learn by doing. Observability, the final innovation attribute, pertained to

the intended adopter's ability to actually see the innovation being used by others. Later adopters might find it easier to learn innovation use by watching it. Innovations that were highly observable might diffuse more quickly than those whose positive consequences were invisible or delayed.

Rogers (2003) claimed that innovations that were perceived by individuals as having greater relative advantage, compatibility, trialability and observability, and less complexity would be adopted more rapidly than other innovations. Of these, relative advantage and compatibility were particularly important in explaining an innovation's rate of adoption.

Markee (1993) identified a slightly different set of five perceived attributes which affected acceptance of innovations. The first four were usefulness, similarity, difficulty, and concreteness, which were parallel to Rogers' relative advantage, compatibility, complexity, and observability. Markee's fifth attribute, likeability, referred to the degree to which the intended user of the innovation was satisfied with the status quo.

Similarly, Stoller (1994), in a sophisticated study within language teaching institutions, analysed perceived attributes of innovations in an attempt to discover which attributes would facilitate and which would inhibit the implementation of innovations. He grouped them in a three-factor model: dissatisfaction, viability and balanced divergence. The dissatisfaction factor, combining dissatisfaction with, and improvement over, past practices, and the viability factor, which included practicality, feasibility and usefulness, were both perceived to play strong facilitative roles in the implementation of an innovation. The third factor, the balanced divergence factor, comprised explicitness, complexity, compatibility, visibility, flexibility and originality. They could be either facilitative or inhibitory, depending on the degree of the attribute presented in a given innovation at a given time. Stoller (1994) concluded that:

When these attributes are 'sufficiently' present and fall within a perceived *zone of innovation*, adoption rates are likely to increase. Conversely, their perceived absence or excess can be detrimental to adoption rates because they lead to unfavourable attitudes towards the innovation and subsequently undermine potential support for the innovation. (p. 314)

A successful innovation would be one that fitted into the narrow 'zone of innovation': the space between 'not too original' and 'not original enough' (Wall, 1996).

Stoller (1994) suggested that there was a three-stage process leading to the formation of favourable or unfavourable attitudes towards an innovation. Ideas for innovation would first emerge from the recognition of dissatisfaction with current practice, followed by a viability factor, whereby innovative ideas were supported by individuals both within and outside the organisation if the innovation was perceived as useful, feasible and practical. Finally, a balanced divergence factor was invoked, whereby the innovation would earn the support of outsiders if it was perceived to fall within the 'zone of innovation'; in other words, if it was perceived as being neither excessively divergent from, nor too similar to, current practices.

Fullan (2007) shifted the setting of educational change to a broader stage, introducing three phases in the innovation process: initiation, implementation and continuation. The initiation stage was the period between the first appearance of an idea for change and the time when it was adopted. After an innovation was adopted it passed through the phases of implementation and continuation. Fullan (2007) did not view change as being linear: "It is not a linear process but rather one in which events at one phase feed back to alter decisions made at previous stages, which then proceed to work their way through in a continuous interactive way" (p. 67). He proceeded to list three sets of interactive factors, which can interact with each other to produce different patterns, related to the characteristics affecting the success of implementing and spreading an innovation:

1. Characteristics of the innovation project
 - Need
 - Clarity
 - Complexity
 - Quality/Practicality
2. Local Characteristics
 - District
 - Community
 - Principal
 - Teacher
3. External Factors
 - Government and other agencies

The first four factors were parallel to Rogers' framework, and it is clear that the 'fit' between an innovation and an institution's needs is essential. Bailey (1992) found that the main reasons for change were teachers' experiences of dissatisfaction with the way things were, including human affective factors and frustration with teaching methods and/or materials. The changes which generally happened and lasted were local and locally adapted, and they were developed and implemented by individual teachers or by small local groups of teachers (Farrell, 2000:87). Stoller (1994) also found that the 'dissatisfaction' factor played the most facilitative role when it came to introducing innovations: potential adopters were more likely to accept an innovation if they were already dissatisfied with the status quo, or if they could see that the proposed change offered some improvement over past practice. In addition, potential adopters would be even more willing to accept an innovation if what was being offered seemed practical, feasible and useful – in short, if it was viable.

Clarity about goals and means is a perennial problem in the change process. At the decision stage, "innovations find greater acceptance if goals are clearly defined" (Hadley, 1999: 7). Fullan (2007) noted that:

Even when there is agreement that some kind of change is needed, as when teachers want to improve some area of the curriculum or improve the school as a whole, the adopted change may not be at all clear about what teachers should do differently. (p. 89)

Lack of clarity – unspecific goals and means of implementation – represents a major problem at the implementation stage; teachers and others have often found that the change concerned was simply not very clear as to what it meant in practice. The result could be what Fullan (2007) called 'false clarity', which occurs "when change is interpreted in an oversimplified way; that is, the proposed change has more to it than people perceive or realise" (p. 89). For example, teachers may use a new textbook in the classroom yet it fails to incorporate significant features of the policy or goals that it was supposed to address. Unclear and unspecified changes can cause great anxiety and frustration to those genuinely trying to

implement them.

Complexity refers to the difficulty and extent of change required of the individuals responsible for implementation. Complex changes “require a sophisticated array of activities, structures, diagnoses, teaching strategies, and philosophical understanding if effective implementation is to be achieved” (Fullan, 2007: 91). While complexity creates problems for implementation, it may paradoxically result in greater change, because more is being attempted. Berman and McLaughlin (1976) found that “ambitious projects were less successful in absolute terms of the percentage of the project goals achieved, but they typically stimulated more teacher change than projects attempting less” (p. 88). Those changes that did occur were more thorough as a result of the extra effort that the project required or inspired. As Berman (1981) stated elsewhere, “little ventured, nothing gained” (cited in Fullan, 2007: 91). In summary, simple changes may be easier to carry out, but they may not make much of a difference. Complex changes promise to accomplish more, but they also demand more effort and such changes take a greater toll when there is failure. Fullan accordingly suggested breaking complex changes into components which can be implemented in an incremental manner.

On the issues of quality, Fullan observed that quality may be compromised, especially in politically-driven projects, simply because the period between the decision to initiate and startup is often too short to allow for adequate quality assurance. When adoption is more important than implementation, decisions are frequently made without the follow-up or preparation time necessary to generate adequate materials. People do not learn or accomplish complex changes by being told what to do. “Deeper meaning and solid change must be born over time” (Fullan, 2007: 92). It is important to attempt substantial change and to do it by persistently working on multilevel meaning across the system over time.

Besides the characteristics of an innovation itself, the context of a proposed innovation also influences the process of diffusing and implementing it. Within the local characteristics listed by Fullan (2007), the teacher factor seems of particular relevance to this study. Fullan stressed the influence of beliefs and attitudes, and Karavas-Doukas (1996) restated that

“attitude change is an essential and inevitable part of any pedagogical innovation” (p. 188). Huberman (1988) and others found that a teacher’s psychological state could make him or her more or less predisposed toward considering and acting on improvements. Some teachers, depending on their personality, and influenced by their previous experiences and stage of career, were more self-actualised and had a greater sense of efficacy, which led them to take action and persist in the effort required to bring about successful implementation. Freeman (1989) also observed that teacher change and development required awareness of the need for change and of available alternatives, which were coloured by that teacher’s experience and philosophy of teaching (cited in Pennington, 1995: 705). Also significant for some was an interest in encountering new ideas which seemed relevant to their situation. Bailey (1992) noted that the experience of the teachers in her sample illustrated “a complex combination of internal and external motivations for change, interacting with the context” (p. 271).

Henrichsen (1989) suggested that a full understanding of the diffusion and implementation of innovation processes required not only an examination of the innovation itself, but also an in-depth examination of (a) the role of the change agent (e.g. policymakers), (b) the role of the adopter (e.g. departmental directors, teachers and students), (c) the various stages of the innovation diffusion process (e.g. decision making, adoption, implementation, diffusion), and (d) the local constraints within which reformers operate. In short, it is important to understand both the sociocultural context and the concerns of the stakeholders in the innovation process, the length of time that is often required for successful innovation, and the odds against actually achieving success (Andrews, 2004). Without an understanding of these varied components of the innovation process, well-intentioned educators may find proposed innovations “seriously threatened or, even worse, thwarted” (Stoynoff, 1989: 17).

Concurring with Fullan’s (2007) three sets of interactive factors affecting innovation implementation, Brindley and Hood (1990) saw political as well as individual and professional factors as being influential in stimulating and promoting innovation. They concluded that for a change to get under way, there had to be an educational need in a political agenda. Rational argument alone would not bring about change; individuals needed to

experience the change personally and innovations needed local advocates. The social and political climate in which the change took place would significantly affect the extent of implementation. They also suggested that the intention of the proponents of change would frequently be misunderstood. Moreover, successful change required adequate support and coordination as well as a positive professional atmosphere. The resentment and resistance that teachers felt towards external attempts to impose change (Goh, 1999) must be compounded when no discussion or collaboration has taken place (Fullan, 2007; Hadley, 1999). Change had to be supported by ongoing professional development. Easen (1985: 71) commented that imposed change itself would not be successful, unless the process of personal change was also considered. Li (2001) also suggested and Carless (1999) confirmed that “without sufficient retraining, even teachers initially enthusiastic about an innovation can become frustrated by problems in implementation and eventually turn against the project” (p. 23). Therefore, teacher development was central to the innovation process.

To summarise, the key aspects in the educational innovation process are as follows: relevance and feasibility; compatibility; knowledge; awareness of the impact of external factors; discussion and collaboration; and adequate support and training. The importance of the role of teachers is crucial. The link with personal change and the process of change itself is also clear. As Fullan (2007) noted, innovation should be seen as a process rather than as an event and all the participants who are affected by an innovation have to find their own ‘meaning’ for the change. He further stated that the ‘subjective reality’ which teachers’ experience would always contrast with the ‘objective reality’ that the proponents of change had originally imagined. According to Fullan (2007), when teachers are expected to carry forward an innovation that is generally not of their own making, their lives can become very difficult because of little consultation with experts, little time to reflect on better solutions, and little time to achieve their goals. This may explain why intended washback did not occur in teaching and learning in some contexts. If educational change is imposed upon those parties most directly affected by the change (that is, learners and teachers), without consultation with them, resistance is likely to be the natural response (Curtis, 2000).

It has been found that in such cases there are discrepancies between the aims of an innovation or curriculum change and the understanding of teachers who have to implement the change (Andrews, 2004; Markee, 1997). The effect is one of two entirely different worlds colliding – the policymakers on the one hand and the local practitioners on the other. To the extent that each side is ignorant of the subjective world of the other, changes will fail. The quality of relationships across this ‘gulf’ is crucial to supporting change efforts when there is agreement, and to reconciling problems when there is conflict among these groups: between governmental ministries, local administrators and teachers. The relationship between local institution systems and external authority agencies is very often in the form of episodic events, rather than continuous processes: submission of requests for money, intermittent progress reports on what is being done, external evaluations – paperwork, not people work. Lack of role clarity, ambiguity about expectations, absence of regular interpersonal forums of communication, ambivalence between authority and the support roles of external agencies, and solutions that prove worse than the original problems combine to erode the likelihood of implementation. The difficulties in the relationship between external and internal groups are central to the problem and process of innovation. Not only is successful innovation hard to come by when two different worlds have limited interaction, but misinterpretation, attribution of motives, the feeling of being misunderstood, and disillusionment on both sides are almost guaranteed.

As has already been noted at several points in this review, bringing about any kind of educational change can be an extremely long, complex and difficult business (Brindley, 2008; Farrell, 2000). Research into washback has consistently shown that tests are not, of themselves, necessarily effective as “levers for change” (Pearson, 1988); successful educational innovations require both concerted system-wide reform and extensive support from those affected (Wall, 2000). It is clear that washback involves complex interactions between tests, teachers and learners, which determine whether individuals will embrace or reject the intended change. The studies in this section make it clear that, to study change in higher education in Taiwan, it is important to try and establish if there are gaps between the

expectations of stakeholders at different levels within the Taiwan educational system regarding the diffusion and implementation of innovation processes, and if gaps are found, to identify the factors involved in the success or otherwise of innovation initiatives.

3.8 Exploring washback within research models of teaching and learning

As has been noted in the previous sections, washback can be a powerful educational phenomenon. It can not only influence different people at different levels within educational contexts, but it can also have an impact on many aspects of teaching and learning in the institutional curriculum. Studies into such educational phenomena accordingly need to draw on research findings about teaching and learning, make references to the existing models of teaching and learning, and also draw on theories in educational change, since the notion of change is the fundamental principle of the washback phenomenon.

3.8.1 Classroom research

A classroom, in its purely locational sense, is a room in which teachers and learners gather together for instructional purposes (Allwright & Bailey, 1991). Van Lier (1988) broadened the characterisation of the language classroom suggesting that “The L2 [second language] classroom can be defined as the gathering, for a given period of time, of two or more persons (one of whom generally assumes the role of instructor) for the purposes of language learning” (p. 47). Therefore, classroom-centred research (Allwright & Bailey, 1991; Dörnyei, 2007) has focused on the classroom, as distinct from research that concentrates on the inputs to the classroom, such as the syllabus and teaching materials, or on the outputs from the classroom, such as learner test scores. It has investigated what actually happens inside the classroom when teachers and learners gather together for instructional purposes. It has focused on the

process variables of the classroom and answered questions such as how and why it is that things happen as they do in the classroom. The notion of classroom-centred research is relevant to the aims of the current study because the objective is to investigate how washback works in the classroom and in the educational context as a whole. The study thus needs to include a focus on teacher and student interaction in the classroom, practice opportunities, and time allocation of teaching and learning activities, in order to understand how these aspects of teaching and learning might be influenced by the English proficiency tests and what differences the examinations may have actually brought about.

On the question of approaches to second language classroom research (see Allwright, 1988; Bailey, 1985; Chaudron, 1988; Dörnyei, 2007; Dunkin & Biddle, 1974; Gibbons, 2006), Chaudron (1988) distinguished four different traditions. The evidence-based approach was applied in early evaluations of L2 instruction. Centring on the use of experimental methods, the evidence-based approach involved pre- and post-treatment tests with experimental and control groups. Context and product components, which investigated the quantitative relationships between various classroom activities or behaviours and language achievement, were essential to this approach. This tradition has been criticised for its neglect of process variables in the classroom and for trying to establish a linear relationship between input and output variables via quantitative analysis (Allwright, 1988; Gibbons, 2006). This approach, accordingly, is unlikely to be ideal for studying washback as washback is far too complex to have a linear input and output relationship.

Interaction analysis, the second tradition, refers to observational schemes used to describe teacher-student interaction in the classroom. It was developed by the mid-1960s when the influence of sociological investigations of group processes led to the development of systems for the observation and analysis of classroom interaction in terms of social meanings and an inferred classroom climate. It was, however, criticised for the limited capacity of those systems to capture the complete picture of classroom interaction (Bailey, 1985; Chaudron, 2000). For the current study, this form of analysis is likely to be too fine-grained to capture and understand the multidimensional phenomenon of washback, but it

could be useful for designing the observation scheme.

Influenced by the second tradition, which viewed interaction as a chain of teacher and student behaviours, discourse analysis was an attempt to analyse the discourse of classroom interaction in structural-functional linguistic terms. This approach involved studying classroom transcripts and assigning all utterances to predetermined discourse categories and frequently involved the assumption that one utterance can only have one function (see Coulthard, 1977; Seedhouse, 2004: 55-66; Sinclair & Coulthard, 1975). Such analysis is likely to be too detailed for the purposes of this research, as the focus of this study is not so much on the language but rather the behaviours.

The fourth tradition, the ethnographic tradition, arose from sociological and anthropological traditions. It attempted to interpret behaviours from the perspective of the individual participant's understanding, rather than from the observer's or analyst's supposedly objective analysis. Classroom ethnography involved naturalistic 'uncontrolled' observation and provided extensive empirical descriptions of what was happening in the classroom by generating a description that approximated the knowledge of participants in a particular event, making, in the words of Robson (2002), the implicit explicit, the invisible visible. Although ethnography has often been criticised for its lack of generalisability to other contexts (Chaudron, 2000; Cohen, Manion, & Morrison, 2007; Denscombe, 2007), it can be a very appropriate approach for gaining a feel for the context of a study. Van Lier (1988) made a convincing case for ethnography as potentially the most useful means to study classroom phenomena. He stressed that "to understand what happens in classrooms, researchers must try to get at the meaning given to these events by the participants themselves" (cited in Allwright & Bailey, 1991: 5). As a result, some sort of ethnographic approach would seem to be useful for the present study, since it is clear from the argument in the previous sections that observing teachers' and students' classroom behaviours and gaining an understanding of the environment surrounding them are central to researching washback and such information can best be gained by observing 'real life' settings.

To this end, no one of the previously discussed approaches alone would be adequate to

satisfactorily carry out a study of washback on teaching and learning. Different approaches to classroom research lead to different techniques for both data collection and analysis. They also lead to the creation of different models of teaching. In addition, Chaudron's (1988) four-way classification of classroom research was somewhat misleading because interaction and discourse analysis were techniques rather than traditions. In fact, these two techniques can be used within either of the other two traditions (Nunan, 2005). Bailey (1999b) thus accepted the notion that there were two dominant traditions, the experimental and the naturalistic. However, like the critiques of the distinction between qualitative and quantitative approaches to research in the social sciences, the two-way distinction is in itself problematic, leading to the consideration of a mixed-methods strategy which will be discussed in detail in Chapter 4.

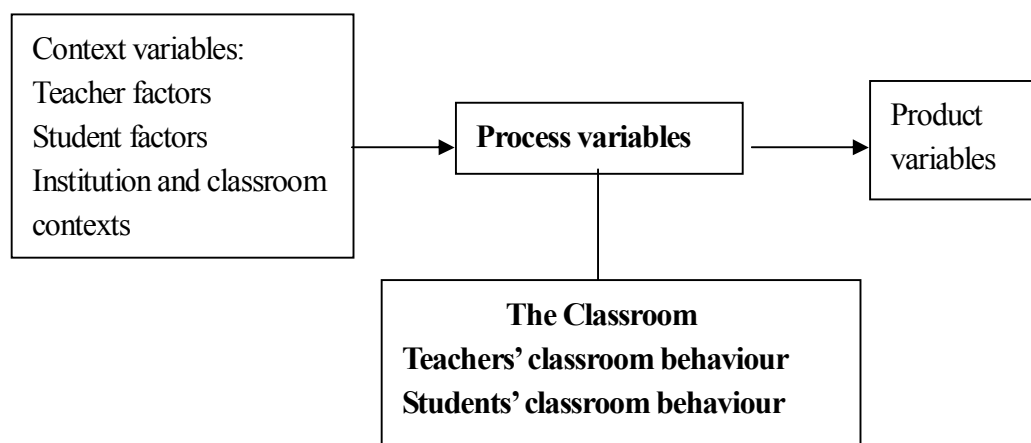
A major methodological dilemma that has bedevilled classroom research for years is how to capture differences in what goes on at the level of the classroom. Classroom-centred research is generally carried out by observation and/or introspection (Dörnyei, 2007; Gibbons, 2006; Mackey & Gass, 2005). Some form of observation is essential for classroom research, and this involves keeping a record of what goes on in the classrooms being observed. Introspection has usually involved either interviews or written questionnaires, to which people respond and where they reflect on their experience, or else diary keeping. Various observation instruments have been developed to explore a range of variables based on different models of teaching. However, I could find no observation instruments publicly available for washback studies of this kind, although there have been washback studies involving classroom observations (see Alderson & Hamp-Lyons, 1996; Cheng, 2005; Hamp-Lyons & Brown, 2007; Wall & Alderson, 1993).

As Rea-Dickins and Scott (2007) pointed out "washback can be viewed as a context-specific shifting process, unstable, involving changing behaviours in ways which are difficult to predict" (p. 5), the current study needs to explore teaching and learning in a classroom setting and a combined research framework is required, drawing on the various strengths of the four traditions mentioned above and taking classroom process as a focus.

3.8.2 Classroom research models

The model for the study of classroom teaching suggested by Dunkin and Biddle (1974: 38) serves as a useful initial guide for the clarification of the variables and behaviours in educationally motivated studies of teaching. Groups of variables concerning classroom processes, which were teachers' and students' classroom behaviours, and outcome products were related to one another, along with variables involving the context variables concerned with teachers' and students' backgrounds and characteristics and the surroundings. Figure 3.2 illustrates the teaching and learning model for the study of classrooms at different levels and the various issues involved.

Figure 3.2 Variables for classroom research
(Source: modified from Dunkin & Biddle, 1974, p. 38)



The study of washback on teaching and learning firstly needs to involve a broad classroom setting (see Wall, Kalnberzina, Mazuoliene, & Truus, 1996). Variables such as context variables, process variables and product variables should then be investigated one by one, if a complete picture of an examination's effect on teaching and learning is to be seen

within the local context (see Dunkin & Biddle, 1974; Bailey, 1996; Hughes, 1993; Stern, 1989). The present study needs to focus primarily on process variables (written in **bold** in Figure 3.2) in the classroom, where teachers and students interact with each other – what teachers do and what students do in the classroom.

Whereas Dunkin and Biddle's (1974) model provided basic concepts for the study of classroom teaching, Stern (1989, 1992) adopted broader political and multilevel interaction perspectives and suggested a set of categories for specifically analysing language teaching. His model, presented in Figure 3.3, consisted of three levels. The basic theoretical or philosophical level related to the fundamental assumptions of language learning and teaching made by the teacher, the evaluator, the administrator, the researcher, and even the learners themselves. It consisted of four concepts: language, society, learning and teaching.

Figure 3.3 A language teaching analysis scheme
(Source: Stern, 1989, p. 210)

Level 3	Teaching and learning activities Classroom behaviour/classroom observation
	Behavioural or surface level
Level 2	General categories of language teaching
	Policy level
Level 1	Fundamental concepts
	Basic theoretical or philosophical level

The second level, the policy level, was rooted in teaching methods and curriculum design theories. It was where professional planning and decision-making occurred, and consisted of four general categories of language teaching: content, objectives, treatment or procedures, and evaluation. The treatment variables, the aspects of language pedagogy, included three strategies which were highly relevant to the current study of the washback effect on aspects of teaching and learning. They were: teaching strategies, such as listening,

reading, speaking, and writing; timing strategies, such as segmenting lessons into larger or smaller chunks; and social and interpersonal strategies, such as teacher- or student-centred, individual- or class-based.



The third level, the behavioural or surface level, revealed the theoretical concepts at the basic level and the policy decisions at the second level through the behavioural acts of teaching and learning. The three levels formed a scheme for language teaching analysis, which reinforced the need to relate observation schemes to the underlying policy decisions and fundamental assumptions of language teaching and learning. This layering approach is particularly relevant to the current study, as one key aim is to investigate how teachers reflect proposed innovations in their teaching in their classrooms. This model can serve as a guide for the data collection and analysis in the current study.

Stern's (1989, 1992) model for analysing language teaching sees washback as bringing changes from the policymakers, involving conceptual change and decision-making, to the teaching and learning level. Washback functions on multiple levels and perhaps at different times. This is consistent with the later views of Fullan (2007). Fullan (2007) strongly argued the need for teaching and learning to be studied within multiple perspectives and made it clear that educational change was multidimensional. His ideas concurred with Bailey's (1996) model of washback mechanisms and Markee's (1997) curricular innovation model along with various innovation diffusion and implementation models mentioned in Sections 3.4 and 3.7.

Drawing on the models of classroom research, the diffusion and implementation of educational innovation and washback mechanisms discussed in this chapter, one can construct a simple model of washback that includes the main characteristics of the Taiwan situation, presented in Figure 3.4. It should be noted that because washback brought new ideas from the top (the theoretical and policy level) for conceptual change and decision-making, down to the classroom (the teaching and learning level), Stern's (1989) model was presented in reverse order. Figure 3.4 shows the working model for the study that follows. The curriculum as planned at the decision-making level refers to the guidelines of The Enhancement of Students' Foreign Language Proficiency Project. The 'curriculum in

action’ and ‘curriculum as outcome’ elements are to be studied at intervening and implementing levels. In between, there may be gaps or mismatches at each level.

Figure 3. 4 Explanatory model of researching washback in the present study

Levels	Participants	Questions to ask
Level 1: Decision-making agencies	Ministry of Education: The Enhancement of Students’ Foreign Language Proficiency Project	What strategies did the Ministry of Education use to implement the project?
Gap 		
Level 2: Intervening agencies	Universities (departmental directors and teachers): Curriculum Departmental policies	1. How did teachers realise the teaching and learning theories behind the implementation of English proficiency tests? 2. How did teachers arrange their lessons and activities accordingly?
Gap 		
Level 3: Implementing agencies	Universities: Teachers Students	How were teaching and learning activities carried out in classrooms?

3.9 Summary

To summarise, an attempt has been made to clarify the definition, scope, and function of washback for the purpose of this study. Washback is at the heart of the intricate relationship between testing, teaching and learning. It also illustrates the impact and power of tests on teaching and learning in educational contexts. However, whether the influences of testing on teaching and learning are positive or negative is still debatable and needs to be studied further.

A large number of studies have dealt with the phenomenon of washback from different perspectives and at multiple levels. There have, however, been few empirical analyses that have investigated how the washback phenomenon actually happens in the classroom. There have been even fewer research studies that have considered washback at both the macro and micro levels, particularly in language education.

Discussions in this chapter have reviewed a number of studies in searching for the meaning and mechanism of the function of washback, including Alderson and Wall's (1993) 15 washback hypotheses, Bailey's (1996) model based on Hughes' (1993) trichotomy, Wall's (1996) and Markee's (1997) models of the mechanism of washback as a phenomenon of change in teaching and learning. These models helped to determine the nature of washback and how washback works in educational contexts and they seem particularly appropriate, as the general aim of the present study is to chart and understand the function of washback in Taiwanese universities of technology. Together, those models have allowed me to formulate the central issues that will be explored in the current study. By combining the models, possible washback effects in an area, or in a number of areas, of teaching and learning affected by tests can be investigated. Accordingly, a study of the effects of washback needs to draw on curriculum and innovation models and explore the phenomenon within a multidimensional context. A working framework for this study has been built on this basis, as presented in Figure 3.4.

In the light of issues raised in previous studies, it was clear that a study looking at what and how English proficiency tests influenced teaching and learning at universities of technology in Taiwan would need to focus on the following dimensions: what possible areas of English teaching and learning have been affected by proficiency tests; how different levels of stakeholders within the Taiwanese educational system have reacted when washback occurred; defining the interrelationship between who changes what, how, when, where, and why (Markee, 1997).

Chapter Four: Research Methodology

4.1 Overall research design

In this chapter I begin by establishing three main research questions, along with some sub-questions and comments. I then discuss how best to answer them, and set up a research design for the study by discussing the theoretical considerations, the methodological considerations, and the research strategies and methods adopted.

4.1.1 Research questions

The general purpose of this study was to investigate if and how the imposition of English proficiency tests as a graduation requirement in universities of technology in Taiwan had influenced classroom teaching, including aspects of teachers' attitudes, teaching content, and classroom interaction, along with gaps or mismatches between stakeholders' expectations and actions. The discussion in the previous chapters suggested that the following three research questions (RQ1, RQ2 and RQ3) would be appropriate:

RQ1. What strategies did the Taiwanese Ministry of Education (MOE) and university departments use to implement the project on The Enhancement of Students' Foreign Language Proficiency (where *strategies* are defined as the processes and methods employed in the decision-making procedures)? This may usefully be split into three sub-questions:

RQ1.1 How did the MOE make the decision to develop the project on The Enhancement of Students' Foreign Language Proficiency?

RQ1.2 Why did some university departments make the decision to implement English proficiency tests as a graduation requirement, while others did not?

RQ1.3 How did different participants act during the various stages of the project?

Participants refers to different parties within the Taiwanese educational context. The reason for asking RQ1.3 was that the nature of the implementation of the English proficiency tests in the various universities of technology might have affected people's perceptions and attitudes towards teaching and learning in different ways. Therefore, the study needed to look at the Taiwanese educational system as a whole and establish as far as possible the reactions and actions of different participants in the context of the innovation. The main participants that needed to be included in the study were:

- the decision-making organisation – the Department of Technological and Vocational Education within the Ministry of Education;
- implementing agents – departmental directors, teachers, and students.

RQ2. What was the nature and scope of the washback effect on teachers' and students' perceptions of aspects of teaching in relation to English proficiency tests?

Teachers' and students' perceptions are defined as teachers' and students' understanding of and concerns about aspects of classroom teaching in relation to English proficiency tests.

The aspects of teaching that needed to be studied included:

- teachers' perceptions of the rationale for, and formats of, English proficiency tests;
- teachers' perceptions of the teaching methods, activities, use of mock exams, and of textbooks employed in the context of implementing English proficiency tests;
- students' perceptions of their teachers' talk, medium of instruction, and teaching activities, as well as their own learning activities, the use of English inside and outside class, and their motivation and opinions about their English lessons;
- teachers' perceptions of the learning strategies and activities they would like their students to use, or do, in relation to English proficiency tests.

RQ3. What was the nature and scope of the washback effect on teachers' behaviours as a result of implementing the English proficiency tests?

Teachers' behaviours are defined as what teachers do in the classroom. The aspects of teachers' behaviours in the classroom that needed to be studied included:

- teachers' medium of instruction, teacher talk, teaching activities;
- teaching materials used in teaching, aspects of lesson planning.

4.1.2 Research framework

The research questions clearly emphasise "the importance of context, setting, and subjects' frames of reference" (Marshall & Rossman, 2006: 54). The methods used thus needed to capture the reality, variation, and complexity of differences in day-to-day classroom practice, as well as differences within the local Taiwanese educational context. In order to understand the actions and practices involved, the researcher needed to engage directly with the local scene, spend sufficient time to understand actions in their specific social context, and gain, as far as possible, access to participant meanings (Robson, 2002). Without careful grounding in the local context, an adequate degree of understanding would be difficult, if not impossible. This study therefore needed to investigate in depth the educational context, as well as particular areas of teaching and learning. Two educational contexts which were likely to be affected by the imposition of English proficiency tests were particularly important:

- The decision-making organisation, consisting of the Department of Technological and Vocational Education within the Ministry of Education;
- Implementing agencies, particularly teachers, students, and departmental directors.

The Department of Technological and Vocational Education within the Ministry of Education made the decisions about the project on The Enhancement of Students' Foreign Language Proficiency with its advisory committee members, which mainly consisted of language experts from tertiary institutions. After the decision-making stage, the departmental

directors and teachers within universities decided how they were going to carry out their General English curriculum and actual teaching based on the governmental grant project.

Smith et al. (1994), however, found that policymakers' definitions of the problems which a given policy should resolve, as well as the nature of curriculum, assessment, and educational changes which are proposed are often translated imperfectly by practitioners. Teachers and directors have repeatedly been found to redefine and reinterpret the messages about policy that they have received. They have then acted – adapted, taught, learned, and evaluated – according to their own definitions of the situation (Blumer, 1986; Brindley, 2008; Geisinger, 1994). These research studies thus suggested that observing the local educational context was crucial to allow one to identify gaps or mismatches between the top level and the actual classroom teaching. Accordingly, in the present case an understanding of the perception of relevant gaps or mismatches within the Taiwan educational context should greatly improve the knowledge and understanding of how and in what areas English proficiency tests can actually influence the institutional curriculum.

After identifying the parties to be studied, I needed to establish the main research methods of the study. The discussion of teaching and learning models in Chapter 3 suggested that several models could contribute concepts which would be useful to the present study:

- Dunkin and Biddle (1974: 38): three key variables – context, process and product – in the study of teaching;
- Stern (1989: 210): three levels of analysis – teaching and learning activities (surface level), categories of language teaching (policy level), and fundamental concepts (basic theoretical level).

Besides the language teaching models, the curriculum models suggested by Fullan (2007) and Markee (1997), as well as Bailey's (1996) model based on Hughes' (1993) trichotomy of washback mechanisms, were also drawn upon to allow a broader perspective of the complex nature of washback. A detailed description of the models was given in Chapter 3.

As indicated in Chapter 2, educational decision-making in Taiwanese higher education is not wholly determined by current government policies. Decisions about whether to make

English proficiency tests compulsory may be made by a high-level university committee with reference to political considerations, but university departmental directors and teachers have regularly made decisions about parts of the curriculum. Therefore, several different perspectives needed to be considered. One perspective was concerned with what *should* happen in programmes (related to the policymakers' intended washback effects), while another perspective was concerned with what actually *does* happen in programmes (actual washback in the university setting). In short, washback needed to be studied in relation to what should happen and what actually happens in programmes and at different levels of the educational context.

4.1.3 Methodological considerations from other washback studies

The review of washback studies in Chapter 3 showed that methods employed to elicit data in previous research studies tended to involve questionnaires, interviews, testing measures and classroom observations. A number of studies have drawn on teachers' and/or students' self-report responses to questionnaires on the effects of standardised testing (e.g. Herman & Golan, 1991; Shohamy, 1992). Andrews and Fullilove (1994) and Qi (2004) investigated the effects of tests on teachers and classrooms by using interviews. A few researchers, for example Hayes and Read (2004) and Her (2007), administered language tests in order to measure washback on students' learning outcomes. Alderson and Wall (1993) used classroom observations, and argued strongly for an observation component in the data collection, in order to understand the nature of washback (see also Alderson & Hamp-Lyons, 1996; Wall, 1996; Watanabe, 2004). Similarly, Bailey (1999a) and van Lier (1988) both suggested that, if the core of washback has to do with the effects of tests on teaching and learning, it would be necessary to document those effects both by asking about and watching teaching and learning. Wall and Alderson (1993) later pointed out the complexity of washback and emphasised the importance of using combined methods to answer questions such as "why the teachers do

what they do, what they understand about the underlying principles of the textbook and examination, and what they believe to be effective means of teaching and learning” (p. 62).

The review of washback studies also showed that there seemed to be no instruments that had been developed specifically for washback studies. No single uniform questionnaire had emerged as being widely used to survey teachers and/or students about washback (Bailey, 1999a). According to Bailey (1999a), two factors accounting for the lack of a uniform questionnaire were firstly, that students had been surveyed in their native language, and secondly that studies had tended to focus on the washback from a particular local exam. The same was also true of classroom observation schemes for observing washback effects, though this might also be due simply to the limited numbers of washback studies using classroom observations (Cheng, 2008). The lack of existing uniform instruments for studying washback implied the need to design the instruments for the survey and classroom observation for any new study, including the present one. As a result, checking the validity and reliability of the instruments must thus rely, in part at least, on the triangulation of data sets, in order to compensate for the weakness of one instrument through the strength of another.

4.1.4 Research methodology for this study

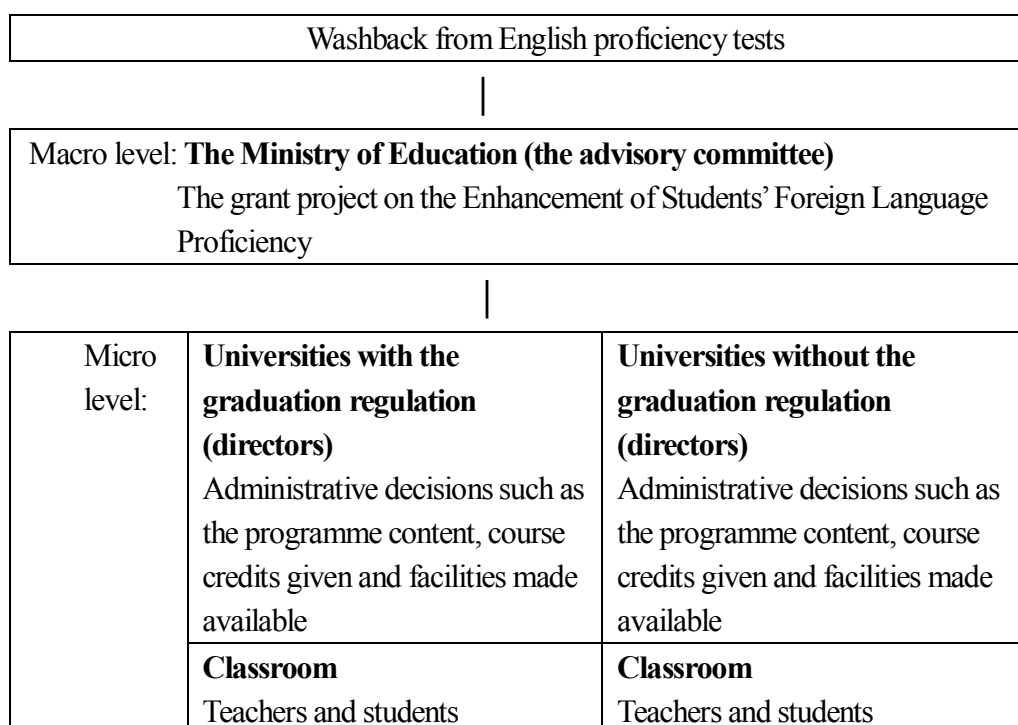
Drawing on the theoretical and methodological considerations described in Sections 4.1.2 and 4.1.3, the investigations for this study were carried out at different levels within the Taiwan educational context, and by employing a mixed-methods strategy.

Because my intention was to explore washback effects at different levels, focusing on perceptions, values, and situational factors in the complex and varying situations of the Taiwan educational context, the design of the study needed to capture both the big and small pictures, that is to say, the macro level of the Ministry of Education and the micro level of the universities of technology concerning different facets of teaching and learning, in which the most significant (or insignificant) differences occur (see Figure 4.1). Accordingly, I decided to

conduct a comparative study. The working framework is set out in Figure 4.1.

In order to find out differences between participants' perceptions, attitudes and opinions, I had to gain some knowledge of the participants' perspectives if I was to understand the actions and reasons of individuals. To comprehend the big picture, I needed to combine the aggregate knowledge of these individual situations with an understanding of institutional factors that influenced the process of innovation, such as the Ministry of Education and universities of technology. As discussed in Chapter 3 and previous sections, washback study was seen as a multivariate, or multifactor, endeavour that required the researcher to think of and address more than one factor at a time. As a result, a mixed-methods approach to data collection seemed to be appropriate, since there was no one all-purpose method of capturing such a complex phenomenon as different methods have different strengths and weaknesses. By using a range of methods, one can put together a more complete picture.

Figure 4.1 The Taiwanese education system relating to the English proficiency tests



Using a questionnaire was one possible method which could be employed. One advantage of questionnaires is that they tend to be reliable because they are anonymous. Questionnaires tend to be suited to involving a large number of subjects, as they allow a structured approach, but the result is that they can generate a large volume of numerical data. They have the possible advantage that they can be administered without the presence of the researcher, particularly when they consist largely of closed questions. These have the further administrative advantage that the answers to a questionnaire are often comparatively straightforward to analyse (Cohen et al., 2007; Gillham, 2007). The results are more comprehensive and representative than data obtained from studying individuals, but because the questions and possible answers are predetermined, the results can often be superficial without revealing unexpected answers and thus show a weak relation between what people said in the questionnaire and what they actually did (Brown, 2001: 77; Gillham, 2007; Robson, 2002: 233). Indeed, this lack of a relation between attitude and behaviour has become notorious (Dörnyei, 2003). As one of the research aims in the present context is to explore teachers' and students' reactions to, and perceptions of, implementing English proficiency tests as a graduation requirement, as well as classroom activities which might be influenced by the introduction of English proficiency tests, questionnaires would be an appropriate method in order to allow a considerable number of teachers and students to voice their opinions. Questionnaires could also be used to indicate where a more focused investigation should take place at later stages of the study.

Interviewing was another possible method which could be employed. Interviews usually provide qualitative descriptions and can, if done well, present reality in considerable detail. Although conducting interviews is time-consuming, there are several advantages to interviews that are applicable to this study. First, Gillham (2007) has suggested that interviews often have a higher response rate than questionnaires because respondents become more involved and motivated; they enable more to be said about the research study than in a written form. Thus interviewing is more likely to foster spontaneous and informative answers compared with questionnaires. Gillham (2007) has also pointed out that respondents may

leave questions unanswered while filling out questionnaires. Interviewers can, however, ensure that most questions are answered in interviews and can instantly clarify any misunderstandings experienced by the interviewees and probe further questions if necessary. Third, interviews allow interviewers to observe and gain insights from non-verbal responses (Denscombe, 2007). For example, a pause may contain implicit meanings that interviewers can detect and, if necessary, ask interviewees to explicate. In addition, interviews can draw out 'deep' information about participants' opinions, experiences, beliefs, and feelings about a topic (Denscombe, 2007). Fifth, interviews can be conducted at an appropriate speed and interviewers can rearrange or rephrase questions to adapt to each interviewee. Another important advantage is that interviews allow participants to provide historical information (Creswell, 2003). In the context of this study, departmental directors can describe what their programmes used to include and what they include at present.

Classroom observations can be used to discover further details of the impact of examinations, and to explore the relationship between testing and classroom teaching activities. Observation "refers to the data collected by a researcher who directly observes a study's research participant(s)" (Anastas, 2005: 214). Just like interviews, observation can, if done well, present a more accurate picture of reality and reveal more complexities, but it is a time-consuming process to capture the required behaviour or phenomenon and it can be difficult to interpret what the data mean (Cohen et al., 2007: 412). There are, however, several advantages to observation that are applicable to the present study. First, it would allow me to investigate the extent of the washback on teaching from the English proficiency tests, which may have a consequential impact on students' learning. For example, students may focus more on a skill that emphasized by teachers in classes. Second, observation can uncover unexpected variables which might affect the washback from English proficiency tests on teaching and learning. Third, observation can reveal behaviour that participants take for granted and fail to report in self-report methods, and can disclose information that participants are reluctant to share in the interviews (Patton, 2002). For example, teachers may not prepare students for English proficiency tests because they are loath to change teaching activities

and/or teaching methodology, and they may not be willing to reveal the truth in questionnaires or interviews. Another key advantage is that observation can attenuate the biases in questionnaires and interviews (Denscombe, 2007). Fifth, observation can cast light on interaction among participants and on individual behaviour, in a way that questionnaires and even group interviews or focus groups do not (Cohen et al., 2007), which is especially applicable to classroom-based research (Hays, 2004).

It seems clear that, in the context of the present study, with both teaching and learning variables, no one method by itself would be sufficient to answer all the research questions. Mixed-methods approaches are both feasible and fairly common in educational research. Indeed, Brannen (1992) argued that quantitative and qualitative data analyses are mutually dependent. Wall and Alderson (1993) also suggested that “observations on their own cannot give a full account of what is happening in classrooms”:

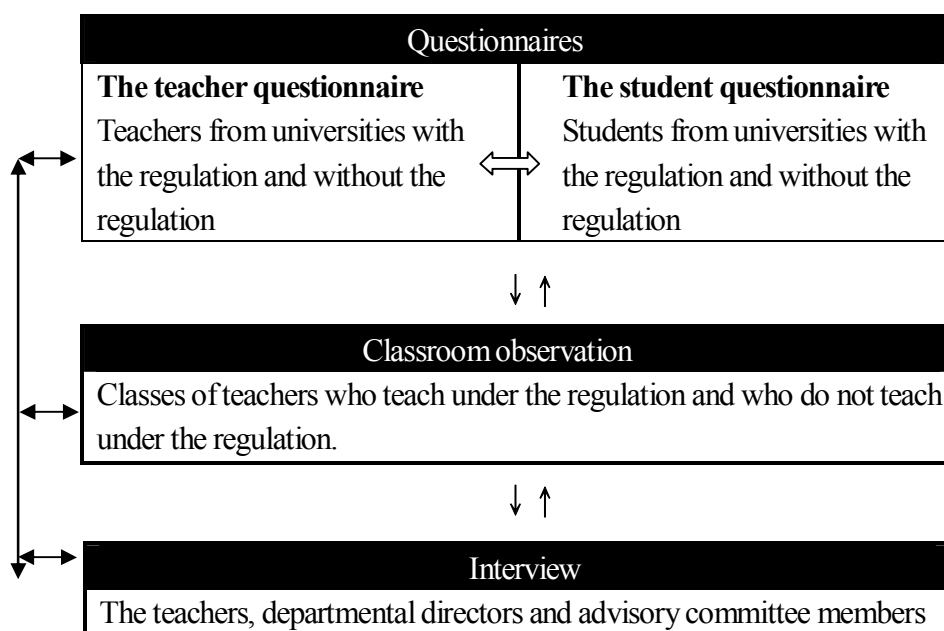
It was important for us to complement the classroom observations with teacher interviews, questionnaires to teachers and teacher advisers, and analyses of materials (especially tests) teachers had prepared for classes. (p. 63)

A mixed-methods approach which has both qualitative and quantitative data would allow questionnaire data to be checked and refined in context. The context, however, could only be understood in depth by study of the situation using qualitative descriptions. Interviews could be used to reflect the participants’ opinions and discover invisible meanings of teaching behaviours. These three methods complement each other and could be integrated relatively easily in practice.

This type of mixed-methods strategy also provides for cross-examination mechanisms, often referred to as triangulation. Denscombe (2007) suggested that a mixed-methods strategy can check on the potential bias in any single research method and can be used to compensate for an individual method’s weaknesses. In addition, since individual research methods can uncover the specific facets of research topics (Patton, 2002), mixed-methods data collection and analysis can provide complementary data and tackle an issue from different angles

(Denscombe, 2007). If different types of data lead to the same conclusion, one can be a little more confident in that conclusion (Cohen et al., 2007; Gillham, 2007).

Figure 4. 2 Data and method triangulation for the study



*Arrows indicate the direction of triangulation.

Various approaches to triangulation can also be employed to increase the quality control and representativeness of the study (Allwright & Bailey, 1991; Cohen et al., 2007; Dörnyei, 2007; Gorard & Taylor, 2004; van Lier, 1989a, 1989b) and two were particularly relevant to the present study. The approaches appropriate to this study were *methodological triangulation* and *data triangulation*. The former refers to using different methods on the same object of study, whereas the latter refers to data from more than one source being brought to bear in answering different aspects of one research question. Denzin (1970, 1997) further identified two categories in his typology of the methodological triangulation: 'within methods' triangulation and 'between methods' triangulation. The former involves the replication of a study as a check on reliability and theory confirmation, and the latter concerns the use of more

than one method in the pursuit of a given objective. Of them, between-methods triangulation was particularly relevant to the present study. As a check on validity, the between-methods approach embraces the notion of convergence between independent measures of the same objective (Cohen et al., 2007).

Accordingly, in order to complement each other and compensate for an individual method's weaknesses, a mixed-methods strategy is employed in the present study. The methods used are questionnaires, interviewing and classroom observations. Both between-methods triangulation and data triangulation approaches are used to examine the methods and the data. The triangulations involved are set out in Figure 4.2.

4.1.5 Research phases

Taking into account the time constraints of a doctoral study, the empirical parts of the study were carried out in three stages (see Figure 4.3).

Phase I lasted from June to July, 2007. This was the pilot study stage, which was designed to test out the research methods and instruments. The major research methods used in this phase were:

- key informant interviews
- classroom observations
- questionnaire administration

Phase II lasted from February to June, 2008. This phase was the perception stage, the findings for which are reported in Chapter 5 (Teacher Questionnaire) and Chapter 6 (Student Questionnaire). The major research methods used in this phase were the two parallel surveys comprising the teacher and student questionnaires.

Phase III of the study lasted from April to June, 2008 and overlapped with *Phase II*. It focused on the classroom to observe closely what was actually happening inside the 'black box' (Long, 1980) in relation to the implementation of English proficiency tests. Detailed

follow-up interviews with teachers, interviews with department directors, and interviews with advisory committee members in the Ministry of Education were also carried out. This phase was the implementation stage and the findings are reported in Chapter 7 and 8. The major research methods used in this phase were thus classroom observations and interviews.

Figure 4.3 The three phases of the research study

<p>Phase I</p> <p><i>Stage:</i> Pilot study</p> <p><i>Time:</i> June 2007 – July 2007</p> <p><i>Method:</i></p> <ul style="list-style-type: none"> ● key informant interviews ● classroom observations ● administration of questionnaires to teachers and students
<p>Phase II</p> <p><i>Stage:</i> Perception of English proficiency tests</p> <p><i>Time:</i> February 2008 – June 2008</p> <p><i>Method:</i></p> <ul style="list-style-type: none"> ● survey of attitudes towards the implementation of English proficiency tests, and their implementation in the teaching and learning of English in Taiwanese universities of technology. ● survey of university students' attitudes towards the implementation of English proficiency tests, in relation to the activities they carried out in their English lessons.
<p>Phase III</p> <p><i>Stage:</i> Implementation of English proficiency tests</p> <p><i>Time:</i> April 2008 – June 2008</p> <p><i>Method:</i></p> <ul style="list-style-type: none"> ● classroom observations of teachers' behaviours in the actual teaching and learning contexts. ● detailed interviews with the teachers who participated in Phase II, interviews with departmental directors, and interviews with advisory committee members.

4.2 Research design for the pilot study and main study – the two questionnaires

Because even minor differences in the wording of a questionnaire can change the response pattern, piloting a questionnaire at various stages of its development on a sample of people who are similar to the target sample is necessary and important in survey studies (Cohen et al., 2007). These trial runs allow the researcher to collect feedback about how the instrument works and whether it performs the job it has been designed for. Based on this information, alterations can be made and the final version of the questionnaire fine-tuned.

At Phase I, two parallel questionnaire surveys, the teacher survey and the student survey, were piloted. Both studies explored the washback effect of English proficiency tests on aspects of classroom teaching and learning: perception and behavioural differences in the context of passing English proficiency tests as a graduation requirement, from both teachers' and students' points of view. The teacher questionnaire was conducted in English as the participants were teachers of English, whereas the student questionnaire was in Chinese to avoid any problems caused by the use of English as a foreign language. Both questionnaires are reported in the following way:

1. rationale and aim
2. structure and content
3. design and validation procedures

4.2.1 Teacher questionnaire

Rationale and aim

The purpose of the teacher questionnaire was to explore any perception and behavioural differences among teachers in relation to the implementation of English proficiency tests as a graduation requirement. It investigated whether or not any washback effect had occurred on

teachers' perceptions of aspects of classroom teaching and learning.

Structure, content, and sources

The teacher questionnaire (TQ) for the pilot study consisted of three parts and was designed in English (see Appendix 3.1). Part One (TQ 1-6) consisted of six questions about background information covering (a) demographic details, such as gender, age, academic qualifications and years of teaching experience, and (b) current teaching situations, such as English levels that were currently being taught and numbers of teaching periods per week.

Part Two consisted of 8 questions and comprised 72 items altogether (TQ 7-14). Each item was designed using a four-point Likert scale. The Likert scale was employed, as the questions need to be answered on an equal-interval scale in order to derive comparable indices and the Likert scale is one of the most commonly accepted scales in the education field (Cohen et al., 2007: 326). Using a four-point rather than a traditional five-point or seven-point scale is a way of controlling for the 'central tendency', that is, a tendency for participants with no strong feelings to opt for the mid-point of a scale. This has been found to be a particular problem with East Asian respondents, where the 'doctrine of the mean' is advocated in Confucian culture, so there are tendencies to respond on the basis of the form rather than the content of items (Aiken, 1996: 40; Cohen et al., 2007: 327). This part of the questionnaire mainly dealt with teachers' perceptions of aspects of teaching, learning, assessment, and evaluation in universities of technology in relation to the implementation of English proficiency tests.

Part Three consisted of 8 questions, which dealt with teachers' reactions to the new policy and to aspects of relevant classroom teaching and learning activities. Questions 15 to 19 were designed on a multiple-choice basis. Teachers were required to choose the appropriate answers according to their own classroom teaching and learning situations. Questions 20 to 22 were again designed using a four-point Likert scale of frequency.

Table 4.1 Structure and themes of the teacher questionnaire in the pilot study

Structure	Content	Items
Theme One	Personal particulars (TQ 1-6) <ul style="list-style-type: none"> ● demographic information, such as gender, age, academic qualifications, and years of teaching experience ● current teaching situations, such as English levels currently taught and the numbers of teaching periods per week 	6
Theme Two	Teachers' reactions and perceptions in relation to the implementation of English proficiency tests (TQ 15, 7-9) <ul style="list-style-type: none"> ● teachers' reactions to the implementation of proficiency tests ● the reasons behind the implementation of proficiency tests ● possible extra work and pressure resulting from the implementation of proficiency tests ● teaching methods teachers would like to change as a result of the implementation of proficiency tests 	4
Theme Three	Teaching materials (TQ 19, 22) <ul style="list-style-type: none"> ● textbook arrangements related to teaching materials (who and how) ● teaching and learning resources 	2
Theme Four	Teachers' classroom behaviours (TQ 16-18, 20-21) <ul style="list-style-type: none"> ● medium of instruction ● teaching arrangements (who plans and how to plan) ● lesson preparation ● teacher talk ● teaching activities 	5
Theme Five	Assessment and evaluation (TQ 11, 13-14) <ul style="list-style-type: none"> ● the use of mock exams ● the evaluation of teaching in Taiwanese technical universities ● factors that influence teaching 	3
Theme Six	Aspects of learning (TQ 10, 12) <ul style="list-style-type: none"> ● learning strategies ● learning aims 	2

There were six main research themes within the questionnaire. These are summarised in Table 4.1. Theme Two was mainly derived from departmental director and teacher comments collected in e-mail interviews at the preliminary information-gathering stage, lasting from November 2006 to April 2007. These interviews aimed to find the main areas that might be

influenced by the new policy. Themes Three and Five were derived from e-mail interviews with teachers at the preliminary information-gathering stage, departmental documents such as organisational projects and relevant documents announced by the Ministry of Education (see Chapter 2). Themes Four and Six were derived from personal experience in talking to departmental directors, teachers and students. They also drew on other relevant research studies (see Burrows, 2004; Cheng, 2005; Ferman, 2004; Qi, 2004; Watanabe, 2004).

Design and validation procedures

Two main methods were used in the questionnaire survey study, namely, qualitative input and piloting, to increase the validity of the items in the questionnaire. While qualitative input ensures content validity, piloting procedures ensure construct validity and reduce problems with the procedure of administering the questionnaire.

Method One: qualitative input. Qualitative input consisted of (a) theoretical considerations from related research studies reviewed in Chapter 3, which suggested modifications or additions to the research questions, instrument development, and plans for analysis of data, and (b) qualitative data from e-mail interviews with teachers and departmental directors. The emphasis of the qualitative input was to achieve a focus on the research questions and on the development of questionnaire questions.

Method Two: piloting. The pilot study was carried out in June 2007 in two universities of technology, which were members of the target population but not part of the final sample. Two Taiwanese teachers of English were asked to trial the questionnaire by attempting to answer the questions and individual interviews were used to allow them to comment on any problems they had and any issues that occurred to them as they did so. The purpose was to determine whether (a) the various items in the questionnaire were understood as intended by the researcher, and (b) they were understood in the same way by each teacher. It also aimed to pre-test the questions and items, along with ease of use, format and overall instrument

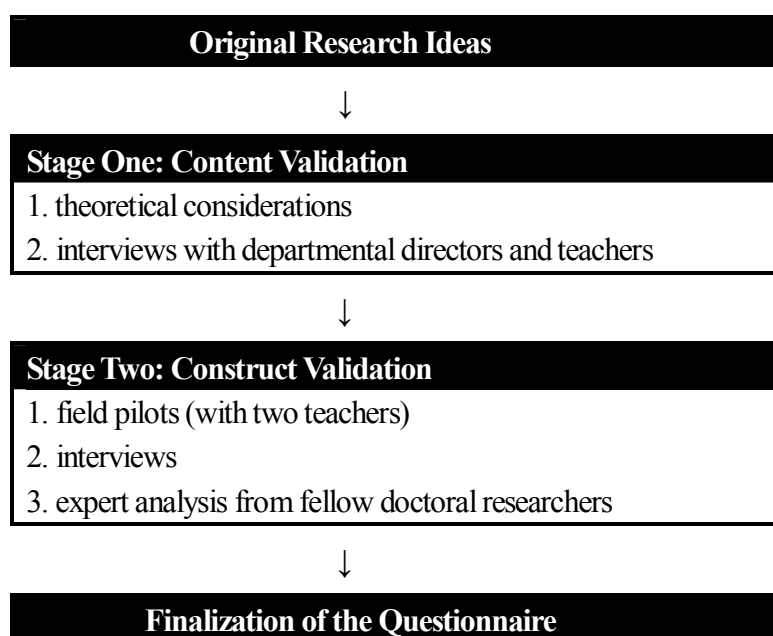
appearance, and to determine what was and was not working in the questionnaire. The procedure served as a method of construct validation for the questionnaire items and to evaluate specific question wordings.

The two teachers involved at this stage expressed doubts about some of the terminology in English and they felt that it took too much time to fill in the questionnaire (around twenty minutes). Both expressed impatience and lost concentration filling in the questionnaire. As a result, the content and layout of the final version of the teacher questionnaire for the main study was kept almost the same as the pilot version, but the terminology was altered, and a number of substantial adjustments and modifications were made to the wording and areas of redundancy. For instance, because students were not placed according to their English level at some universities, teachers might be confused by TQ 5. Therefore, TQ 5 was removed from the final version of the questionnaire. TQ 8 was confusing the teacher whose students did not face the graduation regulation. It was accordingly decided to leave out TQ 8 and I asked the question in the interviews. TQ 11 was rephrased, asking about personal experience rather than colleagues' experience. TQ 13 aimed to investigate whether teachers' teaching was assessed based on students' English proficiency test scores or not. The teachers mentioned that because of the liberalisation in higher education in Taiwan, universities are free to make their own decisions in line with the concept of academic freedom. None of the universities evaluated teaching based on the students' examination performance. As a result, TQ 13 was left off the questionnaire for the main study. Because of academic freedom, it was university teachers who usually made the decisions about lesson content and textbook choice. TQ17 and 19 were therefore removed and rephrased in terms of employing test-oriented materials.

Expert analysis was carried out on the pre-final version of the questionnaire for the main study. Before carrying out the expert analysis, I reviewed the questionnaire to note potential problems (see Brannen, 1992: 11; Bryman, 1992: 63-65) and identified the specific points where the respondent might have difficulties and the types of misinterpretations he or she might make. In the expert analysis, two research colleagues, both doctoral candidates in the area of language education, were invited to complete the questionnaire and voice their

opinions. One of the original two teachers was then asked to comment on the revised version and no further issues were raised at this stage. The final version of the teacher questionnaire in English included 20 questions as shown in Appendix 4.1. Figure 4.4 illustrates the stages of the questionnaire design together with the validation procedures discussed above.

Figure 4.4 Stages of the teacher questionnaire design



4.2.2 Student questionnaire

Rationale and aim

The purpose of the student questionnaire was to explore students' perceptions and behavioural differences that could be linked very closely with classroom teaching and learning in the context of implementing English proficiency tests as a graduation requirement. The student questionnaire explored the perceptions of classroom teaching and learning activities from the students' points of view. It also investigated the role of English proficiency tests in students'

learning. It was designed to provide data from the students' points of view, but also to serve as valid data for comparison (triangulation) with the data from the teacher questionnaire. Therefore, this questionnaire focused on the same main themes as did the teacher questionnaire. The questionnaire was given to university of technology students who would and would not have to pass an English proficiency test as a graduation requirement.

Structure, content, and sources

The student questionnaire (SQ) for the pilot study consisted of two parts (see Appendix 3.2). Part One consisted of six questions (SQ 1-6). It dealt with student demographic information in the survey and the learning contexts of students. Student demographic information included gender and grade. General information about the learning contexts included English level, the medium of instruction of the English lessons, the number of hours spent studying English each week, and the number of hours of private tutorials attended each week for English proficiency tests. Four questions in this part were designed on a multiple-choice basis and two were designed with an open format.

Part Two consisted of three sections with 11 questions and 78 items in total. This part dealt with students' attitudes towards teaching and learning activities inside and outside their English lessons. There were three sections, with a total of three themes. *Section A* (six questions, SQ 7- 12) dealt with classroom teaching and learning activities. Items in this section employed a four-point Likert scale. *Section B* (four questions, SQ 13-16) invited students' opinions about aspects of learning. Items in this section were designed on a multiple-choice format. *Section C* (one question, SQ 17) consisted of one question with 12 items. They were designed on a multiple-choice basis using a scale of agreement. These items were related to students' reactions to and views of the influence of English proficiency tests. There were two sub-themes in relation to English proficiency tests; items 1, 5, 8, 11 and 12 dealt with the impact of examinations on students themselves, while items 2, 3, 4, 6, 7, 9 and

10 dealt with the impact of examinations on students' learning processes and outcomes. The questions in the student questionnaire are described in terms of the main themes in Table 4.2.

Table 4.2 Structure and themes of the student questionnaire in the pilot study

Structure	Content and themes	Items
Part 1	Demographic information and learning contexts: gender, grade, English level, medium of instruction in English lessons, the number of hours spent studying English each week, and the number of hours of private tutorials attended each week for proficiency tests	SQ 1-6
Part 2	Theme 1: Classroom teaching and learning activities <ul style="list-style-type: none"> ● teachers' talk in class ● teachers' teaching activities in class ● teaching resources ● students' use of English in class ● students' use of English outside class ● students' learning activities in class 	SQ 7-12
	Theme 2: Perceptions of aspects of learning <ul style="list-style-type: none"> ● motivation for learning English ● preferred learning strategies ● whether or not students think they are influenced by proficiency tests ● aspects of examination influences on students 	SQ 13-16
	Theme 3: Students' reactions to and views of proficiency tests	SQ 17

The questions in the student questionnaire were mainly derived from a review of the second language classroom teaching literature, from e-mail interviewing with the departmental directors, teachers and students at the preliminary information-gathering stage, from departmental documents such as organisational projects, and from relevant documents announced by the Ministry of Education. The questions were designed parallel to the

questions in the teacher questionnaire with question wording adapted to the perspective of the respondents. The sharing of items across questionnaires enabled comparisons between the students and the teachers. Section C of Part Two of the student questionnaire was adapted from Gullickson (1984). His study focused on teachers' perspectives of their instructional use of tests and these were closely related to the students' learning in this study.

Design and validation procedures

To match the questions asked in the teacher survey, the student questionnaire was designed parallel to the teacher questionnaire. Therefore, the basic considerations for validation were the same as for the teacher questionnaire. Accordingly, both student and teacher questionnaires were designed to comprise two stages – qualitative input and the piloting procedure.

Qualitative input was applied in the same way as in the teacher questionnaire and consisted of theoretical considerations and interviews. The piloting and translation phases were carried out rather differently from those in the teacher questionnaire, however, as the student questionnaire was administered in Mandarin rather than English. The same two universities of technology that participated in the teacher questionnaire survey in the pilot study also participated in the student questionnaire survey in June 2007. 75 students, 31 of whom would sit for English proficiency tests and 44 of whom would not, were involved in the pilot study. The students were asked to trial the questionnaire by attempting to answer the questions and by writing comments on any problems they had, or any issues that occurred to them. The presence of their teachers and the researcher enabled students to ask questions about the questionnaire or the task while they were doing it. The purpose was to determine whether (a) the question items in the questionnaire were understood as intended, and (b) they were understood in the same way by all students. It also aimed to pre-test the questions and items and to determine what was and was not working in the questionnaire. The procedure

served as a method of construct validation of the questionnaire items and to evaluate specific question wording.

65 of the students involved at this stage did not mention any problems, while just 10 indicated minor problems with wording. The question wording was then adjusted and a number of substantial adjustments and modifications were made to the questionnaire items. For instance, SQ 6, 13, 15, 16 and 17 were rephrased. One item in SQ16 was ‘motivation to learn’ and the students said the meaning was vague. The item was thus revised as ‘more interested in learning English’ and ‘less interested in learning English’. The students also mentioned that items in SQ 9 overlapped with items in SQ 14 and they were reluctant to answer questions with similar items. SQ 9 was therefore removed since it would not affect the researcher’s intentions. The final version of the student questionnaire for the main study is presented in Appendix 4.2.

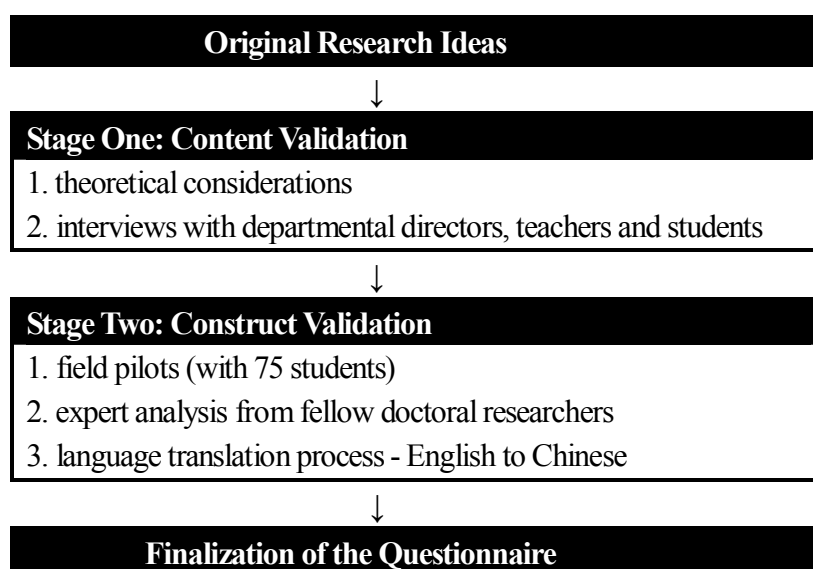
Translation procedures

The student questionnaire was originally designed in English, while it was being co-developed with the teacher questionnaire in the study. It was, however, given to the students in Chinese. The rationale behind this was to minimise misunderstanding and reliability and validity problems caused by the language factor (Geisinger, 1994). Therefore, in addition to using the same procedures that were used in the teacher questionnaire, extra efforts were made in the validation procedure of the translated version.

The Chinese version of the questionnaire for the pilot study was read by three Chinese-speaking doctoral researchers at the University of York to see whether anything was missing or could be misinterpreted. Expert analysis was also carried out on the pre-final version of the questionnaire for the main study. Two research colleagues, both doctoral candidates in the area of language education, were invited to complete the questionnaire and voice their opinions. Their views on the questionnaire structure, the design of the categories,

items, and wordings helped to shape the final questionnaire and prepare it for the main study. Ten of the original 75 students were then asked to comment on the revised version and no further issues were raised. The final version of the student questionnaire in Chinese for the main study included 17 questions. The design stages and validation procedures of the student questionnaire are summarised and illustrated in Figure 4.5.

Figure 4.5 Stages of the student questionnaire design



4.2.3 Sampling for both teacher and student surveys in the main study

It will be recalled from Chapter 2 that there are 38 universities of technology, located in northern, central and southern Taiwan. Of these, 9 out of the 38 are national universities and 29 are private.

In Taiwan, research access to tertiary educational institutions depends very much on the existence of prior personal contacts. It was accordingly considered that a fully randomised sample was unlikely to be feasible, however desirable it might be as a means of reducing bias. It was therefore decided to go for convenience sampling of 20% of the total (i.e. six

universities), but applying three constraints: the sample should be small enough to allow detailed observations, yet remain manageable, but large enough to allow a degree of generalisability, and varied enough to cover the two main institutional differences involved: geographical and use of tests as a graduation requirement.

As this study was a comparative one, a degree of generalisability could be achieved with three institutions per sector and a balanced design involving 3 public and 3 private universities would permit comparability across sectors. The aim was for this sample to involve two universities from northern Taiwan, two from central Taiwan and two from the south. Three of the universities would require non-English major students to pass a certain level of one English proficiency test as a degree requirement, whereas the other three did not have a similar regulation. The most striking contrast between the two groups of universities was the English graduation requirement. Within institutions sampling becomes more difficult for the researcher to control. I decided to approach language departments, language centres and English divisions which delivered General English lessons to non-English majors, but to ask the departments to select the teachers and students themselves. The sole requirements would be that the teachers were full-time and the classes would comprise a range of English levels.

Considerable effort was expended to meet the above aims. In the event, 55 teachers and 621 students from five universities, one public and four private located in central and southern Taiwan, responded to the questionnaires. Later, four of the teachers in four of the five universities agreed to participate in interviews and classroom observations. Three departmental directors and three advisory committee members within the Ministry of Education also agreed to be interviewed.

4.2.4 Data collection procedures for the main study

The teacher questionnaires were distributed through two channels. I personally delivered the questionnaires to teachers who had agreed in their e-mails to participate in the survey and they

distributed the questionnaires to their colleagues. Their colleagues returned the questionnaires to them. The researcher then collected the questionnaires personally from the teachers. In cases where personal delivery was difficult to implement, the questionnaires were mailed to departmental directors or the teachers concerned, with stamped self-addressed envelopes. Mutual agreement about return procedures was reached via telephone and e-mail before the mailing. These departmental directors and teachers subsequently arranged for the surveys to be carried out in their universities and mailed back to me.

As for the student surveys, student questionnaires were distributed mainly by three methods. First, questionnaires were issued to students in class when I was in their classroom. Second, questionnaires were mailed, together with the teacher questionnaires, to the teachers who participated in the teacher surveys. Oral consent was obtained from the teachers and students. Questionnaires were administered directly by the teachers. The teachers then sent the questionnaires back using self-addressed return envelopes. A small proportion of the questionnaires was mailed to departmental directors after they had given oral consent. They arranged for the survey to be carried out in their universities and for the questionnaires to be sent back to the researcher using self-addressed return envelopes.

4.2.5 Data analysis procedures for the main study

As the two questionnaires were designed using parallel procedures, the data analysis procedures were the same for both the teacher and the student survey. The survey data was analysed using SPSS. Frequency distributions were first calculated for all the questionnaire items. All percentages were reported as valid percentages with missing data excluded. Figures to illustrate the findings were created in Microsoft Excel and the tables were created in Microsoft Word.

The comparative study was the key to this study. The major aim of the teacher and student questionnaires was to discover and compare differences between responses from the

two groups of teachers and students. The differences in the survey findings were tested for statistical significance by using chi-square tests and independent sample t-tests. A probability of less than 0.05, as commonly used in educational research (Cohen et al., 2007), was taken as statistically significant for both surveys.

4.3 Research design for the pilot study and main study – classroom observation

This section presents the major considerations underlying the research design for the classroom observations in this study. The purpose of the classroom observations was to find out whether the implementation of English proficiency tests as a graduation requirement had any washback effect on classroom teaching in Taiwanese universities of technology. This section first describes the rationale behind the research design and the classroom observation scheme, then discusses the teacher participants and the data collection and analysis procedures.

4.3.1 Rationale for the classroom observation study

“Classroom observation views the classroom as a place where interactions of various kinds take place, affording learners opportunities to acquire the L2” (Ellis, 1994: 565). To reiterate, this study dealt with possible impacts that the implementation of English proficiency tests as a graduation requirement might bring about in the classroom teaching and learning in Taiwanese universities of technology over a period of time. Therefore, observation was an essential element.

There are essentially two different approaches to classroom observation: structured observation and unstructured observation. Highly structured observation involves going into the classroom with a specific purpose and with an observation schedule with pre-determined

categories and is usually linked with the production of quantitative data and the use of statistical analyses (Denscombe, 2007). With the observation schedule, the observer records what participants do, as distinct from what they say they do. Because the observer is not required to make inferences during the data collection process, the schedule effectively eliminates any bias from the observer and appears to produce objective data. Therefore, with structured observation it is possible to achieve high levels of inter-observer reliability, in the sense that two or more observers using the same schedule should record very similar data. Unstructured observation, on the other hand, is less clear on what it is looking for and usually requires the researcher to observe first what is taking place before deciding on its significance for the research study (Cohen et al., 2007). Thus it involves recording detailed field notes and produces qualitative data. It allows observers to gain rich insights into the situation and it is suited to dealing with complex realities.

The weaknesses of the two approaches have been debated (Allwright & Bailey, 1991; Denscombe, 2007). Structured observation records what happens, but not why it happens. It does not deal with the intentions that motivated the behaviour. In addition, unless a researcher is very clear about what exactly to observe and designs a well-tested observation scheme, the subtleties of the situation can easily be ignored. With detailed field notes in an unstructured observation, researchers may be less sure about what exactly the focus of the observation is at the beginning and try to clarify what behaviours to focus on as they gain experience during the observations. The data from an unstructured observation usually relies heavily on the researcher's inferences and detailed field notes in a particular context, which creates problems with respect to the reliability and representativeness of the data.

As the two approaches to classroom observation have their individual advantages and disadvantages, they would be better used complementarily rather than exclusively. The investigation into the washback effect of English proficiency tests on teaching and learning presented a complicated research situation. It was clear from the start that there would be many intervening factors that interacted in teaching and learning as a result of the implementation of English proficiency tests. This seemed to require a combined approach

using both observation approaches, resulting in what might be called a semi-structured observation.

4.3.2 Research focus on classroom teaching

The review of the literature in Chapter 3 and in previous sections in Chapter 4 provided a basis from which to make assumptions about the nature of classroom instruction in relation to the implementation of English proficiency tests in universities of technology. The following research focuses were thus decided on for exploring the washback effect of the implementation of English proficiency tests as a graduation requirement on classroom teaching. Compared with teaching in the context without the new regulation, teachers whose students had to comply with the new regulation would (it was hypothesised):

1. frequently refer to the English requirement.

If washback occurred, the teachers would talk about the content and format of English proficiency tests and frequently remind students of the graduation requirement to reflect the existence and importance of the policy in the classroom.

2. assign more practice opportunities to students.

If washback occurred, there would be a more student-centred classroom in which teachers provided opportunities for students to engage in activities to develop their knowledge of English and their ability to communicate in the language.

3. assign more class time to student activities such as pair work and group discussion which would increase language practice opportunities.

Pair work and group discussion provide the kind of input and opportunities for output that enables rapid L2 acquisition (see Mitchell & Myles, 2004). If washback occurred, the teacher would assign more time to these activities to develop oral competence.

4. talk less.

Conversely, the students would talk more. If there is a tendency for students to contribute more to the classroom interaction, it can be assumed that students have taken a more active part in learning and receive more practice opportunities to develop their oral competence.

5. use more test-oriented materials.

If washback occurred, it would lead teachers to change what they decide to teach. The teachers would employ exam practice workbooks and/or mock exams to prepare students for the tests, or assign test-oriented magazines to students as extra reading.

4.3.3 The observation scheme

As discussed in Chapter 3, a number of observation instruments have been developed based on developments in language teaching (Allwright & Bailey, 1991: 202-223). However, there appear to be no instruments that have been developed specifically for washback studies. Consequently, a classroom observation instrument was designed for the purpose of the present study. The classroom observation scheme adopted in this study initially borrowed from various observation schemes, such as the Communicative Orientation of Language Teaching (COLT) (Spada & Fröhlich, 1995) and the Brown Interaction Analysis System (BIAS) (Brown, 1975). It was also developed from various data sources, such as the preliminary e-mail contacts with departmental directors and teachers, and my personal teaching experiences. The purposes of this initial approach were to become more familiar with the context of the Taiwanese universities of technology and to obtain participants' opinions and experiences, in order to explore aspects of teaching and learning that might be influenced by the graduation regulation. The preliminary contacts and personal experiences enabled me to identify variables related to the study, and a structured classroom observation scheme was designed. Such structured observation led to quantification and comparison of individual teachers teaching with or without the graduation regulation.

Table 4.3 Classroom observation scheme

Time	Participant organization			Activity type	Activity content	Materials used						
	Whole		Group			Individual	Type			Purpose		
	T to S/C	S to S/C					W	A	V	P	S	N

T to S/C = teacher to students or class as a whole

S to S/C = students to students or class as a whole

W = written

A = audio

V = visual

P = pedagogical

S = semi-pedagogical

N = non-pedagogical

The general categories of the observation scheme were based on the category definitions in the COLT, as it focuses on what teachers and students actually do in the classroom and how they interact (Allen, Fröhlich, & Spada, 1984: 232). It consisted of two parts. Part A described classroom events at the level of activity. Part B focused on the verbal exchanges between teachers and students or among students themselves as they occurred within each activity (Spada & Fröhlich, 1995). Part A of the COLT was employed in this study, as a classroom analysis at the level of activity matched the nature of the research questions to be answered. Part B was not employed, since the focus of this study was not on the language used. The categories based on Part A of the COLT were designed to (a) capture significant features of classroom events in General English classrooms in Taiwanese universities of technology, and (b) provide a means of recording classroom interaction. As shown in Table 4.3, the observation scheme for the study consisted of five major categories: time, participant organisation, activity type, content, and material used. They were all coded in the classes; I simply ticked under the category of participant organisation and materials used during the observations, but made detailed field notes under the category of time, activity type and

activity content.

1. *Time*: How is time segmented within the lesson as a percentage of class time? This category related to instructional behaviours in the classroom. The unit of analysis chosen was a ‘segment’. A segment is defined by Mitchell, Parkinson, and Johnston (1981: 12-14) as “a stretch of classroom discourse having a particular topic and involving participants (both the teacher and students) in carrying out an activity or task through interaction”. A change of topic/activity type or a mode of interaction indicates a completion or the start of a new segment (Gibbons, 2006: 95).

The segment was selected as the basic unit of analysis because it has distinctive features, both linguistic and pedagogic, and therefore can be readily divided into categories as a percentage of class time. Segment boundaries were identified on the basis of ‘focusing moves’ and ‘framing moves’ (Gibbons, 2006; Sinclair & Coulthard, 1975), which are indicators of the completion of one stage of a lesson and the beginning of another. Therefore, the first step in analysing any lessons observed was to divide the lesson into segments.

2. *Participant organisation*: Who is doing the talking during the segments of the lesson as a percentage of class time? Participant organisation covers three basic patterns of organisation for classroom interactions (Allen et al., 1984: 235). The three patterns are:

- whole class – involving teacher to students, or student to students
- pair work or group work
- individual work

These categories describe how a lesson is carried out in terms of the participants in the classroom interaction. The categories reflect different theoretical approaches to teaching. For example, Allen et al. (1984) and Gibbons (2006) considered group work as an important factor in the development of fluency skills. Allen et al. (1984) claimed that:

In classes dominated by the teacher, students spend most of their time responding to questions and rarely initiate speech. Moreover, student talk in a teacher-centred classroom is frequently limited to the production of isolated sentences, which are assessed for their grammatical accuracy rather than for their communicative

appropriateness or value. Because the emphasis in group interactions is more likely to be on the expression of meaning, and less likely to be on the linguistic accuracy of utterances, classes which can be shown to provide more group activities may affect the L2 development of learners in ways which are different from those that represent a teacher-centred lock-step approach to instruction. (p. 236)

Highly controlled, teacher-centred approaches are thought to impose restrictions on the growth of students' productive activity. Participant organisation is one of the rationales behind the imposition of English proficiency tests in order to encourage more practice opportunities for students. Therefore, it is necessary to observe the participant organisation of classroom interaction patterns in this study. The findings enable a comparative investigation of the interaction patterns in classes, to see if there are any differences between different groups.

3. *Activity type*: What teaching and learning activities are realised through various activities as a percentage of class time? After each lesson had been segmented and interaction patterns of classroom activity analysed, the aim was to look more closely at the types of activity carried out within the segments. 'Activity type' was an open-ended construct with no predetermined descriptor. Each activity was separately described, such as discussing, lecturing, or singing.

4. *Content*: What are the teacher and the students talking, reading and writing about, or what are they listening to? 'Content' refers to the subject matter of the activities. It was studied in relation to the area of management. 'Management' refers to classroom procedure.

5. *Material used*: What types of teaching materials are used and for what purpose?

(1) Types of materials:

- written materials, such as textbooks, worksheets, and mock exam papers
- audio materials, such as songs
- visual materials, such as films

(2) Purposes of materials:

- pedagogical (e.g, main textbooks specifically designed for L2 learning)
- semi-pedagogical (e.g, mock exam papers)

- non-pedagogical (materials originally intended for non-school purposes, such as English songs and films).

Brown's (1975) Interaction Analysis System (BIAS) was not designed specially for analysis of language classrooms, but it is a useful tool for analysing verbal interaction (Harrison, 1996). The instrument examines the classroom climate and the degree of direct and indirect teacher influence (Flanders, 1970). I modified the system, to leave just two main categories, 'teacher talk' and 'student talk'. The four categories on the pro forma are shown in Table 4.4.

The system represents a simple post-hoc analysis of audio-taped recordings. The time-line display sheet was marked every three seconds for the duration of the observation. Once a whole lesson had been coded, percentages could be calculated for each of the categories. The purpose was essentially to investigate how much teacher talk and student talk occurred in the classrooms.

Table 4.4 Modified Brown Interaction Analysis System

Teacher																			
Student																			
Silence																			
Unclassifiable																			

Each box represents a three-second time interval.

4.3.4 Data collection and analysis procedures

The classroom observations for the pilot study were carried out in June 2007. Because many decisions within universities are made by their faculties, permission to collect data was obtained orally at the level of individual teachers. The scheduling of each classroom observation was agreed in advance between the teachers, students, and the researcher. All

university visits were pre-arranged with the teachers involved. Reminder telephone calls were made one day before each university visit.

Due to the difficulties of obtaining permission from teachers to video record lessons, both audio-recording and field notes were employed in its place for the pilot study. For each observation, a digital tape recorder was set up before the lesson in one corner of the classroom. The rationale for fixing the tape recorder in a corner was to minimise any interruptions and disturbance to the lessons, as well as to ensure normal teacher and student interaction. I sat at the back of the classroom and recorded interactions via the observation scheme in the classroom as lessons progressed.

Two teachers, who were members of the target population but not part of the final sample, agreed to participate in the observation at this pilot stage. They were both female teachers and had a master's degree in TESOL. One teacher's students faced the new regulation while the other's students did not. Each teacher was observed for two lessons, 50 minutes each. The purpose of the observations was to pilot the observation scheme and to gain familiarity with recording methods. At this stage, no problems were found in the procedures for using the observation scheme or recording methods. As a result, the modified observation scheme based on Part A of the COLT and the recording methods were employed for the main study without any changes. After a trial analysis of the interaction patterns, the BIAS was also employed for the main study.

4.3.5 Reliability of classroom observations

Reliability for classroom observations is concerned with consistency (Allwright & Bailey, 1991: 46). "One is concerned with whether or not an independent researcher could achieve consistent results if working in the same or similar context. The other aspect is concerned with the consistency of the procedure of data collection, analysis and interpretation" (He, 1996: 108).

To enhance the reliability of the data collection and analysis in the main study, I went back to the audio-recorded lessons and recoded the previously analysed lessons six months later. The purpose of this approach was to make sure that I was consistent with the criteria for analysis. Agreement was reached in about 98% of the classroom activities. In addition, a fellow doctoral colleague was invited to be a second coder to examine a sample of observed lessons – 12% of the total lessons observed in the main study. Agreement was reached in about 95 % of the cases (where a ‘case’ means an activity).

4.4 Research design for the pilot study and main study – interviewing

The general aim of conducting interviews was to explore the breadth and range of views represented by the participants on the topic of the complexity of washback phenomena in relation to the implementation of English proficiency tests in universities of technology. At the pilot stage, interviews were also used for the collection of straightforward factual information. Oral consent was obtained from all participants prior to interviews. Face-to-face interviews were then conducted with two teachers and one departmental director, and in the pilot study a telephone interview was conducted with one advisory committee member within the Ministry of Education. Those participants were members of the target population but not part of the final sample in the main study. The purpose of interviews with teachers was to explore the teachers’ beliefs: whether teachers believed that their teaching had been influenced by the introduction of the English proficiency tests; and whether they had made changes in their teaching practices as a result of the implementation of the proficiency tests. The interviews also provided an opportunity for the teachers to give their impressions of the lessons, to describe the rationale behind their choices of activities and materials, and to express their opinions regarding the imposition of English proficiency tests as a graduation requirement. A copy of the interview schedule is given in Appendix 3.3.1.

The interviews with advisory committee members and with departmental directors were

mainly designed to answer the first research question: what strategies did the Taiwanese Ministry of Education (MOE) and university departments use to implement The Enhancement of Students' Foreign Language Proficiency Project? which was split into sub-questions about the rationale and strategies used by the MOE and reactions of departmental directors in relation to the imposition of the new policy. The interview schedules are given in Appendices 3.3.2 and 3.3.3.

All of the interview questions were derived from the review of the literature and e-mail contacts at the preliminary information-gathering stage, and there were parallels between questions in the questionnaires and interviews. All of the interviews were semi-structured with prompts whenever necessary and they were conducted in Mandarin, the official language of Taiwan and also the medium of instruction of the educational system, and hence the language in which all participants would most likely feel comfortable communicating. The interviewees were asked to trial the questions by answering them and they were also invited to comment on any problems they had as well as on any issues that occurred to them after the interviews. The purpose was to determine whether the questions were understood as intended. The aim was also to pre-test the questions and to determine what was and was not working. The procedure served as a method of evaluating specific question wordings and gathering information for designing further questions which did not occur to the researcher. All the interviews were audio-recorded and backed up by written field notes in order to trial the data collection procedure and the equipment.

At this pilot stage, interviewees expressed no particular difficulties in answering any of the questions. Therefore, the interview schedules were employed for the main study with just occasional minor corrections of wording. The interview schedules for the main study are given in Appendix 4.3. In the main study, four teachers, three departmental directors and three advisory committee members were interviewed. Oral consent was obtained from the participants prior to the interviews. All of the interviews were conducted in Mandarin and lasted about 20 to 30 minutes. Each participant was interviewed once and the interviews were audio-recorded.

All of the interviews in the main study were transcribed in full in the original language and then translated into English by the researcher. As suggested by Gillham (2005), the transcripts were edited by avoiding repetitions and putting substantive statements in chronological order to make grammatical sense, which facilitated further levels of analysis and provided a relatively tidy and accessible form for interpretation. Morse and Richards (2002) distinguished between three kinds of coding: descriptive coding, topic coding, and analytic coding. The process of analysis began with topic coding. The topics were designated according to the categories previously used in designing the interview schedules. The categories were used as preliminary ways of understanding the data as “at the beginning of a study the researcher is uncertain about what will ultimately be meaningful” (Merriam, 1998: 179). The researcher then looked for patterns across each of the categories, seeking to identify recurrent analytical categories. The transcripts were then grouped and edited again according to the new analytic categories.

For the purpose of examining the reliability of the interview data, the researcher went back to the audio-recorded interviews and recoded the previously analysed interviews six months later. The purpose of this approach was to make sure that I had been consistent with the criteria for analysis. Agreement was reached in about 98% of the contents. In addition, a fellow doctoral colleague was invited as a second coder to examine a sample of the ten interviews – 10% of the total interviews. Agreement was reached in about 95% of the sample. The main study interview data are summarised and reported in Chapters 7 and 8.

4.5 The ethical issues in the study

There is a practical, as well as a moral, point that, as applied researchers examine and experiment with issues that directly affect people’s lives they must respect the interests of subjects and their communities. Unless all parties concerned are recognised and respected, it is likely that research questions may be inappropriately framed, participants may be

uncooperative and findings may have limited usefulness. Sieber (2009) highlighted two major aspects of ethical responsibility towards conducting research: (1) voluntary informed consent, and (2) privacy/confidentiality/anonymity.

Before obtaining consent, research participants should be informed about the research that is to be undertaken, including who the researcher is, why the study is being done, the length of time required and how many sessions are involved. In addition, respondents should not be forced to participate in a study, even though compulsory participation might increase the response rate. Participants also need to know that they can quit at any time without repercussion. De Vaus (2002: 60) suggested that a general explanation of purpose was preferable to a more detailed one, which might discourage participation, distort responses, and undermine the validity of the findings. Accordingly, in this study, the consent procedure began as a conversation – an ongoing, two-way communication process between research participants and the researcher – by e-mail or by telephone. In this way, the information about the purposes of the study was given to the advisory committee members and departmental directors in the pilot and main studies in order to get permission to interview them. Similarly, teachers who participated in both pilot and main studies were informed about a range of matters concerning the purposes of the study, data collection procedures, or any foreseeable risks or discomfort that might arise during the study in order to gain access to the classrooms or to interview them. With respect to the students, it was explained to them orally by their teachers what the study involved and they were given general information concerning the questionnaire survey. The students were then asked whether or not they were prepared to participate further by answering the questionnaire and by being observed in lessons.

Signed consent proves that consent was obtained. Although informed consent plays an important role in gaining permission to collect data, the situation in Taiwan is somewhat different from that in the UK or USA. According to the ESRC Research Ethics Framework (Economic and Social Research Council, 2005), where research is to be conducted outside the UK, researchers should establish whether a local ethical review is required for the host country for local cultural considerations to be taken into account. In this study, although the

researcher asked the participants to sign a written consent form, the participants preferred to give oral consent because of the local norms. It was therefore decided that ethical concerns could be satisfied in all cases by obtaining verbal agreement. Asking potential participants to sign consent forms can sometimes be problematic, as it may make them unnecessarily defensive and suspicious about the study. Indeed, Singer's (1993) survey, which using informed consent to study interviewees' response rates and quality of response, discovered that information about consent had no perceptible effect on either. In other words, he concluded, there was no point in asking participants to sign a consent form unless sensitive or private topics were involved.

The fundamental principle of the UK Data Protection Act (1998) is the protection of the rights of individuals in respect of personal data held about them by a data controller, including academic researchers who will present and publish their materials (Grinyer, 2002). Researchers are responsible for protecting their respondents' identities by keeping them anonymous and keeping the data confidentially. Grbich (1999; see Grinyer, 2002) suggested that respondents should also be told how confidentiality and anonymity will be maintained before conducting data collection. Following the rules of the UK Data Protection Act and in line with the ethical guidelines for educational research proposed by the BERA (British Educational Research Association, 2004), the procedures used to assure confidentiality were specified on the questionnaires (see Appendices 4.1 and 4.2) and described orally by myself to the participants. The teachers and students who completed the questionnaires did not give their names when I collected their responses, so that they could not be identified. Pseudonyms were adopted for advisory committee members, departmental directors and teachers who participated in the interviews and classroom observations and no information from one group was disclosed to the other.

4.6 Summary

The first part of this chapter summarised the theoretical and methodological considerations for this study. It indicated that a mixed-methods strategy was appropriate to this study, since each single method has its individual weaknesses. The second part described the development of the two questionnaires that were employed. The design and validation procedures, as well as the rationale, aim, structure, content and sources of the two surveys, were discussed. Part Two also discussed the modifications to the questionnaires and the sampling for the teacher and student surveys in the main study, which consisted of 55 teachers and 621 students from 5 Taiwanese universities of technology divided into two groups for comparative purposes. Part Three of this chapter discussed the development of the classroom observation scheme and the data collection and analysis procedures. Finally, the development of interview schedules and data collection and analysis procedures were discussed. Both the classroom observation scheme and the interview schedules were kept, with minor modifications, for the main study.

After analysing all types of data, the researcher made a comparison among data from different methods in order to triangulate and complement the findings. If findings were congruent, interpretations could be made based on the consistent results. When the data showed inconsistency, I tried to speculate on the underlying reasons and interpreted the divergent results.

Chapter Five: Teachers' Perceptions of the New Policy

5.1 Introduction

The comparative findings from one stage in Phase II of the study – the results of the Teacher Questionnaire (TQ) – are discussed in this chapter. The questionnaire was issued to two groups of teachers in universities of technology. One group consisted of teachers whose students had to comply with the graduation regulation, classified as Group 1 (G1); and the other group comprised teachers whose students did not face the regulation, classified as Group 2 (G2). The purpose of administering the questionnaire was to explore possible differences in teachers' perceptions of various aspects of teaching in relation to the new policy.

This chapter first reports the findings on the teachers' perceptions of their General English lessons and their opinions on the graduation requirement, and then compares the results for the two groups. Differences in the teachers' perceptions are tested for statistical significance using independent sample t-tests and chi-square tests. A probability of less than 0.05 is taken as statistically significant for both tests.

There were three parts to the Teacher Questionnaire (see Appendix 4.1). Part One consisted of five questions of nominal variables related to demographic information about the teachers who responded to the questionnaires. Part Two consisted of eight questions on a four-point Likert scale of extent or concern, designed to explore teachers' attitudes towards the new policy. Part Three consisted of seven questions related to teachers' reactions to the new policy and their classroom behaviours. The first four questions were nominal variables, inviting the teachers to make a choice under each question. The remaining three questions were designed on a four-point Likert scale of perceived frequency. For clarity and simplicity of reporting, notation such as TQ 2 is used to describe the Teacher Questionnaire, Question 2.

5.2 Demographic information on the teachers

The survey focused on teachers from five universities of technology in Taiwan, roughly 20% of the total technical university population. Teachers who taught General English to non-English majors were included in the survey. The return rate was 93% for teachers (28 out of 30 questionnaires issued) whose students had to comply with the new regulation and 90% for teachers (27 out of 30 questionnaires issued) whose students did not face the regulation.

A chi-square test was used to evaluate the discrepancy (the degree of relativity) between teachers' characteristics in the two sample groups. Five teacher characteristics were included in the study: gender, age, academic qualifications, number of years of teaching, and number of lessons taught per week. The chi-square results showed that there was no significant difference among all categories of the two groups of teachers' characteristics (see Table 5.1), which indicated that the two samples were similar.

Table 5. 1 Characteristics of the respondents (TQ1-5)

Items	Variables	G1 (%) N=28	G2 (%) N=27	Value	Chi-square significance
Gender	Female	89.3	80.8	.775	.379
	Male	10.7	19.2		
Age	31-40	25.9	40.7	2.489	.288
	41-50	48.1	48.1		
	Above 50	25.9	11.1		
Academic qualifications	Masters	67.9	66.7	.009	.925
	PhD/EdD	32.1	33.3		
Years of teaching	4-6	21.4	37.0	1.623	.203
	7 and more	78.6	63.0		
Teaching periods per week	3-6	32.1	18.5	3.649	.302
	7-10	25.0	25.9		
	11-14	32.1	25.9		
	Other	10.7	29.6		

G1 = teachers whose students had to comply with the regulation

G2 = teachers whose students did not face the regulation

* = significant at $p < 0.05$

Table 5.1 shows that two thirds of the teacher participants were female and one third were male. Ages ranged between 31 and 50 years old and most of the teachers had taught at college level for more than 7 years. The majority of the teachers were teaching General English to non-English majors for between 7 and 14 periods per week. The sampled teachers were qualified academically (around 67% held a master's degree and 32% held a doctorate). It should perhaps be pointed out that there are teachers of English in Taiwanese universities of technology who have a master's or a doctoral degree, and are not subject trained in English, but who are still permitted to teach English. By surfing websites of the five universities along with several others, I was able to ascertain that in most of the cases the non-subject trained teachers' degrees were obtained from overseas universities. In one of the universities studied in Phase III, there were seven teachers of English, none of whom were subject trained in English, including the departmental director, who had a doctorate in science education.

As the majority of the categories in Part Two and Part Three of the Teacher Questionnaire were designed with a four-point Likert Scale of extent, concern, or frequency in order to compare the difference in teachers' attitudes, independent sample t-tests were carried out. Certain questions, such as TQ 13 in Part Two and TQ 14-17 in Part Three, were designed as multiple-choice questions, depending on the nature of the variables. In this case, chi-square tests were performed to test the significance of differences among these variables. The results are reported and discussed in the following order:

- teachers' reactions to and perceptions of the new policy;
- washback on teaching materials;
- washback on teachers' classroom behaviours;
- washback on assessment and evaluation;
- washback on teachers' attitudes towards aspects of learning.

5.3 Teachers' reactions to and perceptions of the new policy

In this section, one question referred to the teachers' reactions (TQ14) to the new policy and was designed using a multiple-choice format with a space for further explanation. The other two categories referred to the teachers' perceptions of the reasons behind the implementation of the new policy (TQ8) and the possible teaching methods teachers had adopted as a result of the new policy (TQ10). These two categories employed a four-point Likert scale.

5.3.1 Teachers' reactions to the new policy

Group 1 and Group 2 teachers were asked to describe their reactions to the new policy – passing English proficiency tests as a graduation requirement for non-English majors. The results are presented in Table 5.2.

Table 5. 2 Teachers' reactions to the new policy (TQ14)

Teachers' reactions	Group 1 (%) N = 27	Group 2 (%) N = 26	Value	Chi-square significance
Sceptical about the policy	29.6	30.8	2.212	.331
Neutral about the policy	18.5	34.6		
Welcome the policy	51.9	34.6		

Group 1 = teachers whose students had to comply with the regulation

Group 2 = teachers whose students did not face the regulation

* = significant at $p < 0.05$

Even though chi-square test results showed that there was no significant difference between Group 1 and Group 2, Table 5.2 shows that a noticeably higher proportion of Group 1 teachers (51.9 per cent of 27) welcomed the policy, compared with Group 2 teachers (34.6

per cent of 26). The results were a possible indication that Group 1 teachers generally held relatively positive and supportive attitudes towards the new policy. Some teachers believed that the new policy could stimulate students' motivation to learn, push them to study harder, and prepare them well for a future career. Therefore, the policy was welcomed.

Although the survey showed a generally positive attitude towards the new policy, this does not necessarily mean that Group 1 teachers had changed or would change their behaviour. As discussed in Chapter 3, it is only when the teachers come across problems and difficulties in actual teaching that they confront the pressure for change. A reluctant attitude to change might simply be due to the practical aspects of teaching. It is a general phenomenon that teachers are seldom happy to abandon what they have been doing and embrace completely some new philosophy, methodology, or new approaches, even if they do have a positive attitude towards the new policy (see Chapter 3). It is only to be expected that over time teachers would modify what they had been doing to prepare their students for the new regulation as it is so important to both students and teachers alike. Further discussion of this issue will be presented in Phase III of the study.

5.3.2 Teachers' perceptions of the reasons behind imposing the new policy

TQ7-9 were designed to explore the extent to which teachers perceived the reasons behind the imposition of the new policy. When Group 1 teachers were asked what they saw as the major aims of the universities in establishing the regulation, the result was as shown in Table 5.3. This gives the mean scores on a scale of concern, where 1 = not a major concern at all and 4 = main concern.

It can be seen that Group 1 teachers on average regarded 'to meet the government regulation' as the most important reason for creating the new regulation, followed by 'to prepare students for their future career' and 'to cope with the present decline in English standards', suggesting that Group 1 teachers were aware of the underlying aims behind the

new policy and that this understanding might encourage them in the direction of teaching to English proficiency tests.

Table 5.3 Group 1 teachers' perceptions of the reasons behind imposing the new policy (TQ8)

	Mean	SD
● to meet the government regulation	3.12	.881
● to prepare students for their future career	3.07	.766
● to cope with the present decline in English standards	3.04	.676
● to direct students' learning	2.78	.751
● to motivate students to use integrated skills	2.54	.647
● to enable students to communicate more with foreigners	2.19	.849
● to assess teachers through students' performance	2.13	.680
● to encourage better textbooks	2.04	.841

N = 27

TQ9 was an open-ended question for Group 2 teachers to provide reasons why their university had decided not to impose the regulation. Eleven teachers provided answers. Six of the eleven thought that the university was not imposing the regulation because the students' English standard was too low for them to pass English proficiency tests. The other five teachers thought that imposing the regulation would decrease students' willingness to apply for the university.

It can be seen that students' English standard was perceived as one of the major concerns directing university authorities' attention to establishing the graduation regulation. In Group 1, the English graduation requirement was considered as a strategy for coping with the present decline in students' English standards. In Group 2, however, students' low English standard was considered as a reason which impeded the development of the initiative. The results appeared to reflect mixed feelings and the complexity of the problems involved in the adoption of the policy.

5.3.3 Alterations to teaching methods due to the new policy

There were twelve items in TQ10, regarding the changes the teachers had made to their teaching possibly in line with the new policy. They are listed in Table 5.4 in descending order according to the mean scores on a four-point Likert scale of extent in Group 1, where 1 = not at all and 4 = to a very large extent.

Table 5. 4 Alterations to teaching methods (TQ10)

Variables	Group	Cases	Mean	SD	T-Value	df	2-Tail Prob.
Encourage more student participation in class	1	26	3.23	.652	.828	50	.412
	2	26	3.08	.688			
Put more emphasis on reading	1	27	3.19	.681	.981	51	.331
	2	26	3.00	.693			
Put more emphasis on listening	1	26	3.04	.720	1.826	50	.074
	2	26	2.65	.797			
Employ more real life language tasks	1	26	2.96	.774	1.008	50	.318
	2	26	2.73	.874			
Use a more communicative approach in teaching	1	26	2.81	.694	-.755	50	.454
	2	26	2.96	.774			
Put more emphasis on the integration of skills	1	26	2.77	.765	-.172	50	.864
	2	26	2.81	.849			
Put more emphasis on speaking	1	25	2.72	.792	1.680	49	.099
	2	26	2.35	.797			
Teach according to English proficiency test format	1	27	2.63	.792	1.326	51	.191
	2	26	2.31	.970			
Employ new teaching methods	1	25	2.60	.645	-.864	48	.392
	2	25	2.76	.663			
Do more mock exam papers*	1	24	2.54	.721	2.194	48	.033*
	2	26	2.12	.653			
Put more emphasis on group discussion	1	26	2.31	.736	-.817	49	.418
	2	25	2.48	.770			
Put more emphasis on writing	1	26	2.31	.736	.177	50	.860
	2	26	2.27	.827			

Group 1 = teachers whose students had to comply with the regulation

Group 2 = teachers whose students did not face the regulation

* = significant at $p < 0.05$

In one item in Table 5.4 (marked by an asterisk) there was seen to be a significant difference between the groups. The result seems to indicate a washback effect of the new policy on teachers' teaching activities. Group 1 teachers were aware of the new policy and had organised more mock exams in their teaching.

The remaining items did not show significant differences between the groups. However, it was still of considerable interest to observe the order of changes that Group 1 teachers claimed to have made in terms of the new policy. Table 5.4 shows that 'more student participation in class', 'more emphasis on reading', and 'more emphasis on listening' were high on the list, with mean scores above three on a four-point Likert scale of extent in Group 1. However, when it came to a 'more communicative approach in teaching', 'more emphasis on the integration of skills', 'new teaching methods', and 'more group discussion', the means were lower than two. This might suggest a discrepancy between teachers' attitudes towards adopting new classroom activities and actually employing new teaching methods.

Analysis of teachers' reactions to and perceptions of the new policy showed that Group 1 teachers appeared to have a positive reaction to the new policy. Moreover, Group 1 teachers' perceptions of the major aims behind the new policy in relation to the changes they had made in teaching appeared to reflect their awareness of the underlying theories of teaching and learning behind the new policy, English proficiency tests, and the intended washback. However, whether teachers actually altered their teaching to reflect the theories is still a question to be answered.

It can be seen that washback on specific aspects of teaching in the context of the new policy was complicated and showed a mixed picture. On the one hand, the Group 1 teachers perceived encouraging 'more student participation in class' as the most important activity, but on the other, they perceived adopting 'more group discussion' as a less important activity. The actual impact of the new policy on the teachers' attitudes cannot be directly determined from the survey results due to the nature of the survey study and the relatively short period of the survey. This will be discussed further in Phase III.

5.4 Washback on teaching materials

There were two questions covering teaching materials (TQ16 and 20). Items in TQ16 were designed on a multiple-choice scale. TQ 20 was designed on a four-point Likert scale of frequency, where 1 = seldom or never and 4 = every class. The purpose of the questions was to explore possible relationships between teaching materials and the implementation of English proficiency tests.

5.4.1 Textbook arrangements related to teaching materials

TQ16 was related to teaching materials in the context of implementing English proficiency tests as a graduation requirement and it was presented in a multiple-choice format as shown in Table 5.5. The results did not differ significantly between the two groups according to a chi-square test.

Table 5.5 Textbook arrangements in Taiwanese universities of technology (TQ16)

Items	Variables	G1 (%) N=28	G2 (%) N=27	Value	Chi-square	significance
Have you chosen textbooks for General English classes because of English proficiency tests?	Yes	46.4	44.4	.022	.883	
	No	53.6	55.6			

G1 = teachers whose students had to comply with the regulation

G2 = teachers whose students did not face the regulation

* = significant at $p < 0.05$

It can be seen that, for the majority of teachers, textbook choice did not appear to be affected by the new policy. As one of the Group 1 teachers explained later in the interview, the purpose of General English classes was to develop students' English skills and prepare them

for the world of work. Therefore, English teaching in universities of technology should not be test-oriented.

5.4.2 Teaching and learning resources

TQ20 was aimed at finding out whether the teachers believed that the implementation of English proficiency tests would involve extra teaching and learning resources. Teachers were asked to state how often they used a list of eight resources in their teaching. The list in Table 5.6 is arranged in descending order according to the mean scores for Group 1.

Table 5.6 Teaching and learning resources (TQ20)

Variables	Group	Cases	Mean	SD	T-Value	df	2-Tail Prob.																																																																																
Textbooks	1	27	3.41	.572	-.671	52	.505																																																																																
	2	27	3.52	.643				Language laboratory	1	26	2.12	.909	-.760	51	.451	2	27	2.30	.823	English proficiency test-oriented materials	1	28	1.93	.766	.339	53	.736	2	27	1.85	.907	Pictures and/or cards	1	26	1.88	.816	-.341	51	.734	2	27	1.96	.854	Magazines	1	27	1.74	.594	-1.312	52	.195	2	27	1.96	.649	Radio	1	27	1.67	.961	1.575	47	.122	2	26	1.31	.679	Newspapers	1	26	1.50	.583	-.110	51	.913	2	27	1.52	.643	Television	1	27	1.48	.753	-.284	51	.778
Language laboratory	1	26	2.12	.909	-.760	51	.451																																																																																
	2	27	2.30	.823				English proficiency test-oriented materials	1	28	1.93	.766	.339	53	.736	2	27	1.85	.907	Pictures and/or cards	1	26	1.88	.816	-.341	51	.734	2	27	1.96	.854	Magazines	1	27	1.74	.594	-1.312	52	.195	2	27	1.96	.649	Radio	1	27	1.67	.961	1.575	47	.122	2	26	1.31	.679	Newspapers	1	26	1.50	.583	-.110	51	.913	2	27	1.52	.643	Television	1	27	1.48	.753	-.284	51	.778	2	26	1.54	.706								
English proficiency test-oriented materials	1	28	1.93	.766	.339	53	.736																																																																																
	2	27	1.85	.907				Pictures and/or cards	1	26	1.88	.816	-.341	51	.734	2	27	1.96	.854	Magazines	1	27	1.74	.594	-1.312	52	.195	2	27	1.96	.649	Radio	1	27	1.67	.961	1.575	47	.122	2	26	1.31	.679	Newspapers	1	26	1.50	.583	-.110	51	.913	2	27	1.52	.643	Television	1	27	1.48	.753	-.284	51	.778	2	26	1.54	.706																				
Pictures and/or cards	1	26	1.88	.816	-.341	51	.734																																																																																
	2	27	1.96	.854				Magazines	1	27	1.74	.594	-1.312	52	.195	2	27	1.96	.649	Radio	1	27	1.67	.961	1.575	47	.122	2	26	1.31	.679	Newspapers	1	26	1.50	.583	-.110	51	.913	2	27	1.52	.643	Television	1	27	1.48	.753	-.284	51	.778	2	26	1.54	.706																																
Magazines	1	27	1.74	.594	-1.312	52	.195																																																																																
	2	27	1.96	.649				Radio	1	27	1.67	.961	1.575	47	.122	2	26	1.31	.679	Newspapers	1	26	1.50	.583	-.110	51	.913	2	27	1.52	.643	Television	1	27	1.48	.753	-.284	51	.778	2	26	1.54	.706																																												
Radio	1	27	1.67	.961	1.575	47	.122																																																																																
	2	26	1.31	.679				Newspapers	1	26	1.50	.583	-.110	51	.913	2	27	1.52	.643	Television	1	27	1.48	.753	-.284	51	.778	2	26	1.54	.706																																																								
Newspapers	1	26	1.50	.583	-.110	51	.913																																																																																
	2	27	1.52	.643				Television	1	27	1.48	.753	-.284	51	.778	2	26	1.54	.706																																																																				
Television	1	27	1.48	.753	-.284	51	.778																																																																																
	2	26	1.54	.706																																																																																			

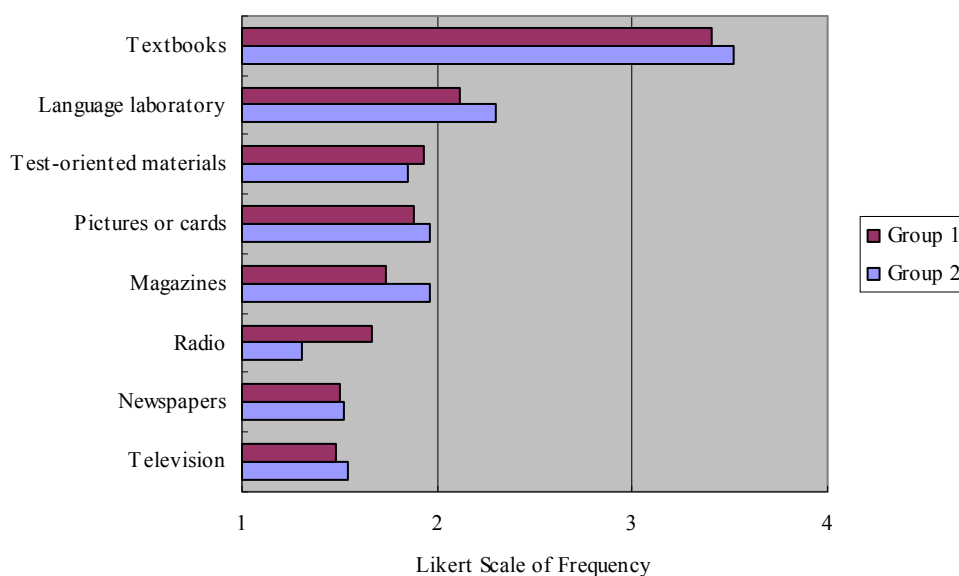
Group 1 = teachers whose students had to comply with the regulation

Group 2 = teachers whose students did not face the regulation

* = significant at $p < 0.05$

There was no significant difference between the two groups for any of the teaching and learning resources. Figure 5.1 shows that textbooks were the most often used resources, followed by language laboratories. It was quite evident throughout the study that textbooks produced by publishers were the most direct form of support on which teachers relied.

Figure 5.1 Teaching and learning resources (TQ20)



In summary, the results in the area of washback on the choice of teaching materials (TQ16 and 20) showed that there was no clear evidence of washback. The overall pattern was not seen to differ in the context of implementing English proficiency tests as a graduation requirement. The assumption by the researcher (see Chapter 4, Section 4.3.2) that the new policy would require more teaching and learning resources was not proven. The teachers continued to use General English textbooks produced by publishers in their teaching.

5.5 Washback on classroom teaching behaviour

Questions related to aspects of daily teaching were investigated in this section. The purpose was to find out whether the implementation of English proficiency tests would influence the teachers' decisions about, and perceptions of, their teaching activities. There were four questions. The first two questions (TQ15 and 17) were designed with a multiple-choice format. A chi-square test was carried out to test the significance of differences between the two groups. The next two questions (TQ 18 and 19) were designed on a four-point Likert scale of frequency, where 1 = seldom or never and 4 = every class. An independent sample t-test was used for statistical analysis of differences between the group means. Results of this section are reported in two parts: (a) teaching planning and medium of instruction in General English lessons (TQ 15 and 17), and (b) teacher talk and teaching activities in General English lessons (TQ 18 and 19).

5.5.1 Teaching planning and medium of instruction

There were two questions (TQ15 and 17) in this section. As they were designed on a multiple-choice scale, a chi-square test was used to clarify the general teaching situation as shown in Table 5.7. Neither of them produced answers that differed significantly between the groups.

Table 5.7 shows that the majority of teachers arranged their teaching according to the arrangement of content in a General English textbook, which was consistent with the results of TQ20 regarding teaching and learning resources, namely that textbooks were the most used resource. As shown in Figure 5.2, the majority of teachers used half English and half Chinese as their medium of instruction. One possible reason for this might be teachers' concerns and worries over their current students' English level. The key aim was to get the meaning of the language across to students; this point was frequently mentioned by many teachers and

departmental directors during university visits in all phases of the study.

Table 5. 7 Chi-square tests of differences in aspects of daily teaching (TQ15 and 17)

Items	Variables	G1 (%)	G2 (%)	Value	Chi-square significance
How do teachers arrange their teaching at their universities?	1. Followed a General English textbook's arrangement	37.0	50.0	1.124	.570
	2. Followed English proficiency test-oriented materials	33.4	30.8		
	3. Divided into separate skills such as reading or listening	29.6	19.2		
What is the medium of instruction?	1. English only	14.8	18.5		.574 (Fisher's Exact Test [^])
	2. Half English, half Chinese	59.3	44.4		
	3. Mainly Chinese	25.9	37.0		

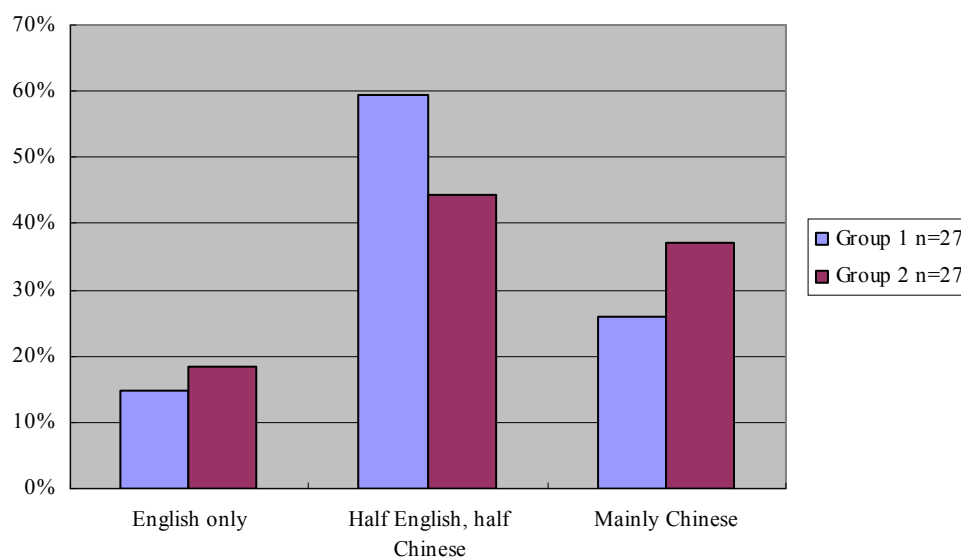
G1 = teachers whose students had to comply with the regulation

G2 = teachers whose students did not face the regulation

[^]Due to the small sample size, Fisher's Exact Test was used because it is more accurate than the chi-square test.

* = significant at $p < 0.05$

Figure 5. 2 Reported medium of instruction in Taiwanese universities of technology (TQ17)



5.5.2 Teacher talk and teaching activities in General English lessons

TQ18 was related to teachers' talk. Teachers were asked how often they carried out the following activities in class: (a) talk to the whole class, (b) talk to groups of students, or (c) talk to individual students. The general pattern of teacher talk in General English lessons for the two groups is shown in Table 5.8 and Figure 5.3. Of the three items, 'talk to groups of students' and 'talk to individual students' showed significant differences between the groups.

Table 5.8 Teacher talk in General English lessons (TQ18)

Variables	Group	Cases	Mean	SD	T-Value	df	2-Tail Prob.
Talk to the whole class	1	28	3.46	.508	.390	53	.698
	2	27	3.41	.572			
Talk to groups of students*	1	24	2.67	.482	2.095	40.749	.042*
	2	26	2.27	.827			
Talk to individual students*	1	26	2.58	.578	2.837	50	.007*
	2	26	2.08	.688			

Group 1 = teachers whose students had to comply with the regulation

Group 2 = teachers whose students did not face the regulation

* = significant at $p < 0.05$

There was a higher frequency of teacher talk to groups and individual students in Group 1. It seemed that Group 1 teachers provided more practice opportunities for the students than Group 2 teachers did. However, in the great majority of lessons *both* groups of teachers talked to the whole class, and much less to groups or to individuals. The results suggest that Group 1 teachers still dominated and controlled the classroom talk for most of the lesson time, though the new policy did appear to encourage teachers to be more learner-centred. All in all, there seemed to be an impact of the new policy on the area of teacher talk, but it was not as great as had been expected (see Chapters 2 and 4).

Figure 5.3 Teacher talk in General English lessons (TQ18)

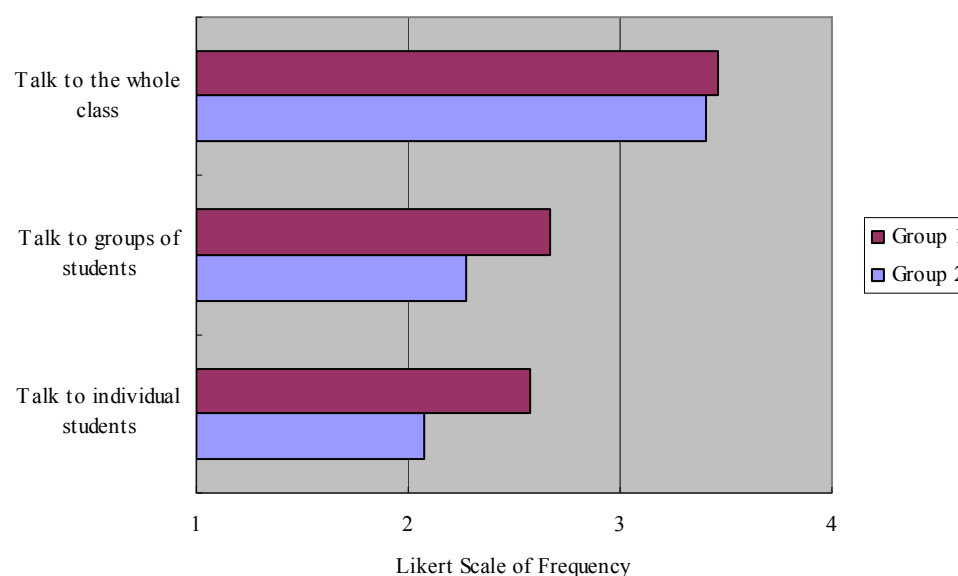


Table 5.9 Teachers' teaching activities (TQ19)

Variables	Group	Cases	Mean	SD	T-Value	df	2-Tail Prob.
Explain specific language items	1	28	3.18	.670	-.417	53	.678
	2	27	3.26	.764			
Explain the meaning of the text	1	27	3.11	.698	-1.375	52	.175
	2	27	3.37	.688			
Go through textbook exercises	1	27	2.81	.834	-.755	46.697	.454
	2	27	2.96	.587			
Organise pair work or group discussion	1	26	2.50	.812	-.215	51	.831
	2	27	2.56	1.050			
Organise integrated language skill tasks	1	26	2.50	.860	.407	51	.686
	2	27	2.41	.797			
Go through mock exam papers	1	26	2.23	.863	1.277	51	.207
	2	27	1.93	.874			
Talk about English proficiency tests	1	26	1.96	.528	-.419	51	.677
	2	27	2.04	.759			

Group 1 = teachers whose students had to comply with the regulation

Group 2 = teachers whose students did not face the regulation

* = significant at $p < 0.05$

There were seven activities listed under TQ19 to explore how often teachers carried out the activities in class. The items were designed with a four-point Likert scale of frequency, where 1 = seldom or never and 4 = every class, and they are listed in Table 5.9 in descending order, according to the mean scores for Group 1. None of the activities were found to differ significantly between Group 1 and Group 2 teachers.

It can be seen from Table 5.9 that ‘explain specific language items’ such as words or sentences was carried out most often by Group 1 teachers. This was closely followed by ‘explain the meaning of the text’ and ‘go through textbook exercises’. All these activities were related to explaining, which might suggest why Group 1 teachers spent so much of their time talking to the whole class (see Table 5.8). This finding was also supported by Phase III of the study, when classroom observations were carried out (see Chapter 7). The teachers spent a great deal of time explaining language items and the meanings of the text to their students in class. ‘Organise pair work or group discussions’ and ‘organise integrated language skill tasks’ were less common, indicating that Group 1 teachers still adhered to more traditional methods. Furthermore, ‘talk about English proficiency tests’ was ranked the lowest among the items. It was interesting that although ‘go through mock exam papers’ was done more often by Group 1 teachers than Group 2 teachers, Group 1 teachers said they talked about English proficiency tests *less* often than the Group 2 teachers.

To summarise, there was not much of an indication of washback on aspects of teaching at the micro level, which was perhaps not surprising considering the short period of the survey. The two groups showed, at a general level, similar approaches to teaching in terms of teacher talk, the nature of teaching, and delivery modes. However, some differences were observed. There was a tendency in Group 1 for teachers to pay more attention to talking to groups and individuals, which indicated more practice opportunities for students and might be directly related to the new policy of increasing oral components.

5.6 Washback on assessment and evaluation

5.6.1 Factors that influence teaching

In TQ12, teachers' perceptions of a list of eight factors that influenced their teaching were explored. The results are listed in Table 5.10 in descending order according to the mean scores for Group 1.

Table 5. 10 Factors that influence teaching (TQ12)

Variables	Group	Cases	Mean	SD	T-Value	df	2-Tail Prob.
Teaching experience and beliefs	1	26	3.50	.583	1.344	51	.185
	2	27	3.26	.712			
Past experience as a language learner	1	25	3.32	.557	.585	50	.561
	2	27	3.22	.641			
Learners' expectations*	1	26	3.23	.514	2.735	43.365	.009*
	2	26	2.73	.778			
Professional training received	1	24	3.08	.717	.945	48	.349
	2	26	2.88	.766			
The need to obtain satisfaction in teaching*	1	24	3.08	.584	2.261	46.453	.028*
	2	27	2.63	.839			
English proficiency tests	1	25	2.84	.624	1.743	47.641	.088
	2	27	2.48	.849			
Academic seminars or workshops taken	1	25	2.76	.523	-.109	50	.914
	2	27	2.78	.641			
Textbooks	1	24	2.75	.676	-1.097	49	.278
	2	27	2.96	.706			

Group 1 = teachers whose students had to comply with the regulation

Group 2 = teachers whose students did not face the regulation

* = significant at $p < 0.05$

The t-test results showed that Group 1 and Group 2 teachers perceived two factors –

'learners' expectations' and 'the need to obtain satisfaction in teaching' – differently (marked by asterisks), both factors influencing Group 1 more than Group 2. The results might suggest that the English graduation regulation had the effect of making teachers pay more attention to learners' expectations and stressing teachers' need to obtain satisfaction in teaching. The remaining items perceived by the teachers did not show differences. 'Teaching experience and beliefs' were perceived as the most influential factor, and 'textbook' was perceived as the least influential factor. In TQ15 and 20, textbooks were reported by the teachers as providing the most direct teaching support, but teachers regarded them as relatively less important than the other factors. The teachers might rely on textbooks directly and extensively, but they did not perceive textbooks as an important factor in teaching.

5.6.2 The use of mock exams

TQ13 was included to explore the number of mock exams that teachers gave to their students every semester. As shown in Table 5.11, there was no significant difference between the two groups. The majority of teachers assigned their students 1-3 mock exams every semester. It was interesting that more Group 2 teachers (92.6 per cent of 27) assigned 1-3 mock exams than did Group 1 teachers (75 per cent of 28), even though Group 2 students did not have to comply with the new regulation. More Group 1 teachers (25 per cent of 28), however, assigned 4 or more exams than did Group 2 teachers (7.4 per cent of 27). The results suggested that there was a direct washback effect on the use of mock exams in both groups. As one of the Group 1 teachers explained, mock exam practice was employed as a result of the new policy and teachers perceived preparing students for English proficiency tests as one of their responsibilities. On the other hand, one of the Group 2 teachers explained that, although taking English proficiency tests was not compulsory in the university, teachers liked to provide more learning opportunities to students and hoped students' English standard could be improved before graduation, so they, too, assigned students mock exam practice.

Table 5. 11 The use of mock exams (TQ13)

Items	Variables	G1 (%) N=28	G2 (%) N=27	Fisher's Exact Test [^]
How many English proficiency mock exams do non-English majors take every semester?	1-3 4 and more	75.0 25.0	92.6 7.4	.143

G1 = teachers whose students had to comply with the regulation

G2 = teachers whose students did not face the regulation

[^]Due to the small sample size, Fisher's Exact Test was used because it is more accurate than the chi-square test.

* = significant at $p < 0.05$

To summarise, the findings regarding washback from the new policy on assessment and evaluation (TQ12 and 13) showed that there was in all probability a washback effect. The attitudes that differed between Group 1 and Group 2 teachers were those influencing factors such as learners' expectations and the need to obtain satisfaction in teaching. This might suggest that the new policy was having the effect of making teachers pay more attention to learners' expectations and stressing teachers' need to obtain satisfaction in teaching. Having said that, increased mock exam practice was employed in *both* groups. It seemed that when there was an intention to use tests as part of teaching and learning, there was very likely a re-focusing of the classroom activities to reflect the tests in the classroom. In this study, Group 1 teachers viewed mock exam practice as a necessary activity to prepare students for English proficiency tests and Group 2 teachers viewed mock exam practice as a learning resource, suggesting an indirect washback effect of implementing English proficiency tests.

5.7 Washback on teachers' attitudes towards aspects of learning

Two questions in Part Two of the teacher questionnaire (TQ6 and 11) explored teachers' attitudes regarding the aims of learning English in Taiwanese universities of technology and the learning strategies the teachers had recommended to their students. There were 12 items in

all in these two questions. All items were designed using a four-point Likert scale. Both questions dealt with teachers' attitudes towards how aspects of learning might have been influenced in the light of the implementation of English proficiency tests.

5.7.1 Learning aims

Teachers were asked about their perceptions of students' major reasons for learning English in Taiwanese universities of technology. None of the four items listed in Table 5.12 showed a significant difference between the two groups.

Table 5.12 Teachers' perceptions of students' major reasons for learning English (TQ6)

Variables	Group	Cases	Mean	SD	T-Value	df	2-Tail Prob.
To satisfy university requirements	1	28	3.68	.476	1.313	42.019	.196
	2	27	3.44	.801			
To pass examinations	1	27	3.22	.801	-1.882	52	.065
	2	27	3.59	.636			
To obtain a 'good' job	1	26	2.54	.761	-1.000	51	.322
	2	27	2.74	.712			
To pursue further studies	1	24	2.13	.797	-.109	49	.914
	2	27	2.15	.718			

Group 1 = teachers whose students had to comply with the regulation

Group 2 = teachers whose students did not face the regulation

* = significant at $p < 0.05$

Table 5.12 shows that Group 1 teachers' responses to 'satisfying university requirements', which came highest on the list, were followed by 'passing examinations', and Group 1 teachers gave relatively more weight to satisfying the university's requirements than Group 2 teachers as the reason for learning English in Taiwanese universities of technology,

both indicating possible pressure on Group1 teachers' teaching whenever there are new policies and new tests introduced to the education system. The remaining reasons were rated relatively lower by Group 1 teachers than Group 2 teachers, though one of them was relatively important to both groups. The mean scores for 'passing examinations' were above three on the Likert scale, indicating that both groups regarded this as an important aim of learning English in Taiwan.

5.7.2 Learning strategies

Eight strategies were listed to explore teachers' attitudes towards the learning strategies they had recommended to their students in the context of implementing English proficiency tests as a graduation requirement. They were listed in Table 5.13 and mean scores were shown in descending order for Group 1.

Only one strategy showed significant differences between the groups (marked by an asterisk). Table 5.13 shows more Group 1 teachers had recommended that their students should put emphasis on listening than had Group 2 teachers. This finding might suggest a change in Group 1 teachers' attitudes towards how learners should study in the context of the new regulation.

Even though there was no significant difference in the remaining strategies, the priority that teachers gave to strategies they regarded as important was revealed from the results. Both groups of teachers tended to recommend those strategies that were directly related to the English proficiency tests and to language learning.

Table 5. 13 Learning strategies that teachers had recommended to students (TQ11)

Variables	Group	Cases	Mean	SD	T-Value	df	2-Tail Prob.
To put more emphasis on reading	1	26	3.19	.634	1.019	51	.313
	2	27	3.00	.734			
To put more emphasis on listening*	1	27	3.11	.641	2.726	52	.009*
	2	27	2.67	.555			
To expose themselves to a range of English media	1	27	3.07	.675	.591	52	.557
	2	27	2.96	.706			
To take notes	1	26	2.96	.774	.183	50	.856
	2	26	2.92	.744			
To use English more in their daily life	1	27	2.81	.834	.166	52	.868
	2	27	2.78	.801			
To put more emphasis on speaking	1	26	2.77	.652	.346	51	.731
	2	27	2.70	.724			
To express their opinions in English in class	1	27	2.67	.734	-.180	52	.858
	2	27	2.70	.775			
To put more emphasis on writing	1	25	2.32	.627	-.253	50	.801
	2	27	2.37	.792			

Group 1 = teachers whose students had to comply with the regulation

Group 2 = teachers whose students did not face the regulation

* = significant at $p < 0.05$

5.8 Summary

It can be seen from the findings of the teacher questionnaire that identifying washback effects of the new policy is complicated. The following aspects of the teachers' attitudes in the context of the implementation of English proficiency tests as a university graduation requirement were explored in this chapter:

- teachers' reactions to and perceptions of the new policy;
- washback on teaching materials;
- washback on teachers' classroom behaviours;
- washback on assessment and evaluation;
- washback on teachers' attitudes towards aspects of learning.

Firstly, exploring teachers' reactions to and perceptions of the new policy, it seems that Group 1 teachers had a generally positive reaction to the new policy. There was a significantly higher proportion of Group 1 teachers (51.9 per cent of 27) who 'welcomed the policy' compared with Group 2 teachers (34.6 per cent of 26), and a correspondingly lower proportion of Group 1 teachers (29.6 per cent of 27) who were 'sceptical about the policy' compared with Group 2 teachers (30.8 per cent of 26). In addition, teachers' perceptions of the reasons for implementing the new policy were consistent with the underlying theories behind it. Group 1 teachers were aware that the main purposes of imposing the graduation requirement were to prepare students for their future career and to counter the decline in English standards.

In terms of teaching, differences in teaching practice were closely related to doing more mock exam papers. This suggested a washback effect in the context of implementing English proficiency tests as a graduation requirement, but at the surface level in accordance with the implementation of the policy. Moreover, there was a discrepancy between teachers' attitudes towards the relative importance of classroom activities and employing new teaching methods. Both groups of teachers gave a higher weighting to the activities of 'encouraging more student participation in class', 'putting more emphasis on reading', and 'putting more emphasis on listening'. However, they considered it less likely that they would employ new teaching methods to increase communicative competence, indicating a degree of reluctance by teachers to make changes in certain aspects of their teaching.

Secondly, the majority of the teachers in the two groups did not employ test-oriented textbooks and their teaching mainly relied on the content and organisation of General English textbooks. The results suggested that there was no washback effect on teaching materials. Although Group 1 teachers tended to provide students with more practice opportunities in terms of teacher talk, the general pattern of the nature of teaching and teaching activities did not differ significantly. In addition, the majority of the teachers in the two groups used half English and half Chinese as the medium of instruction due, according to the teachers, to the low level of their students' language proficiency. These results together indicated that there

was a minimal washback effect on teachers' overall classroom behaviours.

Thirdly, two factors influencing teaching proved to be significantly different; these were learners' expectations and the need to obtain satisfaction in teaching. This might indicate that using English proficiency tests as an agent of change could possibly make Group 1 teachers pay more attention to learners' expectations. In addition, there was a discrepancy that textbooks were reported by the teachers as providing the most direct teaching support, but teachers regarded them as relatively less important than the other factors in teaching.

Finally, when aspects of learning were explored from the teachers' perspectives, it appeared that the Group 1 teachers were more willing to encourage their students to cope with the new policy rather than change their own teaching methodology. Explaining the teachers' reluctance to make changes was complicated, and is explored further through detailed classroom observations and in-depth interviews in Phase III of the study.

Chapter Six: Students' Perceptions of the New Regulation

6.1 Introduction

In this chapter, the findings from the other stage in Phase II of the study – the results of the Student Questionnaire (SQ) – are discussed. As explained in the methodology chapter, the questionnaire was issued to two groups of university of technology students (one group was scheduled to take English proficiency tests as a graduation requirement, whereas the other group did not face the regulation). The student questionnaire explored possible differences in the students' attitudes in relation to their classroom activities within the context of the new regulation.

This chapter first reports the findings on the students' perceptions of their English lessons, and then compares the results for the two groups. For the purpose of in-depth investigation, differences between students in different study years were compared at some points. Differences in the students' perceptions were tested for statistical significance using independent sample t-tests and chi-square tests according to the nature of the variables concerned. A probability of less than 0.05 was, as before, taken as statistically significant.

The Student Questionnaire was divided into two parts (see Appendix 4.2). Part One consisted of seven questions related to demographic information on the students who responded to the questionnaire and various elements of their learning contexts. Six questions used nominal scales and one category employed a fill-in format; students needed to tick the appropriate choice and fill in answers based on their own situations. Part Two consisted of three sections (Sections A, B & C) including 10 questions. Section A employed a four-point Likert scale of frequency; sections B and C had a nominal scale of agreement. Part Two of the questionnaire aimed to explore the students' attitudes towards teaching and learning activities, inside and outside the classroom, in relation to the new regulation. For clarity and simplicity

of reporting, a label such as SQ8 is used here to refer to Student Questionnaire, Question 8.

6.2 Students' characteristics and their learning contexts

The survey focused on students from five universities of technology in Taiwan, roughly 20% of the technical university population. The return rate for the student questionnaire was 80.3% of students (321 out of 400 questionnaires issued) who had to comply with the regulation and 75% of students (300 out of 400 questionnaires issued) who did not face the regulation.

Table 6. 1 Students' demographic information and their learning contexts (SQ1-7)

Items	Variables	G1 (%)	G2 (%)	Value	Chi-square significance
1. Gender*	Female	59.8	67.7	4.132	.042*
	Male	40.2	32.3		
2. Year of study*	First	63.2	97.7	115.913	.000*
	Second	30.2	1.0		
	Third	3.7	1.0		
	Fourth	2.8	0.3		
3. Major*	Commerce	39.6	60.0	130.037	.000*
	Health and life sciences	35.4	0.0		
	Ocean Engineering	25.0	40.0		
4. Medium of instruction*	English only	11.9	20.3	23.805	.000*
	Half English and half Chinese	71.1	52.0		
	Mainly Chinese	17.0	27.7		
5. Study hours per week*	0	34.1	24.3	13.689	.003*
	1-2	55.9	63.0		
	3-4	5.0	10.0		
	Other	5.0	2.7		
6. Perception of the regulation*	Yes	95.6	43.1	204.407	.000*
	No	4.4	56.9		
7. Test-oriented courses	Yes	19.9	23.7	1.267	.260
	No	80.1	76.3		

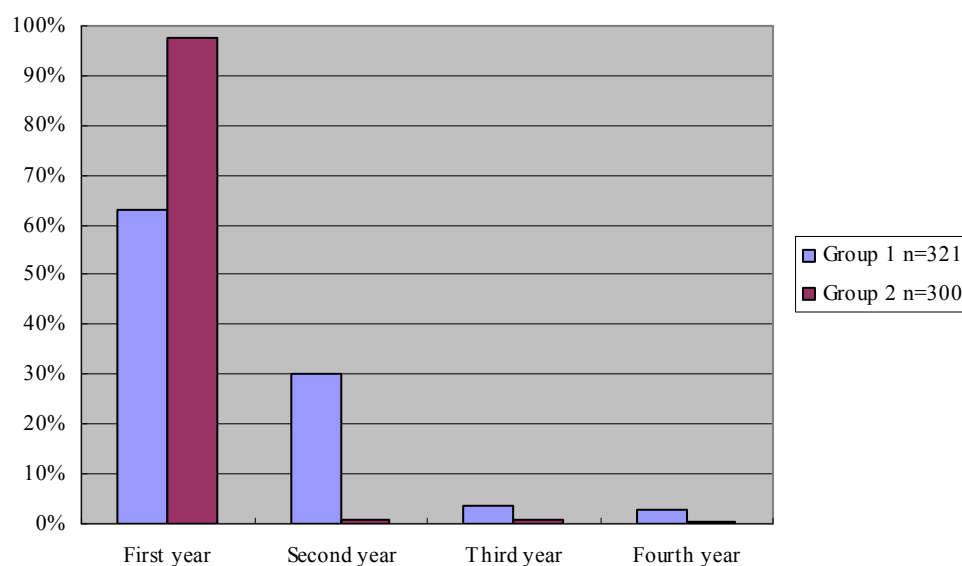
G1 = students who had to comply with the regulation

G2 = students who did not face the regulation

* = significant at $p < 0.05$

A chi-square test was used to evaluate the discrepancy (the degree of relativity) between students' demographic information and their learning contexts in the two sample groups. Seven student characteristics and learning contexts were explored in the study: gender, year of study, major, medium of instruction, weekly English study hours, perception of the existence of graduation regulation, and test-oriented courses taken. Six questions in this section were designed on a nominal scale and had a multiple-choice format, where students were required to choose one item from each question. The chi-square results in Table 6.1 showed that there were significant differences in six students' characteristics and learning contexts between the two groups (marked by asterisks).

Figure 6. 1 Students' year of study (SQ2)



There was a significant difference in the category of gender and the majority of the students were in their first or second year at university (see Figure 6.1). Such an imbalance was expected due to the data collection procedures (see Chapter 4) in which teachers were approached first and they then distributed the questionnaires to their students. Since the teachers only taught first and/or second year non-English majors, who were the parties

required to take General English courses, I had no control over the choice of student participants. With regard to the variety of majors, students in different majors had an opportunity to give their opinions on the new regulation, providing a range of different viewpoints.

In terms of the medium of instruction, there was a significant difference between the groups as shown in Table 6.1 and Figure 6.2. It can be seen that there was a higher proportion of teachers using half English and half Chinese in Group 1 (71.1 per cent of 318) than in Group 2 (52 per cent of 300). There was a lower proportion in Group 1 teachers' use of mainly Chinese (17 per cent of 318) compared with Group 2 teachers (27.7 per cent of 300). However, there was a lower proportion in Group 1 teachers' use of English only (11.9 per cent of 318), compared with Group 2 teachers (20.3 per cent of 300). Overall, the dominant trend among their teachers when students were taught English was to teach half in English and half in Chinese. These findings were consistent with the results from the same parallel question in the Teacher Questionnaire (TQ17).

Figure 6.2 Medium of instruction that teachers use in the classrooms (SQ4)

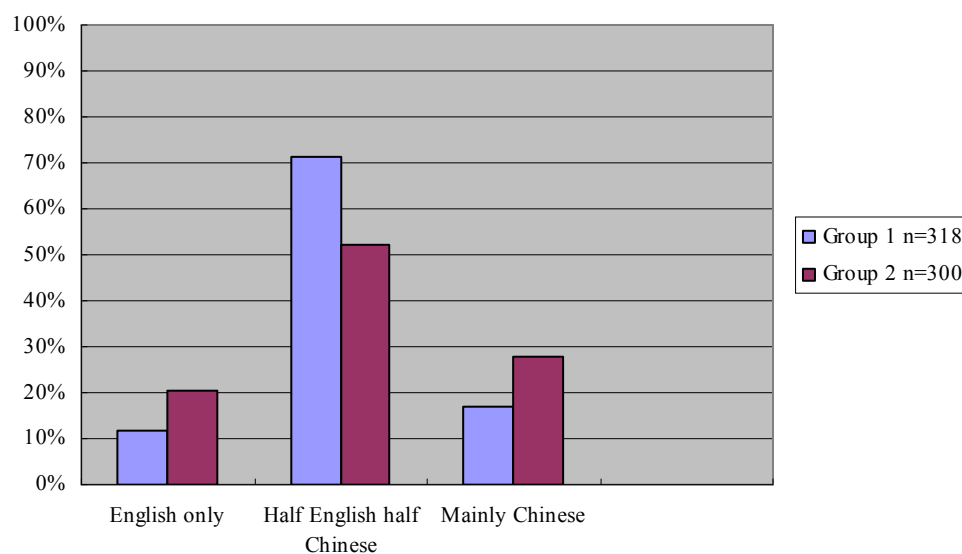


Table 6.1 and Figure 6.3 show that there was a significant difference in students' English study hours per week, but the pattern was nevertheless generally similar. The majority of students studied English less than two hours a week after class. In the items 1-2 and 3-4 hours per week, Group 1 students (55.9 per cent and 5 per cent of 320) reported spending less time on studying English compared with Group 2 students (63 per cent and 10 per cent of 300), and a higher proportion (34.1 per cent of 320) reported zero hours. The number of study hours reported might suggest that there was no increase in exam preparation in response to the new regulation that the Group 1 students should sit for English proficiency tests. It might also indicate that with respect to its aim of motivating students to do self-study in English, the graduation regulation had not been so effective.

Figure 6.3 Students' study hours per week (SQ5)

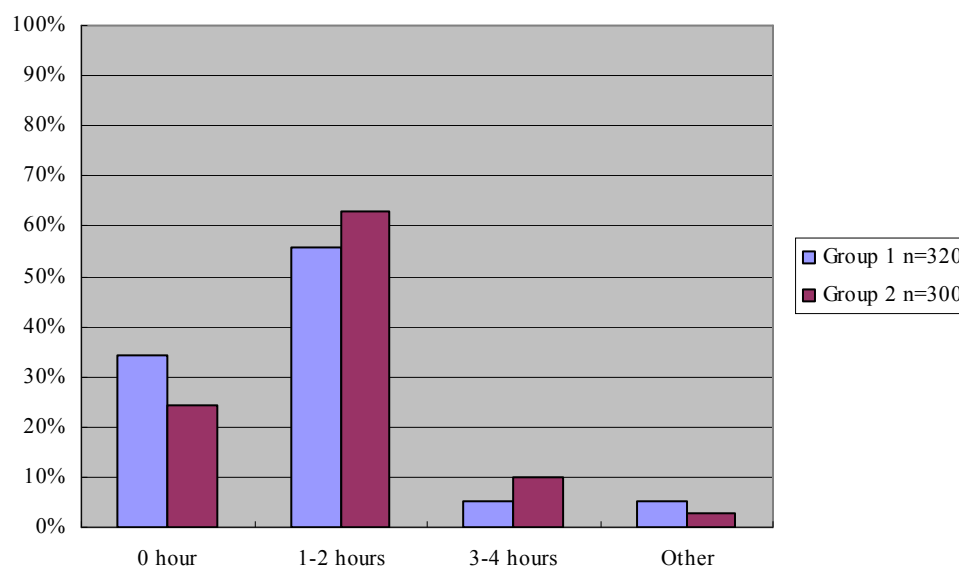
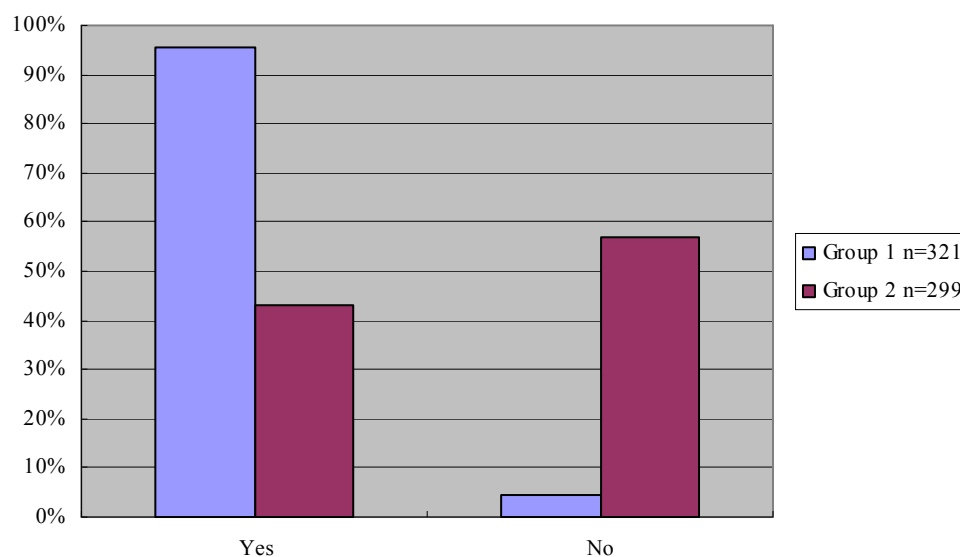


Table 6.1 and Figure 6.4 show that there was a significant difference in students' perceptions of the existence of the English graduation regulation. The majority of Group 1 students (95.6 per cent of 321) said they knew that they had to comply with the new regulation, suggesting that the dissemination of the new regulation had been successful.

Nevertheless, it can be seen that the majority of the students were not taking courses in preparation for English proficiency tests, either in the university or at cram schools (see Table 6.1). There was no significant difference regarding the test-oriented courses, suggesting that there was no real washback effect.

Figure 6.4 Students' acknowledgement of the regulation (SQ6)



In summary, half English and half Chinese was the dominant medium of instruction in Group 1 classrooms. Students who had to comply with the new regulation did not spend more time on studying English and they did not attend test-oriented courses. Both findings suggested that little attention was given to English language input in lessons and the students paid relatively little attention to the new regulation, even though they were well aware of its existence.

6.3 Students' attitudes towards teaching activities inside and outside class

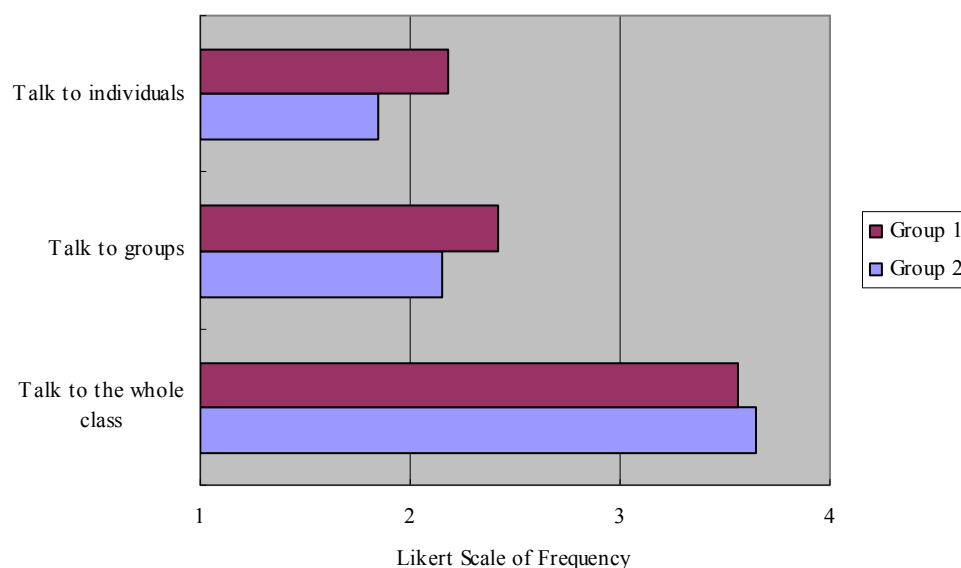
This section explored the students' attitudes towards teaching activities carried out inside and

outside the classroom by their teachers and by themselves in order to understand possible changes that might have taken place in the context of the new regulation. There were five questions, which were Student Questionnaire: Part Two, Section A, Questions 8, 9, 10, 11 and 12. All these questions were designed on a 4-point Likert scale of frequency, where 1 = seldom or never and 4 = every class.

6.3.1 Teachers' talk in class

SQ8 explored the students' perceptions of how much the teachers talked in their English lessons. The aim was to identify whether the new regulation reduced the amount of teacher talk and encouraged more communicative teaching approaches, and consequently provided more practice opportunities for the students. The students were asked to grade how often their teachers talked to the whole class, groups of students, and individual students in their English lessons.

Figure 6. 5 Students' perceptions of their teachers' talk (SQ8)



The general pattern of teachers' talk in the two groups is shown in Figure 6.5. Both groups of teachers seemed to talk to the whole class the most, followed respectively by talking to groups, and individual students. This situation was more or less the same in both groups. However, some differences were observed in teacher talk in the two different learning contexts. Table 6.2 shows that there was a significant difference in the students' perceptions of the last two items regarding their teachers' talk in their English lessons (marked by asterisks).

There was a higher frequency of teacher talk to the groups and to individual students in Group 1, which consisted of students who had to comply with the regulation. It seemed that Group 1 teachers provided more practice opportunities for the students than Group 2 teachers did. However, from the mean scores in Table 6.2, it can be seen that the general pattern of teacher talk in class in Group 1 was still in the order of, from most to least, 'talking to the whole class', 'talking to groups', and finally, 'talking to individuals' (see Figure 6.5). Group 1 teachers were still very much in control, though the new policy did appear to encourage them to be more learner-centred and to provide students with more practice opportunities.

Table 6.2 Differences in students' perceptions of their teachers' talk (SQ8)

Variables	Group	Cases	Mean	SD	T-Value	df	2-Tail Prob.
Talk to the whole class	1	320	3.56	.600	-1.893	611.310	.059
	2	300	3.65	.506			
Talk to groups of students*	1	301	2.42	.686	5.028	585.632	.000*
	2	289	2.15	.618			
Talk to individual students*	1	300	2.18	.758	5.573	580.391	.000*
	2	288	1.85	.659			

Group 1 = students who had to comply with the regulation

Group 2 = students who did not face the regulation

* = significant at $p < 0.05$

6.3.2 Students' perceptions of their teachers' teaching activities in class

SQ9 explored the students' perceptions of their teachers' teaching activities in their English lessons. Students were asked to grade the frequency with which their teachers organised five different activities in class (see Table 6.3 and Figure 6.6). Figure 6.6 shows the general pattern of the responses for both groups.

Figure 6.6 Students' perceptions of their teachers' teaching activities (SQ9)

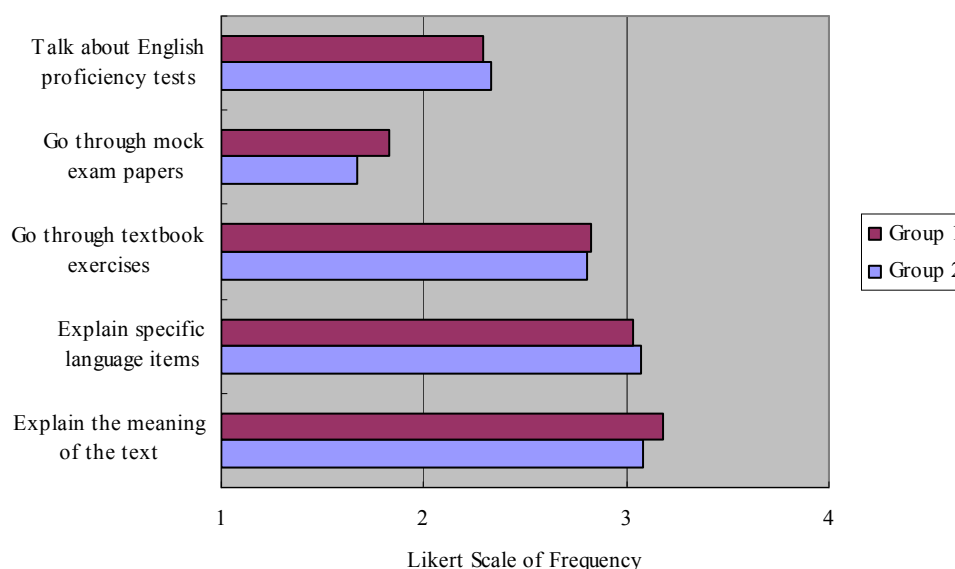


Table 6.3 shows that only one activity differed significantly and it was carried out more often in Group 1 than in Group 2: 'Going through mock exam papers' (marked by an asterisk). Clearly, this can help students prepare for the English proficiency tests, suggesting a washback effect of the regulation on classroom activities. On the other hand, Group 1 teachers mentioned the English proficiency tests *less* frequently than Group 2 teachers did. This difference, however, was not significant. The remaining activities, such as explaining the meaning of the text, explaining specific language items, and going through textbook exercises were relatively ordinary teaching activities and there was no significant difference between

the groups.

Table 6.3 Students' perceptions of their teachers' teaching activities (SQ9)

Variables	Group	Cases	Mean	SD	T-Value	df	2-Tail Prob.																																												
Explain the meaning of the text	1	319	3.18	.705	1.675	617	.094																																												
	2	300	3.08	.713				Explain specific language items	1	317	3.03	.834	-.615	609.418	.539	2	300	3.07	.717	Go through textbook exercises	1	315	2.83	.862	.226	613	.821	2	300	2.81	.826	Go through mock exam papers*	1	313	1.83	.900	2.359	608	.019*	2	297	1.67	.775	Talk about English proficiency tests	1	317	2.29	.892	-.537	599.439	.591
Explain specific language items	1	317	3.03	.834	-.615	609.418	.539																																												
	2	300	3.07	.717				Go through textbook exercises	1	315	2.83	.862	.226	613	.821	2	300	2.81	.826	Go through mock exam papers*	1	313	1.83	.900	2.359	608	.019*	2	297	1.67	.775	Talk about English proficiency tests	1	317	2.29	.892	-.537	599.439	.591	2	295	2.33	.726								
Go through textbook exercises	1	315	2.83	.862	.226	613	.821																																												
	2	300	2.81	.826				Go through mock exam papers*	1	313	1.83	.900	2.359	608	.019*	2	297	1.67	.775	Talk about English proficiency tests	1	317	2.29	.892	-.537	599.439	.591	2	295	2.33	.726																				
Go through mock exam papers*	1	313	1.83	.900	2.359	608	.019*																																												
	2	297	1.67	.775				Talk about English proficiency tests	1	317	2.29	.892	-.537	599.439	.591	2	295	2.33	.726																																
Talk about English proficiency tests	1	317	2.29	.892	-.537	599.439	.591																																												
	2	295	2.33	.726																																															

Group 1 = students who had to comply with the regulation

Group 2 = students who did not face the regulation

* = significant at $p < 0.05$

6.3.3 Students' classroom learning activities

SQ10 investigated the students' learning activities in their English lessons in order to see whether the new regulation had brought differences in their classroom learning activities. This question explored what the students did and how much opportunity for practice they had in their lessons. From the classroom observations (to be discussed in Chapter 7) it was clear that very often students did not have much control over the choice of learning activities in the classroom. The activities they actually carried out in the class comprised the practice opportunities (learning opportunities) their teachers assigned to them. There were ten activities, listed in Table 6.4. Three of the activities showed no significant difference between the groups: reading, listening, and writing. The remaining seven showed significant differences (marked by asterisks).

The seven learning activities concerned were speaking, practising grammar items, learning vocabulary, carrying out pair work, carrying out group discussions, doing language

games, and doing mock exam papers, and all were carried out less often in Group 1 than in Group 2 (see Figure 6.7). Among the activities, speaking, pair work and group discussions were expected by the authorities to be carried out more often in Group 1 because they believed that those activities provided students with more oral communication practice. However, the expected washback effect was not proven by the results. Moreover, contrary to my expectations, but consistent with the result of TQ13, mock exams were taken less often in class or in the self-study centre by Group 1 students, who had to comply with the regulation. The results indicated that there was some focus on test-oriented classroom activities to reflect the tests in the universities, but less than one might have expected (see Chapters 2 and 4).

Table 6. 4 Students' perceptions of their classroom activities (SQ10)

Variables	Group	Cases	Mean	SD	T-Value	df	2-Tail Prob.
Reading	1	320	2.65	.887	-.066	612.366	.948
	2	298	2.65	.764			
Listening	1	315	2.34	.915	-1.406	613	.160
	2	300	2.44	.854			
Writing	1	315	1.67	.770	-1.375	612	.170
	2	299	1.75	.777			
Speaking*	1	315	2.10	.879	-6.154	610.346	.000*
	2	299	2.54	.879			
Practising grammar items*	1	314	2.16	.852	-4.943	611	.000*
	2	299	2.48	.774			
Learning vocabulary*	1	313	2.42	.871	-2.682	608.636	.008*
	2	299	2.61	.793			
Carrying out pair work*	1	313	1.75	.832	-10.439	575.742	.000*
	2	300	2.54	1.025			
Carrying out group discussion*	1	314	1.68	.829	-9.019	598.437	.000*
	2	299	2.31	.912			
Doing language games*	1	312	1.56	.732	-9.386	562.463	.000*
	2	298	2.20	.935			
Doing mock exam papers*	1	315	1.55	.718	-4.236	613	.000*
	2	300	1.80	.732			

Group 1 = students who had to comply with the regulation

Group 2 = students who did not face the regulation

* = significant at $p < 0.05$

Figure 6. 7 Students' perceptions of their classroom activities (SQ10)

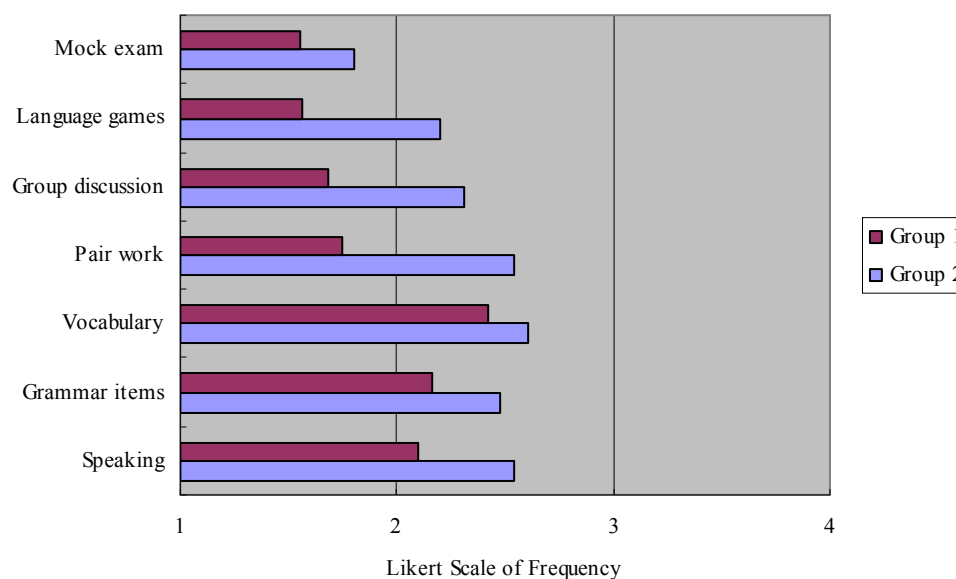


Table 6. 5 Differences in the students' classroom activities between Year1 and Year2 (SQ10)

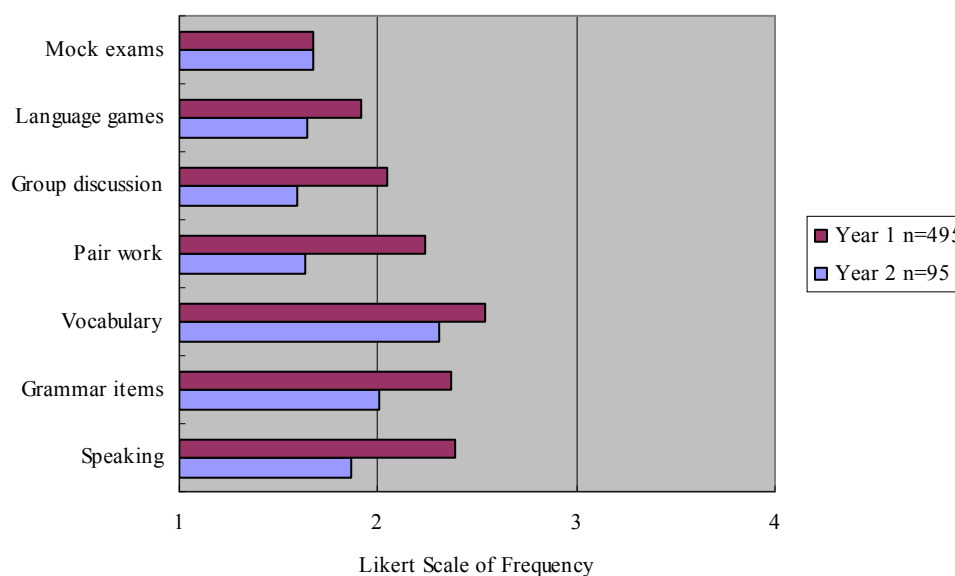
Variables	Year	Cases	Mean	SD	T-Value	df	2-Tail Prob.
Speaking*	1	494	2.39	.901	5.534	140.243	.000*
	2	95	1.87	.828			
Practising grammar items*	1	494	2.37	.827	4.207	141.517	.000*
	2	95	2.01	.751			
Learning vocabulary*	1	492	2.54	.832	2.520	585	.012*
	2	95	2.31	.839			
Carrying out pair work*	1	495	2.24	1.017	6.339	151.368	.000*
	2	94	1.63	.829			
Carrying out group discussion*	1	494	2.05	.920	4.599	586	.000*
	2	94	1.59	.835			
Doing language games*	1	492	1.92	.918	2.759	584	.006*
	2	94	1.64	.774			
Doing mock exam papers	1	496	1.67	.717	-.060	588	.952
	2	94	1.67	.781			

* = significant at $p < 0.05$

Investigation of whether teachers assigned activities differently because of students' study year showed that, of the seven activities, six (marked by asterisks) were carried out significantly differently between Year 1 and Year 2 students. It appeared that Year 1 students carried out those six activities more often than Year 2 students (see Table 6.5 and Figure 6.8).

The results might suggest that Year 1 students obtained more practice opportunities than Year 2 students.

Figure 6. 8 Differences in the students' classroom activities between Year1 and Year2 (SQ10)



The above results regarding students' perceptions of their teachers' teaching activities and their own learning activities in the classroom revealed a discrepancy. According to the students in SQ9, teachers explained mock exams more often in Group 1 classes, yet Group 1 students reported they did mock exams less in SQ10. The discrepancy perhaps demonstrated a weakness in using a questionnaire to obtain information regarding actual teaching and learning behaviours in the classroom. However, what was not discovered via the questionnaire was further explored via the use of classroom observations in the study, which will be discussed in the next chapter.

6.3.4 Students' use of English in relation to their classroom activities

SQ11 explored the activities the students carried out in English in their English lessons. There were five activities listed under the question. Three activities were perceived as being carried out significantly differently by the students in the two groups: doing pair work, holding group discussions, and asking the teacher questions (marked by asterisks in Table 6.6 and see Figure 6.9). Group 1 students asked the teacher questions more often in English in class, while pair work and group discussions were carried out less often in English. This would seem to indicate that the Group 1 students were more confident about using English in class while learning activities involved their teachers, but they were reluctant to use English while doing pair work and holding group discussions.

Figure 6.9 Students' use of English in relation to their classroom activities (SQ11)

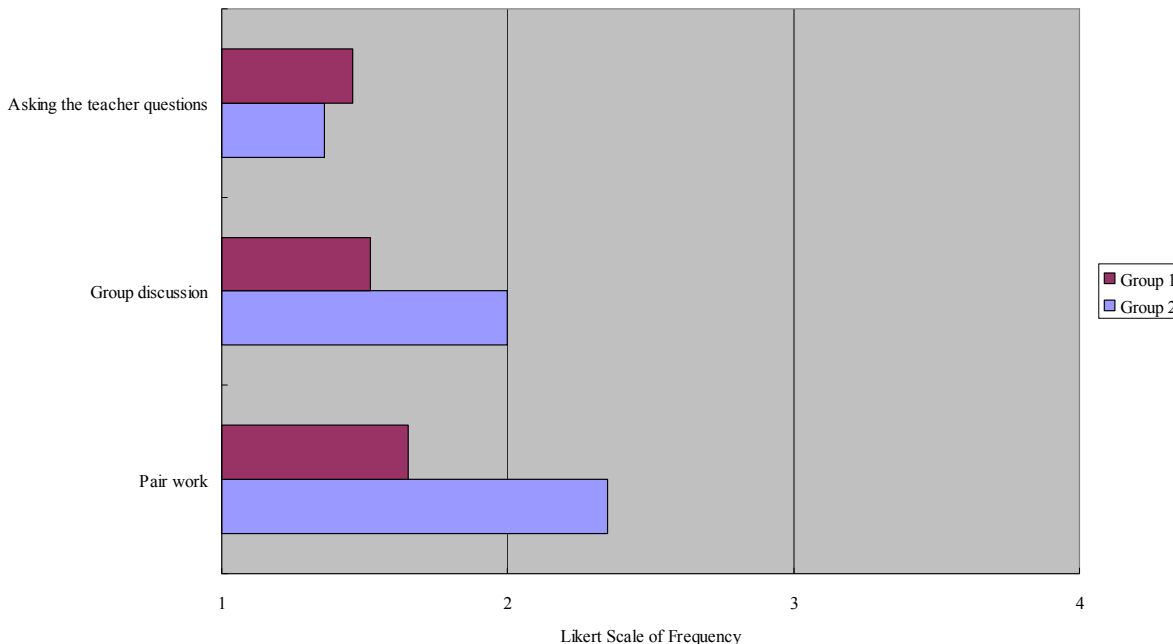


Table 6. 6 Students' use of English in relation to their classroom activities (SQ11)

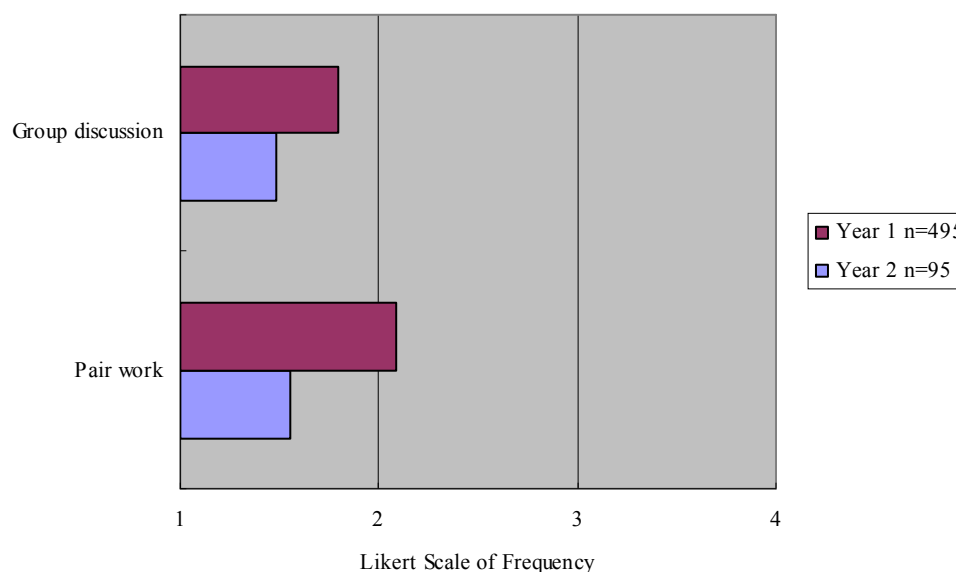
Variables	Group	Cases	Mean	SD	T-Value	df	2-Tail Prob.
Asking a fellow student questions	1	316	1.66	.791	-.010	613	.992
	2	299	1.66	.708			
Doing pair work*	1	318	1.65	.814	-9.425	575.478	.000*
	2	300	2.35	1.005			
Holding group discussions*	1	315	1.52	.767	-7.236	613	.000*
	2	300	2.00	.864			
Asking the teacher questions*	1	317	1.46	.700	2.015	597.795	.044*
	2	299	1.36	.559			
Expressing your own ideas	1	315	1.35	.643	.576	612	.565
	2	299	1.32	.554			

Group 1 = students who had to comply with the regulation

Group 2 = students who did not face the regulation

* = significant at $p < 0.05$

Figure 6. 10 Year 1 and Year 2 students' use of English in relation to their classroom activities (SQ11)



As shown in Figure 6.10, there was evidence of differences between Year 1 and Year 2 students in the frequency of pair work and group discussions. On further inspection it was found that Year 1 students carried out pair work and group discussions in English in class more often than Year 2 students. The results suggested that, while carrying out activities in

class, Year 1 students used English more often than Year 2 students. In short, there appeared to be no direct washback effect on learning activities in the classroom.

6.3.5 Students' use of English in relation to their learning activities outside class

SQ12 investigated aspects of the students' learning activities carried out in English outside their English lessons. There were seven activities listed under this question. Four of these activities were carried out significantly differently: talking to classmates, listening to the radio or CDs, reading newspapers or magazines, and doing online mock exams (marked by asterisks), all of which were carried out less frequently in Group 1 than Group 2 (see Table 6.7). The results suggested that the graduation regulation did not motivate Group 1 students to make the increased effort to learn English expected by the university authorities.

Table 6.7 Students' use of English in relation to their learning activities outside class (SQ12)

Variables	Group	Cases	Mean	SD	T-Value	df	2-Tail Prob.
Watch TV or movies	1	318	2.52	.843	.551	616	.582
	2	300	2.48	.765			
Listen to the radio or CDs*	1	314	1.81	.726	-2.160	604.604	.031*
	2	300	1.92	.621			
Read newspapers and magazines*	1	317	1.72	.641	-3.217	614.455	.001*
	2	300	1.88	.625			
Talk to classmates*	1	316	1.67	.661	-2.870	611.842	.004*
	2	299	1.83	.653			
Do online mock exams*	1	315	1.55	.628	-4.670	610.294	.000*
	2	299	1.79	.629			
Talk to teachers	1	317	1.43	.583	-.308	615	.758
	2	300	1.44	.554			
Talk to other people outside university	1	317	1.38	.542	-.650	615	.516
	2	300	1.41	.538			

Group 1 = students who had to comply with the regulation

Group 2 = students who did not face the regulation

* = significant at $p < 0.05$

In the remaining activities outside class there was no significant difference between the two groups. They are listed in Table 6.7 in descending order according to the mean scores for Group 1. This gives a general picture of the frequency of activities that students carried out in English outside class. From the list, it can be seen that watching TV was the most frequently carried out activity the students did in English outside the class, whereas talking to other people outside university was carried out the least frequently. This is probably due to the limited English communication needs in Taiwan outside the classroom.

To sum up, the results of Part Two: Section A – Students’ attitudes towards teaching activities inside and outside class – showed that teachers talked more often to the groups and individual students in Group 1, which might suggest more practice opportunities for students in class. Secondly, there was also a higher frequency of teachers carrying out classroom activities in relation to English proficiency tests (going through mock exam papers) in Group 1, but Group 1 students reported that they did mock exams less often than Group 2 students. Thirdly, there was a lower frequency of learning activities such as speaking, pair work, group discussions, and language games in Group 1, which might suggest fewer communicative practice opportunities for students in class.

As for student learning activities involving English in class and outside class, there was a much lower frequency of carrying out pair work and group discussions in Group 1, showing no positive washback effect on classroom learning activities. Group 1 students, however, did report talking more to their teachers in English. This result supports the information obtained from the teacher interviews (see Chapter 7). One teacher mentioned that an ‘English Corner’ was set up to encourage students to communicate with teachers in English. However, it can be seen from the mean score of SQ 12 – Students’ use of English in relation to their learning activities outside class – that Group 1 students still did not carry out many activities in English outside class. Nevertheless, overall Year 1 students seemed to show more concern for, and make more effort at, learning English than the Year 2 students.

6.4 Students' attitudes towards aspects of learning

There were four questions in Part Two: Section B of the Student Questionnaire, designed using a nominal scale and a multiple-choice format. This section aimed to explore possible reasons for the students' attitudinal differences in relation to learning in the context of taking English proficiency tests as a graduation regulation.

6.4.1 Students' motivation to learn English

SQ13 explored the students' reasons for learning English, to see whether the regulation brought about different perceptions in the students' motivation. Only three reasons out of the eight in the question differed significantly between the two groups (marked by asterisks).

It can be seen from Table 6.8 and Figures 6.11 – 6.13 that satisfying university requirements was seen as a more motivating reason to learn English in Group 1 (89.4 per cent of 320) than in Group 2 (75 per cent of 300). A lower proportion of Group 1 students (89 per cent of 281) agreed that meeting the expectations of society at large was an important reason to learn English compared with Group 2 students (93.7 per cent of 271). Similarly, being able to communicate with foreigners was seen to be less of a reason to learn English in Group 1 (84.7 per cent of 321) than in Group 2 (91.3 per cent of 300). The results, consistent with the results of TQ 6, would suggest that the Group 1 students were largely instrumentally motivated to learn English.

Table 6. 8 Students' perceptions of their reasons for learning English (SQ13)

Variables	Groups	Cases	Agree (%)	Disagree (%)	Value	Chi-square significance
To obtain a 'good' job	1	321	95.6	4.4	1.303	.254
	2	300	97.3	2.7		
To be able to make a success of your career	1	313	93.3	6.7	1.525	.217
	2	295	95.6	4.4		
To pass examinations	1	320	91.6	8.4	.039	.843
	2	300	92.0	8.0		
To satisfy university requirements*	1	320	89.4	10.6	22.082	.000*
	2	300	75.0	25.0		
To meet the expectations of society*	1	281	89.0	11.0	3.935	.047*
	2	271	93.7	6.3		
To be able to watch English movies and listen to English radio programmes	1	319	86.2	13.8	1.403	.236
	2	300	89.3	10.7		
To be able to communicate with foreigners*	1	321	84.7	15.3	6.358	.012*
	2	300	91.3	8.7		
To pursue further studies	1	319	65.2	34.8	.236	.627
	2	300	63.3	36.7		

Group 1 = students who had to comply with the regulation

Group 2 = students who did not face the regulation

* = significant at $p < 0.05$

Figure 6. 11 Learning reasons – To satisfy university requirements (SQ13)

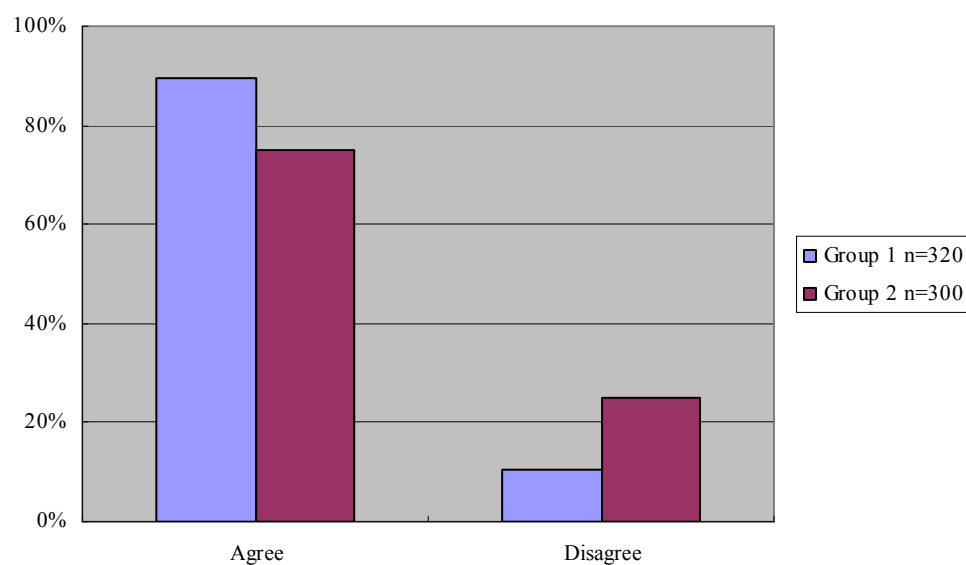


Figure 6. 12 Learning reasons – To meet society's expectations (SQ13)

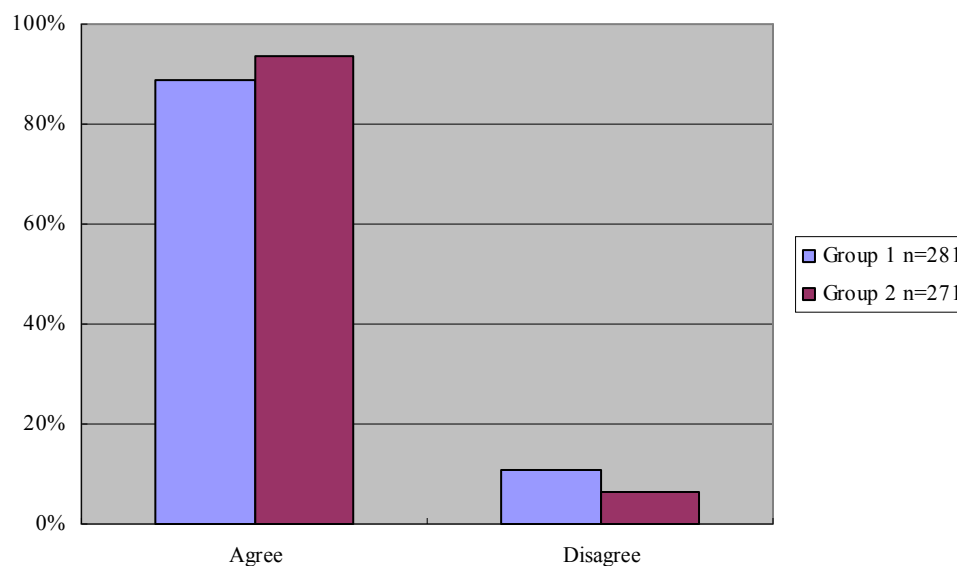
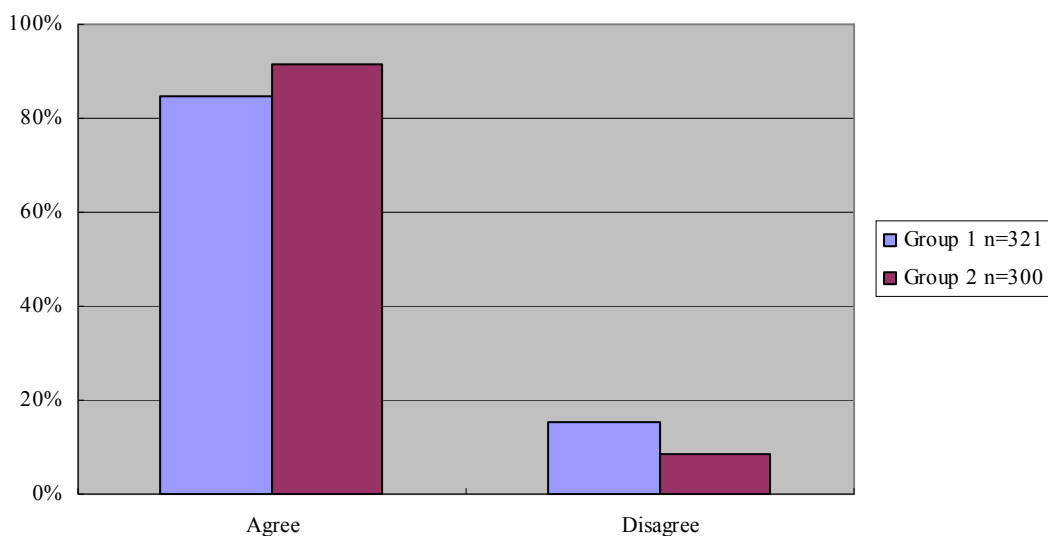


Figure 6. 13 Learning reasons – To be able to communicate with foreigners (SQ13)



The remaining reasons did not differ significantly. Table 6.8 also provides a general picture of the students' reasons for learning English and serves as a comparison of the reasons that students perceived as being important. The factors perceived by Group 1 students as being most important were related to instrumental reasons, such as getting a job, passing examinations, and meeting university regulations. The reasons perceived as less motivating

were related to functional uses of the English language and pursuing further studies. The situation appeared to partially fit the spirit of the new regulation.

6.4.2 Students' preferred learning strategies

SQ14 investigated whether or not the graduation regulation had had an influence on the students' learning strategies. There were 12 strategies listed under this question. Seven learning strategies differed significantly between the two groups (marked by asterisks). These learning strategies were, however, seen to be *less* used by the students in Group 1, who had to comply with the regulation, than those in Group 2, who did not face the regulation (see Table 6.9).

The remaining learning strategies did not differ significantly. Table 6.9 presents a general picture of the situation. The least used learning strategy was 'expressing opinions in English in class', which showed that there was no washback effect on oral communicative skills, even though these were required skills for dealing with the English proficiency tests and the authorities expected students to develop these skills.

Table 6.9 Students' preferred learning strategies (SQ14)

Variables	Groups	Cases	Yes (%)	No (%)	Value	Chi-square significance
Watching TV or movies in English*	1	321	80.1	19.9	5.885	.015*
	2	299	87.3	12.7		
Learning vocabulary*	1	321	71.3	28.7	18.710	.000*
	2	300	85.7	14.3		
Doing exercises and homework*	1	320	64.1	35.9	4.873	.027*
	2	300	72.3	27.7		
Taking notes*	1	320	62.2	37.8	4.587	.032*
	2	300	70.3	29.7		
Learning grammar rules	1	320	45.9	54.1	2.805	.094
	2	300	52.7	47.3		
Taking part in group activities in class*	1	320	42.2	57.8	35.314	.000*
	2	300	66.0	34.0		
Listening to English radio programmes and CDs	1	320	40.0	60.0	3.387	.066
	2	300	47.3	52.7		
Reading English newspapers and magazines*	1	318	38.1	61.9	7.090	.008*
	2	300	48.7	51.3		
Practising English writing	1	320	30.9	69.1	.000	.987
	2	300	31.0	69.0		
Taking optional English lessons*	1	320	30.6	69.4	8.197	.004*
	2	300	41.7	58.3		
Participating in extracurricular activities	1	320	22.2	77.8	.003	.955
	2	300	22.0	78.0		
Expressing opinions in English in class	1	320	20.6	79.4	.663	.415
	2	300	23.3	76.7		

Group 1 = students who had to comply with the regulation

Group 2 = students who did not face the regulation

* = significant at $p < 0.05$

6.4.3 Students' attitudes towards whether and in which aspects they are influenced by the English proficiency tests

SQ15 was primarily designed to ask students directly whether or not they had ever been affected by English proficiency tests. A significant difference in the students' attitudes towards this question was observed between the groups. As shown in Table 6.10 and Figure 6.14, a lower proportion of Group 1 students (39.7 per cent of 317) reported being affected by the tests, compared with Group 2 students (48 per cent of 298). A clear 'no' was the dominant response from Group 1.

Table 6. 10 Students who had and had not been affected by English proficiency tests (SQ15)

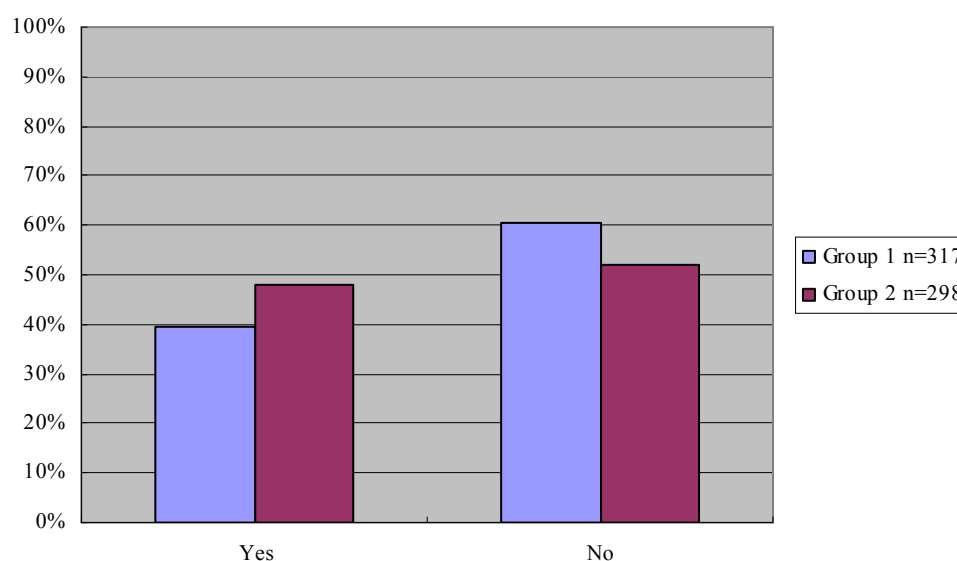
Variables	Groups	Cases	Yes (%)	No (%)	Value	Chi-square significance
Affected by English proficiency tests*	1	317	39.7	60.3	4.237	.040*
	2	298	48	52		

Group 1 = students who had to comply with the regulation

Group 2 = students who did not face the regulation

* = significant at $p < 0.05$

Figure 6. 14 Students who were and were not affected by the English proficiency tests (SQ15)



SQ16 then explored further those aspects of the students’ lives that were affected by English proficiency tests. There were six aspects in this question and two of them showed significant differences (marked by asterisks). It can be seen from Table 6.11 and Figures 6.15 and 6.16 that those students who reported being affected by English proficiency tests in SQ15 reported that taking English proficiency tests did not lead to a lack of motivation to learn English. However, those students in Group 1 (64.2 per cent of 123) still experienced greater anxiety compared with those in Group 2 (47.8 per cent of 138). The results suggested that there may well have been some negative washback effects on students’ emotions.

Table 6. 11 Aspects in which students are affected by the English proficiency tests (SQ16)

Variables	Groups	Cases	Yes (%)	No (%)	Value	Chi-square significance
Future job opportunities	1	114	88.6	11.4	.830	.362
	2	126	92.1	7.9		
Clearer on learning objectives	1	124	69.4	30.6	1.047	.306
	2	140	75.0	25.0		
Anxiety and emotional tension*	1	123	64.2	35.8	7.085	.008*
	2	138	47.8	52.2		
More interested in learning English	1	124	44.4	55.6	1.908	.167
	2	138	52.9	47.1		
Less interested in learning English*	1	124	41.1	58.9	13.467	.000*
	2	138	20.3	79.7		
Self-image	1	112	31.3	68.8	.417	.518
	2	124	27.4	72.6		

Group 1 = students who had to comply with the regulation

Group 2 = students who did not face the regulation

* = significant at $p < 0.05$

Figure 6. 15 Aspect affected – Anxiety and emotional tension (SQ16)

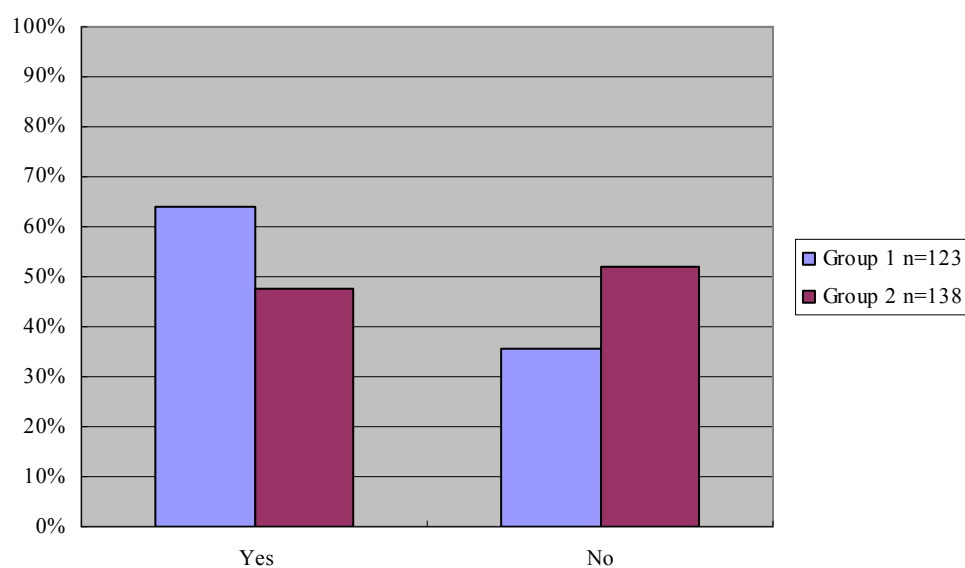
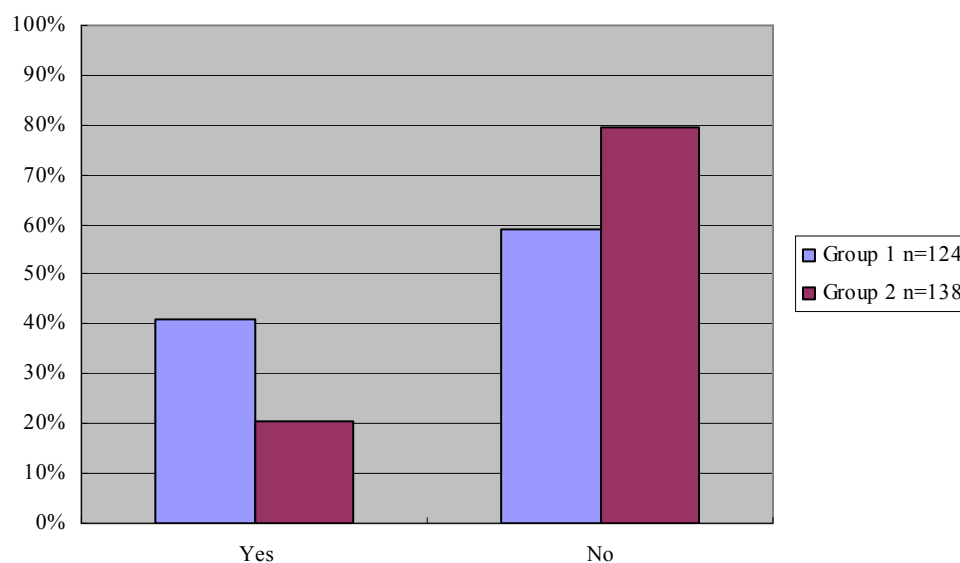


Figure 6.16 Aspect affected – Less interested in learning English (SQ16)



The results from this question demonstrated that there was no significant difference in the test's influence on the other aspects of the students' lives. As shown in Table 6.11, students reported that English proficiency tests had the greatest impact on their future job opportunities compared with other aspects of their lives, which was consistent with the results of SQ13, where students explained that the most important reason for learning English was to obtain a good job. This may well be due to the language requirement demanded by employers when students applied for jobs (see Chapter 2).

In summary, the results in Part Two Section B showed that there were differences in many aspects of student learning. For example, more Group 1 students stated that meeting the requirements of the university, such as passing English proficiency tests, was an important reason to learn English. Students who had to comply with the new regulation perceived more instrumental purposes for learning English than the students who did not face the regulation. The results thus reflected a washback effect on motivation.

Regarding the students' preferred learning strategies, seven out of the twelve strategies showed a significant difference, but they were strongly favoured by Group 2 students rather than Group 1 students. A general pattern could be seen in the percentage of learning strategies

that were preferred by both groups of students; the learning strategies that involved a functional use of English were rated lower in general. In short, there was no apparent relationship between the learning strategies and the implementation of English proficiency tests.

From the results of this section, it was clear that a majority of the students were not influenced greatly by English proficiency tests, and Group 1 students perceived fewer influences of the tests than Group 2 did. However, those Group 1 students who reported having been affected by the English proficiency tests felt more anxiety and emotional tension than those in Group 2. The results indicated that where there was a test impact, it was likely to be negative, especially on emotions. The influences of English proficiency tests will be explored further in the following section.

6.5 Students' attitudes towards aspects of examinations

This section – Student Questionnaire Part Two: Section C – consisted of one question with 9 statements. All statements were designed on a nominal scale and used a multiple-choice format. The category explored the students' attitudes towards English examinations in relation to teaching and learning. Two main themes about examinations were included. One theme dealt with the impact of examinations on the students themselves, while the other dealt with the impact on the students' learning processes and outcomes. Table 6.12 provides a summary overview.

The students' attitudes towards four of the nine statements differed significantly between the two groups (marked by asterisks). The items concerned were 1 and 3, which related to the impact of the examinations on the students themselves, and Items 2 and 9, concerning the processes and outcomes of their learning (see Table 6.12).

Table 6. 12 Students' attitudes towards English examinations (SQ17)

Variables	Groups	Cases	Yes (%)	No (%)	Value	Chi-square significance
7. Examinations should NOT be used as the sole determiner of student grades.	1	297	93.6	6.4	.548	.459
	2	263	95.1	4.9		
3. Taking examinations is a valuable learning experience.*	1	296	80.4	19.6	10.050	.002*
	2	261	90.0	10.0		
4. Examinations force students to study harder.	1	296	79.1	20.9	1.275	.259
	2	262	82.8	17.2		
9. Every university graduating student should pass at least one English proficiency test.*	1	289	72.0	28.0	7.449	.006*
	2	259	81.9	18.1		
2. Students' learning is improved by practising mock exams.*	1	296	70.9	29.1	6.454	.011*
	2	263	80.2	19.8		
1. Examinations increase students' motivation to learn English.*	1	296	69.9	30.1	9.112	.003*
	2	263	81.0	19.0		
8. Mock examinations are important ways to learn.	1	294	67.7	32.3	1.236	.266
	2	261	72.0	28.0		
6. A student's performance on an examination is a good indication of how well she or he will be able to apply what has been learned.	1	296	63.5	36.5	.080	.778
	2	263	62.4	37.6		
5. A student's score on an examination is a good indication of how well she or he has learned the material.	1	296	60.1	39.9	2.455	.117
	2	263	66.5	33.5		

Group 1 = students who had to comply with the regulation

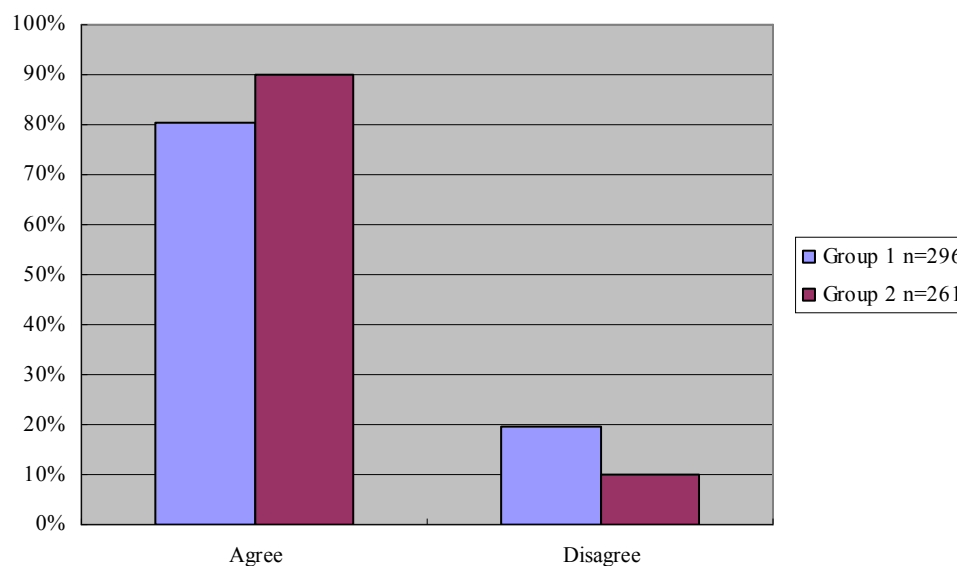
Group 2 = students who did not face the regulation

* = significant at $p < 0.05$

Figures 6.17–6.20 show that the Group 1 students did not seem to feel strongly that taking examinations was a valuable experience or that examinations motivated them to learn English. The results might suggest possible worry or tension in the students' attitudes towards preparing for English proficiency tests. In addition, some Group 1 students did not feel that their English learning was improved by practising mock exams. More Group 1 than Group 2 students thought passing English proficiency tests should *not* be a graduation requirement for

every non-English major student. As some students explained in written answers, although there was a trend towards hiring staff with English proficiency certificates, it was not applied to all professionals. Comparing the two groups, the results showed that Group 1 students generally expressed a more negative attitude towards English proficiency tests and paid less attention to them.

Figure 6.17 Attitudes towards English examinations – Taking examinations is a valuable learning experience (SQ17)



The students perceived the remaining statements in a similar fashion in the two groups. It can be seen from Table 6.12 that Item 7 ('Examinations should **NOT** be used as the sole determiner of student grades') produced the strongest agreement among the students. However, this was followed by Items 3 and 4 ('Taking examinations is a valuable learning experience,' and 'Examinations force students to study harder' respectively). This appeared to reflect mixed feelings about English examinations. On one hand, students did not think examinations were an accurate reflection of all aspects of their study, while on the other hand they were put in a position where they had to work hard to achieve the best examination

scores and comply with the new regulation. This was an understandable instrumental motivation for the students to learn English. On the whole, it can be seen from Table 6.12 and Figures 6.17 – 6.20 that students in both groups did have strong feelings about aspects of examinations in their university lives.

Figure 6. 18 Attitudes towards English examinations – Every university graduating student should pass at least one English proficiency test (SQ17)

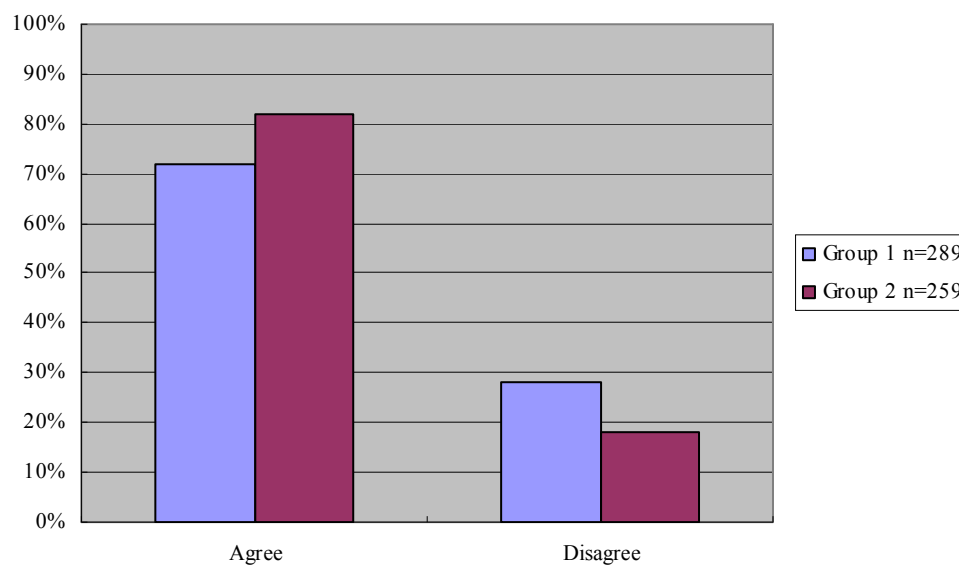


Figure 6. 19 Attitudes towards English examinations – Students’ learning is improved by practising mock exams (SQ17)

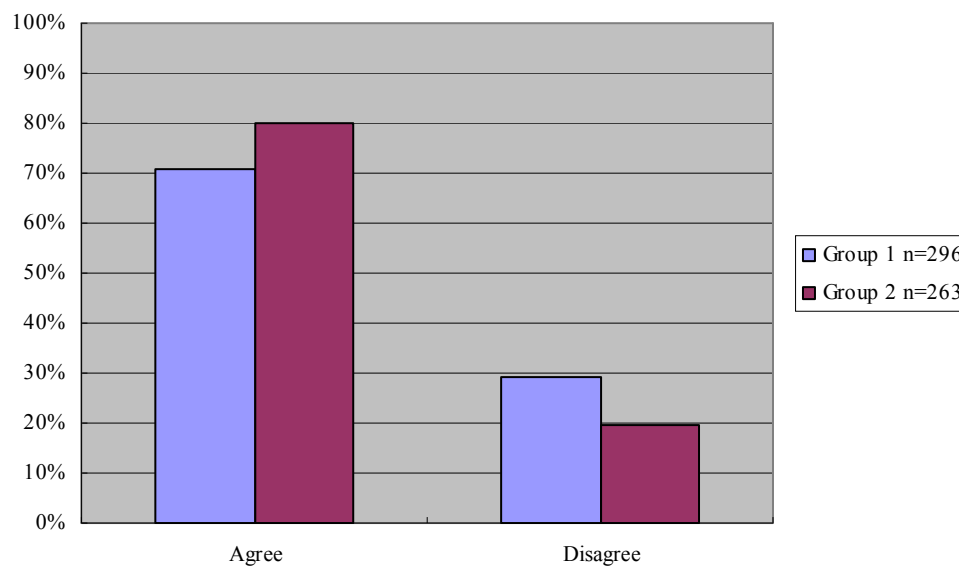
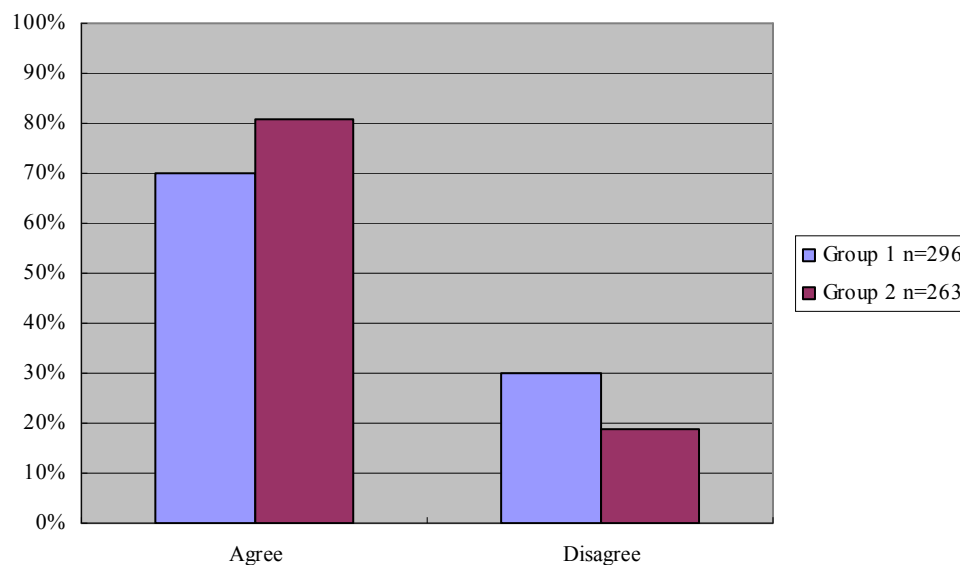


Figure 6.20 Attitudes towards English examinations – Examinations increase students' motivation to learn English (SQ17)



6.6 Summary

Findings from the Student Questionnaire in this chapter respond to the following main themes:

- students' characteristics and their learning contexts;
- students' attitudes towards classroom teaching and learning activities;
- students' attitudes towards aspects of learning;
- students' attitudes towards examinations.

Firstly, with respect to the students' learning contexts, a significant difference can be seen in the medium of instruction used in English lessons. More Group 1 teachers used half English and half Chinese predominantly as the medium of instruction, compared with Group 2 teachers (see Figure 6.2). Furthermore, it can be observed that students who had to sit English proficiency tests did not spend more time studying English and did not attend test-oriented classes, indicating a minimal impact of the new regulation on motivating students to do self-study in English and no more attention being given to examination

preparation.

Secondly, both groups of teachers seemed to talk to the whole class the most, followed respectively by talking to groups, and then to individual students (see Figure 6.5). This situation was relatively equal between the groups. No significant difference can be seen in the item 'talking to the whole class'. However, according to the students, Group 1 teachers talked more often to groups and to individuals than Group 2 teachers did, suggesting Group 1 students received more practice opportunities although their number was still minimal.

Regarding the students' attitudes towards teaching activities inside and outside class, there was a significant difference between the two groups for one item under the question 'Students' perceptions of their teachers' teaching activities in class'. 'Going through mock exam papers' was reported more often for the Group 1 classrooms, an activity which was closely related to the English proficiency tests to be taken. However, Group 1 students reported that they did mock exams less often than Group 2 students. In addition, there was a lower frequency of learning activities such as speaking, pair work, group discussions, and language games in Group 1, which might suggest fewer communicative practice opportunities for students in class. The results thus showed a minimal washback effect on classroom activities from the students' perspectives.

Furthermore, student activities carried out in English both inside and outside class showed significant differences. Pair work and group discussions in English were carried out less frequently in Group 1 compared with Group 2, and Group 1 students used less English outside class than Group 2 students. However, Group 1 students talked to their teachers slightly more in English than Group 2 students did. The results showed that overall there was actually a minimal washback effect on the classroom activities and the activities that students carried out inside and outside class.

The third part of the results relating to the students' attitudes towards learning also showed significant differences between the two groups. With respect to the students' motivation to learn English, there were significant differences in three out of the eight aspects. More Group 1 students thought the reason they learned English was 'to satisfy university

regulations' (see Figure 6.11), and more Group 2 students thought the reason they learned English was 'to meet the expectations of society' (see Figure 6.12) and 'to be able to communicate with foreigners' (see Figure 6.13). It seemed that Group 1 students were instrumentally motivated to learn English. The situation appeared to partially fit the spirit of the new regulation in which functional use of English was encouraged and expected.

In the case of the students' preferred learning strategies, seven out of the twelve strategies differed significantly between the groups, though the preference rates were *lower* for Group 1 than for Group 2 (see Table 6.9). The least used learning strategy was 'expressing opinions in English in class', suggesting there was minimal washback effect on oral communicative skills, even though they were required skills for dealing with the English proficiency tests and the authorities expected students to develop those skills. With regard to perceptions of the impact of English proficiency tests, although the majority of students in both groups reported not having been affected, those who said they had been affected, reported a negative reaction and effect.

The last section of the findings was related to the students' attitudes towards aspects of examinations. There were significant differences in only four out of the nine statements between the groups. Group 1 students generally expressed a more negative attitude towards English examinations. In both groups, the students showed mixed feelings about examinations. On one hand, they really did not like examinations. On the other hand, examinations seemed to be one of their key motivations for learning English. Therefore, imposing the new regulation might have changed certain aspects of the students' attitudes towards teaching and learning, but the regulation did not seem to have changed their attitudes towards the function of the examinations.

Chapter Seven: Teachers' and Students' Actions and Reactions

7.1 Introduction

This chapter reports the findings from Phase III of the study. It first emphasises the importance of the research framework used and then reports on the classroom observation studies of four teachers and the interviews with them. The chapter complements the questionnaire findings reported in Chapters 5 and 6. The main focuses in this phase of the study are related to the washback effect on:

- the English language syllabus within the Taiwanese universities of technology,
- teaching materials used in practice;
- teachers' teaching behaviours;
- teachers' beliefs about teaching and learning.

The observation study investigated how the new policy – passing English proficiency tests as a graduation requirement – influenced classroom teaching. The investigation was carried out with regard to the participants' reactions in the local educational context and to explore the effect of the new policy on aspects of teachers' attitudes, teaching content, and classroom interaction. It emphasised “the importance of context, setting, and subjects' frames of reference” (Marshall and Rossman, 2006: 54). The method used was aimed at capturing the reality, variation, and complexity of changes in day-to-day classroom practice within the Taiwanese educational context.

Smith et al. (1994) pointed out that the statements about educational change made by policymakers, as well as curriculum and assessment, can often be reshaped and translated imperfectly by practitioners, such as programme directors and teachers, and this can impact on both teachers and students. Teachers and programme directors can redefine and reinterpret the messages about policy that they have received. They then adapt, teach, learn, and evaluate

according to their own definitions of the situation (Brindley, 2008; Geisinger, 1994; Markee, 1997). Therefore, in order to discover whether the graduation regulation brought about washback effects on the English teaching and learning in Taiwanese universities of technology, it was important to explore various participants in teaching and learning settings.

In order to understand actions and practices involved, I engaged directly with the local scene, spending a reasonable amount of time to understand actions in the specific social context, and gaining as far as possible access to participant meanings, in line with the suggestions in Miles and Huberman (1994) and Robson (2002). The research question to be answered was:

- What was the nature and scope of the washback effect on teachers' behaviours as a result of implementing the English proficiency tests?

Teachers' behaviours were defined as what teachers do in the classroom. The aspects of teachers' behaviours in the classroom were studied in relation to the following two areas:

- teachers' medium of instruction, teacher talk, and teaching activities;
- teaching materials, and aspects of lesson planning.

As mentioned in Chapter 3, classroom observation was defined as the systematic, purposeful recording of interactions and events in the classroom (Allwright & Bailey, 1991; Chaudron, 1988; van Lier, 1988). The observations in this study accordingly involved audio-taping, taking field notes, and developing and applying a coding scheme.

7.2 Choosing the sample

The classroom observations were carried out between April and June in 2008 to compare two kinds of class: non-English majors' General English classes with the English qualification requirement, and non-English majors' General English classes without the new regulation. Efforts were made to select different types of university based on levels of student proficiency, university size and regional location, as in the survey studies reported in Chapters 5 and 6. As

suggested by other studies, the main reason behind the choice of different types of university was that it was not clear at that time where washback would occur, if it occurred at all, and whether teachers in different universities, such as universities consisting of students of higher or lower levels of English proficiency, would react differently to the new policy (see Fullan, 2007; Herman & Golan, 1991; McDonnell et al., 1990).

During the stage of survey studies (see Chapters 5 and 6), it became clear that it would not be possible to obtain a random sample of six teachers for the classroom observations. I decided to use a convenience sample by approaching survey respondents about participating. The information about this project was outlined to the participants in person or by phone. In the end, a total of four teachers gave oral consent and took part in the observation study. The teachers were from three medium-size universities located in Southern Taiwan. One of the universities was a national university and consisted of students of a higher level of English proficiency, whereas the other two were private universities and consisted of students of a lower level of English proficiency.

The four teachers shared the following characteristics:

- They were currently teaching General English to non-English majors at universities of technology in Taiwan, so they could provide relevant information related to the research topic.
- They had a positive attitude towards observations and were willing to let their teaching be observed.
- They agreed to have their lessons audio-recorded.
- They were willing to express opinions on the topic and agreed to have their interviews audio-recorded.

Efforts were made to include both male and female teachers in order to have a balanced representation, but unfortunately none of the male teachers agreed to be observed. It should be emphasised that the four female teachers were observed in this study to provide four detailed cases of English teachers in Taiwanese universities of technology for an illustrative purpose, not for the purpose of generalising to other teachers. Fortunately, the resulting four teachers

did in fact represent a range of interests and experience, even though I had to rely on a sample of convenience.

The following is a brief introduction to each of the four teachers in the study, who were given the pseudonyms Athena, Brenda, Cathy and Dana. The introduction consists of the teachers' qualifications, teaching experience, and their viewpoint on English language teaching and learning in universities. The information was provided by the teachers themselves in the interviews or by the university webpages. None of the information provided in the interviews contradicted that given on the webpages.

Athena graduated with a Master's degree in TESOL. She was the director of the Applied Linguistics Department in 2001 and had been teaching in the same university for more than ten years at the time of this study. Being a university teacher was her dream and she thought teaching suited her personality. She commented that English language teaching in Taiwanese universities of technology was very important. However, it was difficult for teachers to make much progress with improving students' English learning, as university students tended to be passive and dependent on their teachers. She also felt the students' standard of English seemed to be declining annually.

Brenda had obtained a PhD in Adult and Continuing Education in 2001 and since then she had taught English at college level. She believed that teachers should be facilitators; they should show students how to do self-study and help them realise that learning can be fun. She thought teacher development was very important; teachers needed to re-train themselves in order to provide students with what they needed. She was also not sure about the usefulness of the way that English was currently taught and whether English should or should not be taught through the medium of English. She seemed to believe that it should not be, arguing on the basis of her own success in learning English through Chinese.

Cathy graduated with a Bachelor of English degree in 1978. She worked as a teaching assistant and started teaching English as a Foreign Language (EFL) at college level in 1980. Years later, Cathy became interested in further postgraduate study and research in order to understand teaching contexts and to be able to improve them. She received a Master's degree

in TESOL in 1992 and obtained a PhD in 1999. She had been teaching in her current university for eight years. Her goal as an associate professor of English was to make English an integrated part of the students' everyday and academic experience by using interactive activities and approaches, and not simply a subject to be presented and taught.

Dana was a university graduate of English literature and had obtained a Master's degree in Family Consumer Sciences in 1997. She worked as an English teaching assistant from 1996 to 1998 and started to teach English to adults at cram schools in 1998. She moved to her current university in 2006. The objectives of her teaching were, she said, to improve students' capacity for English learning. She reported that her students' English level was very low and the students' lack of exposure to English and lack of opportunity to use the language limited their motivation and encouragement to learn the subject. She felt that English teachers needed to create more innovative teaching methods and learn how to maintain students' interest better.

7.3 Classroom observation scheme

7.3.1 Rationale behind the observation scheme

Two 50-minute lessons were observed for each teacher. Eight lessons were observed in total; four were delivered in the context of the new regulation, while the other four were delivered in a context where the regulation did not apply. Prior to the observations, discussions were conducted with the teachers to gather information regarding their university and course contexts, such as the organisation of their lessons and the teaching materials they used. The teachers were informed of the purpose of the study, but they were not asked about their attitudes toward the English qualification requirement, lest their awareness should be raised unduly and their teaching affected. During those observations, a digital audio-recorder was set up before each lesson in one corner of the classroom. The researcher sat at the back of the

room and made notes as the lessons went along. The following research assumptions were made about the washback effect of the new policy on classroom teaching. The expectation was that teachers whose students had to comply with the new regulation would (see Chapter 4):

- frequently refer to the English requirement,
- assign more practice opportunities to students,
- assign more class time to student activities such as pair work and group discussions,
- talk less,
- let the students talk more,
- use more test-oriented materials.

7.3.2 The observation scheme

The observation scheme was based on the category definitions of Part A of the Communicative Orientation of Language Teaching (COLT) (Spada & Fröhlich, 1995). The modified scheme provided for time-coding of lessons under the four categories in Table 7.1 to describe classroom activities, in order to investigate whether the lesson was student-centred or teacher-centred, how many learning opportunities were provided, and what materials teachers used in teaching, such as non-pedagogical materials, commercially written textbooks, or exam practice workbooks.

The questions asked for each column were:

1. Time: How is lesson time split into different activities as a percentage of class time?
2. Participant organisation: Who is holding the floor during the segments of the lesson as a percentage of class time?
3. Activity type: What teaching and learning is realised through various activities as a percentage of class time?
4. Content: What are the teacher and the students talking, reading and writing about, or what

are they listening to?

5. Materials used: What types of teaching material are used and for what purpose?

Table 7.1 Classroom observation scheme for the main study (sample sheet)

Time	Participant organization			Activity type	Activity content	Materials used								
	Whole		Group			Individual	Type			Purpose				
	T to S/C	S to S/C					W	A	V	P	S	N		
13:00 13:08				quiz	TOEIC related vocabulary	✓				✓				
13:08 13:25				vocabulary learning	Unit 14, page 83: Copy the meanings of vocabulary into the textbook	✓				✓				

T to S/C = teacher to students or class as a whole

S to S/C = student to students or class as a whole

W = written

A = audio

V = visual

P = pedagogical

S = semi-pedagogical

N = non-pedagogical

Part B of the COLT records the linguistic features of classroom talk, based on a tape recording of the lesson. Because the aim was to establish the extent to which the English requirement had been adopted and the effect of this on teaching, rather than to focus on the language of the classroom, COLT Part B would not in itself be informative and was thus not employed (see Chapter 4, Section 4.3.3). Instead, a modified coding system in Table 7.2, based on Brown's (1975) Interaction Analysis System (BIAS), was adopted (see

Malamah-Thomas, 1987). The system represents a simple post-hoc analysis of audio-taped recordings. The modified categories comprised:

1. Teacher: teacher describes, explains, directs and asks questions about content or procedure.
2. Student: students respond to teacher questions and directions, and give information, comments or questions.
3. Silence: pauses, short periods of silence, and periods of blackboard work without accompanying teacher or student talk.
4. Unclassifiable: listening activities without accompanying teacher or student talk.

The time-line display sheet was marked every three seconds for the duration of the observation. Once a whole lesson had been coded, percentages could be calculated for each of the categories. The purpose was to investigate how much teacher talk and student talk occurred in the classrooms.

Table 7.2 Interaction analysis system for the main study

Teacher																			
Student																			
Silence																			
Unclassifiable																			

Each box represents a three-second time interval.

7.3.3 Observation scale

Data were analysed in terms of the four categories of the classroom observation scheme. The observation scheme required detailed notes to be made in real time on the activities that occurred during the lesson, including the time taken for each activity. The start time of each activity was recorded to the nearest second. The notes also illustrated any differences among

the lessons at the level of practice opportunities, in terms of how the lessons were segmented regarding interaction patterns, the nature of the teacher talk, activity types and content, and the use of teaching materials. Immediately after each observation, while my memory was still fresh, I completed the notes by listening to the audio recording and I wrote my impressions of the lesson in the margins. Subsequently the duration of each activity and the time teachers and students spent on each of the categories were calculated as a percentage of the total lesson time (length of lessons minus breaks). During the observation, it became clear that several activities, such as times when the teacher gave the students information about tests or discussed test-taking strategies, were not specifically identified by COLT. These were recorded and analysed separately. These analyses were followed by an analysis of the classroom interaction patterns and in-depth interviews with the four teachers, which enabled me to explore in detail the intentions behind a variety of classroom activities. Both pre-observation discussions and follow-up interviews were conducted in Mandarin, the transcriptions of which were subsequently summarised in English.

7.4 Classroom observation outcomes

The following section reports the findings from the four teachers in the main study. This section discusses whether there were any noticeable differences in classroom teaching among the teachers in relation to the implementation of English proficiency tests. The percentage of time spent on each of the categories under COLT's major features for Group 1 and Group 2 was compared.

7.4.1 Participant organisation

Table 7.3 summarised the interaction patterns of the lessons with Athena, Brenda, Cathy and

Dana teaching General English to non-English majors. Two aspects were observed: (a) Is the teacher working with the whole class or not? (b) Are students divided into groups or are they engaged in individual work? They were represented in the table as:

- T to Student/Class (teacher to students or class as a whole, e.g. explaining),
- S to Student/Class (student to students or class as a whole, e.g. oral presentations),
- Group work (students are working on a certain task in groups of two or more),
- Individual (individual work, e.g. on exercises or listening).

Table 7.3 Participant organisation of lessons

Context	Name	Percentage of lesson time*			Total	
		Whole class		Group		Individual
Group 1 With the regulation		T to S/C	S to S/C	0%	29%	100%
	Athena	71%	0%			
Brenda	100%	0%	0%	0%	0%	100%
Group 2 Without the regulation		T to S/C	S to S/C	0%	5%	100%
	Cathy	80%	15%			
Dana	59%	0%	41%	0%	0%	100%

T to S/C = teacher to students or class as a whole

S to S/C = students to students or class as a whole

* Percentages are averaged per person and rounded off to the nearest integer.

The findings from Athena's lessons in the group in which students had to take English proficiency tests as a graduation requirement made it clear that Athena (71%) spent the majority of class time talking to the whole class. No student talk or group work was observed in either lesson. The remaining 29% of the class time was spent on individual work. The rationale behind spending a considerable amount of class time on individual work, according to detailed observations and based on interviews with Athena, was to allow a focus on reading and vocabulary learning in class. Athena explained in the interview that because TOEIC tested only students' listening and reading skills, she mainly focused on teaching those skills

in her lessons. She believed that students' English ability and exam scores would be improved in a short space of time if they practised reading skills and vocabulary activities regularly in class.

Also teaching Group 1 students, but at a different university, Brenda (100%) had a total control of the lessons. No student talk, group work or individual work was observed in her lessons. Brenda explained that because her students' English level was quite low, she had to direct her students' learning and explain things in detail.

In Group 2, the findings from Cathy's lessons showed that she talked to the whole class the majority of the lesson time (80%). Just 15% of class time was devoted to student talk in which student(s) spoke to the whole class. It should be noted that this category of activity was treated differently from group work, in which students were involved in discussion (multi-level/party communication) or in conducting a group task. No group work was observed and only 5% of class time was spent on individual work.

Dana held the floor for 59% of the class time. However, unlike the other teachers, she spent 41% of her class time on group work. She explained that the purpose was to provide students with plenty of practice opportunities. She believed that group work helped students learn from each other and gave them more practice opportunities, especially oral practice.

In short, neither student talk nor group work were observed in the Group 1 teachers' lessons, even though the new policy might be expected to have encouraged a more learner-centred classroom. Although student talk and group work were seen in the Group 2 teachers' lessons, the proportion of the time spent varied from teacher to teacher. Moreover, the nature of teacher talk, as a percentage of class time, was similar in both groups. It seemed that the teachers still dominated most of the classroom interaction and that teacher talk was the predominant strategy in their classrooms.

7.4.2 Activity types and content as percentage of class time

The purpose of looking at activity type in classroom teaching was to explore what kinds of teaching and learning were realised through various activities. By investigating the content of the activities carried out in the classroom, the researcher can explore the subject matter of the activities – what the teachers and the students were talking, reading, or writing about, or what they were listening to.

Activity types were grouped into teacher activities and student activities. Findings relating to the content were again reported as a percentage of class time. The analysis of the eight lessons with the four teachers in Table 7.4 showed (a) what types of activity were carried out in the lessons and how lessons were segmented according to the percentage of time devoted to them by the four teachers, and (b) who was holding the floor and in what ways.

Table 7.4 Classroom activities of all four teachers' lessons as a percentage of class time*

Activity type	Group 1		Group 2	
	Athena	Brenda	Cathy	Dana
<i>Teacher activities:</i>				
Pre-lesson activities	4	10	18	2
Lecturing, explaining, and directing	67	85	62	57
<i>Student activities:</i>				
Individual work: Listening	0	5	5	0
Individual work: Reading	4	0	0	0
Individual work: Vocabulary	25	0	0	0
Group work	0	0	0	41
Oral work	0	0	15	0
Total	100	100	100	100

* Percentages are averaged per person and rounded off to the nearest integer.

It can be seen that the teacher was still the predominant focus of the classes in the two groups, but the activity types and contents were varied. 71% of the lesson time was spent on

teacher activities in Athena's lessons. 18% of Athena's activity time was in fact spent on testing students' TOEIC vocabulary knowledge and explaining the meaning of the vocabulary. The remaining 53% was spent on explaining vocabulary, reading text, and reading comprehension exercises in the General English textbook. Similarly, Brenda spent 95% of her lesson time on teacher activities, but focused on explaining phonics and vocabulary. Cathy also spent the majority of the lesson time (80%) on teacher activities in which she directed discussions. Dana spent just 59% of the lesson time on explaining vocabulary and grammar.

One of the research focuses was to explore whether, in the light of the new policy, teachers would provide students with more practice opportunities and fewer teacher-dominant activities. When Athena and Brenda were asked the reason for there being a high proportion of teacher activities in their classes, they both explained that it was because of the students' low English level. As their students' English was at pre-intermediate level, they had to explain more and give detailed explanations, in order to make sure that students could later meet the English proficiency test requirements.

The findings from Athena's and Brenda's lessons showed that reading and vocabulary learning were the main student activities in Athena's lessons and listening was the main student activity in Brenda's lessons. Most of the time students carried out those activities individually, rather than interactively with the teachers or their peers. In addition, those types of activity were usually related to a comprehension exercise with multiple-choice responses or merely consisted of listening to CDs. The results suggested that the activities were not interactive and little was achieved in terms of providing learning opportunities for the students. Moreover, there was an imbalance in activity type, in that students developed only their receptive skills of reading and listening, and not the productive skills of speaking and writing.

As with Athena's and Brenda's groups, Cathy's students spent 5% of the lesson time on listening to the reading text. However, 15% of the lesson time was devoted to oral work, in which students gave a short presentation to other students. Unlike in the other teachers' classrooms, student activities in Dana's classrooms comprised 41% of the lesson time spent on group work, in which students were divided into four groups and competed in

grammar-related and vocabulary-based language games. It could be seen that Cathy's and Dana's lessons were more interactive than Athena's and Brenda's lessons in terms of student activities. Nevertheless, there was also an imbalance in activity type.

A cross-comparison of the percentage of time spent on student activities in the lessons observed showed that there were substantial differences; the total was 29% for Athena, 5% for Brenda, 20% for Cathy, and 41% for Dana. More importantly perhaps, it can be seen that overall Athena's and Brenda's lessons were not interactive and their lessons were more teacher-controlled compared with Cathy's and Dana's lessons, suggesting that the Group 1 teachers, whose students had to meet the graduation requirement, did not provide more practice opportunities.

7.4.3 Teaching materials

This section summarises the findings related to the use of teaching materials in classroom teaching. It was found that Brenda and Dana used university-developed textbooks while Athena and Cathy used commercially-written textbooks, all occasionally accompanied by audio-visual materials, specifically CDs for listening. Besides the pedagogical General English textbooks, Athena also used a commercially-written and test-oriented textbook, she said, due to the implementation of TOEIC by the university.

In summary, the types of teaching materials for both groups were not substantially different. They were all pedagogical General English textbooks. There was perhaps an impact of English proficiency tests on teaching materials in the case of Athena, who used test-oriented materials.

7.4.4 Further analysis

In addition to the categories included in the modified COLT instrument, one other phenomenon was observed during the lessons. With regard to the instances and ways of mentioning English proficiency tests, Athena referred to TOEIC twice in her lessons. She provided the students with factual information about the test and reminded them that there would be a mock exam in week 13 and an official test in week 15 of the semester. On the other hand, the other three teachers did not mention any English proficiency test to their students in the eight lessons observed. This finding indicated that the use of a public test did have some influence on one of the teachers.

7.5 Classroom observations of individual teachers' lessons and interviews

This section reports findings from further analysis of each teacher's first lesson and follow-up interviews with the teachers. Because one purpose of the new policy was to improve university students' communication skills, implying more student talk in the classrooms, it was necessary to explore a sample lesson in detail to see whether there were any discernible washback effects. Fullan and Park (1981: 7-8) highlighted the importance of the belief systems of those affected by educational change, and in line with this, it was felt important here to explore the teachers' beliefs: whether teachers believed that their teaching had been influenced by the introduction of English proficiency tests; and whether they reported changes to their teaching practices as a result of the implementation of the tests. The follow-up interviews also provided an opportunity for the teachers to give their impressions of the lessons and to describe the rationale behind their choices of activities and materials. The interviews were semi-structured with prompts whenever necessary (see Appendix 4). The interview questions, including class-based questions depending on observations, were the following:

1. What do you think are the major reasons why the department has made passing English proficiency tests a graduation requirement?
2. What do you think about making passing English proficiency tests a graduation requirement?
3. Does this new policy influence your teaching?
4. (If 'yes' to 3.) How does it influence you?
5. (If 'yes' to 3.) How do you feel about test-influence on teaching?
6. Do you think the regulation has increased or decreased students' motivation to learn? Why? How?
7. Do you think this regulation should be implemented or rejected? Why?
8. Would you like to add anything concerning the influences of English proficiency tests?

These questions were based on three washback hypotheses taken from Alderson and Wall (1993: 120-121): "A test will influence *what* teachers teach"; "A test will influence *how* teachers teach"; and "Tests will have washback effects for ... *some* teachers, but not for others". There were also parallels between questions in the questionnaires and interviews (see Chapter 4). The questions were piloted (two interviews were held) in June 2007 to obtain information regarding the relevance and clarity of the questions, the format, and the amount of time required to answer the questions. They were revised accordingly: several questions were formulated more clearly as discussed in Chapter 4.

All the interviews lasted about 20 to 30 minutes, were conducted in Mandarin and recorded with a digital audio-recorder and I also took notes throughout the interviews. The data were transcribed in full in the original language and then translated into English. The transcripts were edited, by eliminating repetitions and putting substantive statements in chronological order to make grammatical sense. They were then analysed into major themes (see Chapter 4, Section 4.4).

Table 7.5 Summary of the observations of the four teachers

	Group 1		Group 2	
	Athena	Brenda	Cathy	Dana
Teacher talk	58%	69%	66%	73%
Student talk	3%	18%	13%	15%
Silence	39%	8%	11%	12%
Unclassifiable	0%	5%	10%	0%

Every teacher's first lesson was observed and analysed. It can be seen from Table 7.5 that, during the 50 minute lesson, Athena talked for 58% of the lesson time. The teacher-talk time was spent on explaining reading texts and vocabulary, and commenting on language points and sentence structures. Students talked for 3% of the lesson time and this involved making occasional responses. 39% of the lesson time was silent, when students carried out individual work on reading, reading comprehension exercises and test-oriented vocabulary quizzes. The teaching materials that Athena used came from both a commercially-written General English textbook and an exam practice workbook.

When asked about whether the implementation of English proficiency tests as a graduation requirement had brought about changes in her teaching, Athena said:

Yes, it has. I think the biggest impact is on teaching materials. I usually use a commercially-written General English textbook and teach based on it. This semester, besides the textbook, I have started using a TOEIC practice workbook. I spend more lesson time on using the TOEIC practice book than the General English textbook because students will take a school-wide TOEIC exam in May. If the students don't pass the test, their General English module in this semester is counted as a fail. My teaching is now based on activities and the test format in the TOEIC practice workbook and mainly focuses on listening and reading. I think it is unreasonable to ask students to take exams without helping them prepare for them. Besides, students won't know how to prepare for the test if I don't teach them how to.

Athena later commented on why the changes happened in her teaching due to the new policy:

I think teachers are under a lot of pressure. I don't know about other teachers, but I'm under a lot of pressure to help students prepare for the TOEIC. In my university they won't acknowledge that, but it is there. Teachers have to work for the exam.

She also addressed concerns and opinions on the English graduation requirement:

This is the first semester that we are asking students to take a school-wide TOEIC exam and it is counted as the final exam. I think it is all right to teach towards the test for one semester. But it is inappropriate to do it on a long-term basis, otherwise universities will become cram schools. Although I think it is good to set up a graduation requirement in which students know the university's expectations and goals, most students' English level is quite low and they are not able to pass English proficiency tests and meet the graduation requirement. I have found that more than half of my students in this class have given up learning English and about one-third of the students in this class haven't shown up this semester.

Athena concluded that:

Setting up an English graduation threshold is a trend among universities, so they can gain prestige in the nationwide university evaluation. I think the regulation should be carried on, but the university should not function as a cram school and will need to persuade students to accept the regulation. So far, most students have a positive attitude towards the regulation, but others are waiting for a remedy for failing in tests and thus they do not study harder as expected.

A review of Athena's 100-minute classroom observations and interview proved emphatically that Athena taught for English proficiency tests, in her case, TOEIC. Observations showed that her course became focused on equipping students with specific English abilities in terms of the teaching content, the tests she administered, and her teacher talk. The teaching content was adopted from a commercially-written TOEIC practice book. Quizzes and the final examination were also relevant to the TOEIC. Athena's teacher talk pertained to the TOEIC in terms of TOEIC-pertinent information, such as the deadline and

venue of the registration. There was thus no doubt that Athena did in fact prepare students for the test. The results of Athena's interview were congruent with the classroom observations. It was noted that although Athena agreed with the graduation regulation, she expressed the concern that a handful of students were sometimes absent or late for classes, suggesting a low level of student motivation for learning English due to the high-stakes of testing.

As for Brenda's first lesson, it was discovered that, during the 50 minute period, she talked for 69% of the time, explaining and discussing vocabulary and phonics. 5% of the lesson was devoted to individual student work on listening activities and 18% involved students' occasional responses. The main pattern of the lesson was teacher explanation → student activities → teacher comment and further explanation. The teaching material used was confined to a university-developed textbook.

From a comparison of Athena's and Brenda's lessons, it can be seen that there was more time spent on teacher talk by Brenda than by Athena, which was different from the assumption that there would be constantly less teacher talk and more student practice opportunities. In the interview, Brenda explained that:

It is because the university-developed textbook is very structured. The contents of the textbook focus mainly on phonics, vocabulary and grammar and they are tested school-wide in the mid-term and final exams. I have to teach based on the textbook otherwise students will fail in the exams. Besides, students' English level is quite low, so I have to teach them basic skills. I feel my teaching is limited by the textbook. The lessons become less interactive and more like lecturing.

To Brenda, if a class of students had a high level of proficiency overall, there would be no need for the teacher to talk and explain much, and vice versa.

When asked whether she had taught differently since the implementation of the new policy, Brenda replied:

No, I have not taught differently. I use the textbook which is assigned by the university. I focus on teaching vocabulary and grammar and employ the same

teaching methods I used before. Besides, there is not enough classroom time for me to teach towards English proficiency tests. But I think ultimately the new policy will change teaching. However, whether the change will make English learning at college level better or not is open to question.

Brenda finally commented that:

When I began teaching in this university, the regulation had already been imposed. I don't know why the university set up the requirement, but I have found that students seem to care about their English learning. I think the regulation should be carried on under certain conditions. General English should be a required module for students to take for all four years at university. Teachers' teaching should be more interactive, otherwise students will be demotivated as regards English learning.

In striking contrast to Athena's lessons, Brenda's lessons lacked any direct connection to any English proficiency tests. The textbook she adopted was not specifically relevant to any English proficiency tests. Moreover, she did not offer any English proficiency test-relevant information or test-taking skills in her course. It was clear that Brenda did not prepare students for English proficiency tests in her course. Consequently, the English graduation requirement had no impact on her course. The results were evidenced by Brenda's interview. She, nevertheless, agreed that the graduation regulation should be carried on under certain conditions, such as extending compulsory English study years and improving teachers' teaching.

In Cathy's 50-minute lesson, her talking time represented 66% of the lesson. Student talk took up 13% of the lesson time and consisted of responding to the teacher's questions. 10% of the lesson time was spent on individual work in which students listened to CDs. Although teacher-dominated talk occupied most of the lesson time, students received more opportunities to talk in a more creative manner than did students in Athena's or Brenda's classes. Examples of interactions were observed where Cathy initiated questions and students answered them. In the interview, Cathy explained:

I encourage critical thinking and would like to allow more language in the classrooms, more spoken language. That way, students can allow the language to become a part of themselves. They internalise it. I want to prepare the students for the real world.

Later while asking about whether she would change her teaching methodology when the graduation requirement was imposed, Cathy commented that:

No, I won't. I try to keep a balance in developing students' four skills now and I will do the same whether there is an English graduation requirement or not. I think students should develop the four skills rather than listening and reading skills alone.

Cathy also shared her opinions on setting up the English graduation threshold:

I strongly agree with the university setting up the graduation threshold for the coming year. For now, students' English ability is too low to be at college level. They don't care about their English learning. Although high English proficiency test scores don't mean high proficiency, I think with the graduation threshold, students will study harder and finally obtain college level English performance.

In short, Cathy's courses offered students listening and speaking opportunities overall, but did not prepare students specifically for any English proficiency tests. The teaching materials that she adopted did not pertain to any English proficiency tests. In her lessons, she did not mention English proficiency tests or coach students in test-taking strategies. All evidence demonstrated that Cathy's course was not geared to preparing students for English proficiency tests. The results of the observations were congruent with the interview. Therefore, it can be concluded that Cathy did not prepare students for English proficiency tests in her General English course and the lack of English proficiency test-relevant teaching was because the English graduation requirement had not been established in Cathy's university. However, Cathy mentioned in the interview that she strongly agreed with the policy of establishing a graduation threshold, which would be introduced for the coming year, even though she would

not change her teaching practices.

As for Dana's first lesson, she talked for 73% of the lesson time. Most of that time was spent on explaining vocabulary. Students talked for 15% of the lesson time while carrying out group work with vocabulary games. The main interaction pattern of the lesson was teacher explanation followed by students' short responses. The atmosphere was lively, as students spent plenty of the time working in groups. Their attention was focused on the tasks they were doing. The teacher directed the vocabulary games, went around the classroom giving individual comments and then periodically commented to the whole class.

When asked about the rationale behind her choices of activities and materials, Dana explained that:

The textbook is designed and provided by the university. The course content is prescribed, focusing on vocabulary and grammar, and it is tested in the mid-term and final exams. Therefore, my teaching is based on and limited by the textbook. There is not much flexibility and not enough class time.

Dana also commented on the English requirement and the General English syllabus in the university:

Most of the universities of technology have imposed the regulation and I think the regulation gives students learning goals, so I think the university *should* set up the graduation threshold. It will make students study harder and finally develop basic English ability. In addition, I think the university should adopt streaming rather than mixed-ability grouping. Otherwise, students with a high English level will lack motivation for learning English.

From Dana's 100-minute lessons it was considered that Dana did not teach to English proficiency tests, because the teaching content was irrelevant to any specific English proficiency tests. English proficiency test-taking strategies were not coached. In addition, test-relevant information was not mentioned. The conclusion was congruent with Dana's views as expressed in her interview. As with Cathy, Dana also agreed that the university

should adopt the graduation threshold, but relevant policies on motivating low-proficiency students without demotivating high-proficiency students should be considered and planned carefully.

In conclusion, all four teachers talked for some time in their first lesson and students talked for only a relatively short period of time, suggesting that teachers dominated most of the classroom interaction and students received minimal opportunity for oral communication practice. Furthermore, because the Taiwanese Ministry of Education has announced the adoption of the English proficiency test pass rate as one indicator of the effectiveness of universities' performance (see Chapter 2), the policy of the English graduation requirement will play a critical role in determining the universities' policies in the future. Overall, the teachers interviewed had a positive attitude towards the graduation requirement. They believed that passing English proficiency tests would benefit students in their work, academic performance and subsequent English studies. However, they expressed concern that students' proficiency levels were too low to enable them to pass the relevant English proficiency tests and meet the graduation requirement, and over the effectiveness of implementing the graduation regulation as a means of motivating students' learning and improving their English ability.

7.6 Summary

This chapter has focused on the findings from observing teachers and students in classroom settings and from talking to the teachers about their lessons and their opinions on the graduation requirement. It reports General English lessons carried out by four teachers, two of whom were teaching under the new policy and the other two not. At a general level, the lessons were roughly similar across the four teachers. The activities mainly emphasised reading skills, vocabulary building, language points and sentence structures, which limited students to producing isolated sentences, and limited them to be assessed merely for

grammatical accuracy.

The patterns of teacher interaction in the classroom did not reveal many differences either. All eight lessons were highly teacher-controlled. In both groups, teacher talk dominated for the majority of the lesson time. Nevertheless, there were some differences. There were differences among teachers in the allocation of time in teacher and student talk. In Athena's lessons, there was less teacher talk than in the other teachers' lessons. Brenda's lessons were much more teacher-controlled and consequently showed the least amount of student activity.

The observations revealed that the approaches to teaching did not vary greatly. All four teachers relied on General English textbooks, so they used activities suggested by the textbook writers. Only Athena used test-oriented materials, but she tended to rely on the practice books for the English proficiency tests. As a result, Athena, too, used activities suggested by the practice book writers. Paradoxically, there were more opportunities for student activities in Group 2 classrooms than in Group 1 ones. The Group 2 teachers provided students with much more time for activities such as group work and oral presentations. As a result, class time spent on individual student work was less than in Group 1 classrooms.

All in all, it can be seen that the general pattern of teaching approaches did not differ much between the teachers. It seems clear that, although the decision to implement English proficiency tests was expected to persuade teachers to change their classroom activities, given the importance of the examinations, it did not change them in their fundamental beliefs and attitudes about teaching and learning, about the roles of teachers and students, or about how teaching and learning should be carried out. However, there was a substantial variation in teachers' beliefs about what to teach and what students should learn. The teachers also varied markedly in their classroom practices. A new policy involving the introduction of new exams to an existing educational system can, to a large extent, change the content of teaching and even the way activities are carried out, but overall, very little change in the interaction patterns between teachers and students could be found in this study, and might not be found within the initial couple of years of the implementation of any new policy.

The detailed findings suggested that there was a minor washback effect on teaching

brought about by the implementation of English proficiency tests. Teacher talk as a percentage of class time was still the dominant activity in the classrooms where teachers taught under the new policy. According to teacher interviews, change in classroom teaching had occurred for a few teachers, but not for all, and to different degrees. It also appeared that this change might differ with the passage of time, as Fullilove (1992: 131) commented “the nature and strength of this washback effect and the benefits or disadvantages of the washback depend in large measure upon the educational system as a whole, upon the nature of the other participants involved in writing or establishing the competing curricula, and, of course, upon the types of examination in question”. Noble and Smith (1994b) suggested, in addition, that conceptual changes were not simply brought about by the acquisition of new ideas. They were seldom achieved without attending to the beliefs of those who were the targets of change: teachers, teaching and learning contexts concerned, and students, their levels of proficiency concerned. As a result, efforts at educational innovation directed towards instructional improvement should acknowledge the challenges presented by such conceptual changes.

Chapter Eight: Advisory Committee Members' and Departmental Directors' Reactions

8.1 Introduction

The other stage at Phase III of the study, lasting between April and June 2008, consisted of an investigation of the Taiwanese educational context at both the macro and micro levels. The sample for this stage and the research methods employed were (a) document studies, (b) interviews with three advisory committee members at the macro level of the Taiwanese educational context, and (c) at the micro level, interviews with three departmental directors in Taiwanese universities of technology. This phase of the study was designed to answer the first research question in Chapter 4: what strategies did the Taiwanese Ministry of Education (MOE) use to implement the project on The Enhancement of Students' Foreign Language Proficiency? It was split into sub-questions about the rationale and strategies used by the MOE and reactions of the advisory committee members and departmental directors in relation to the imposition of English proficiency tests.

In this chapter, the general intentions and expectations of the committee members are discussed first, followed by a discussion of the departmental directors' perceptions and reactions. A comparison is made between the two to identify possible matches and mismatches.

8.2 The Taiwanese educational context at the macro level

During the past few years, technological and vocational education in Taiwan has been substantially developed. The Department of Technological and Vocational Education (TVE)

is the central authority within the Ministry of Education. It is in charge of national TVE and has supervisory control of colleges and universities of technology, national and private junior colleges of technology, and national and private vocational high schools. Within the TVE department, the International Cooperation and Exchange Advisory Committee consists of six divisions which are in charge of six different disciplines, namely agriculture and maritime technology, commerce and management, home economics, industry, nursing and medical technology, and arts and humanities. The arts and humanities division consists of three to five members who are selected from university principals and provosts, executives who run subject-related governmental bodies, and external bodies involved in commerce and industry. The advisory committee members have a duty to develop and formulate policies and to advise on and evaluate projects.

In 2002, the advisory committee members of the TVE developed and formulated the Ministry of Education Grant Project on the Enhancement of Students' Foreign Language Proficiency. The purpose of the grant project was to improve university students' foreign language proficiency through diverse activities. Its key components were (Ministry of Education, 2005b):

- to implement foreign language examinations and remedial language instruction in universities;
- to create multimedia teaching programmes and an excellent learning environment;
- to run various English camps and related activities;
- to develop and adapt new English or other foreign language teaching approaches;
- to promote professional English teaching programmes and experiments;
- to cooperate with international universities and invite foreign instructors to Taiwan;
- to develop courses delivered in English and supporting programmes to attract international students.

On the basis of the components, each university in Taiwan submitted one proposal. Approved organisational projects were evaluated twice a year in terms of two criteria: implementation schedule and performance. Each university had to fill in a self-evaluated

performance form and submit it to the TVE department. The advisory committee of the TVE department then selected organisational projects randomly for further inspection. In 2005, three English teaching and learning resource centres were also built in northern, central, and southern Taiwan providing further support for universities.

8.2.1 Interviews with advisory committee members

This section reports findings from the interviews with advisory committee members. The interviews were to investigate the decision-making process in relation to the government grant project. They explored expectations and attitudes of stakeholders at the macro level of the Taiwanese educational context.

Semi-structured telephone interviews were conducted with three advisory committee members between April and June in 2008. One of the three members was currently involved in the management of a university of technology, and the other two interviewees had previous experience of it. The information about this project was outlined to the committee members by phone in advance, but detailed interview questions were not given at that time to avoid the occurrence of prepared answers. Oral consent was obtained from all the participants prior to the interviews. To maintain confidentiality, the informants were given the pseudonyms Michael, Sherry, and Patrick. Being proficient in Mandarin, Taiwanese and English, I allowed the participants to choose any of the three languages for their interviews. All participants chose Mandarin. Each committee member was interviewed once for between 20 and 30 minutes and was encouraged to elaborate on their answers as the aim was to generate in-depth information. All of the interviews were recorded with a digital audio-recorder and I also took notes throughout the interviews. The data were all transcribed in full in the original language and then translated into English. The transcripts were edited, by eliminating repetitions and putting substantive statements in chronological order to make grammatical sense, and they were then analysed into major themes. The interview schedule, translated into English, can be

summarised as follows (see Appendix 4.3.2):

1. Why did the Ministry of Education decide to introduce the project on The Enhancement of Students' Foreign Language Proficiency in universities of technology?
2. Why did the Ministry of Education decide to include the implementation of English proficiency tests in the project?
3. What do you think are the reasons why some universities of technology have brought in passing English proficiency tests as a graduation requirement for non-English majors, while others have not?
4. What do you think about imposing passing English proficiency tests as a graduation requirement for non-English majors?
5. Do you think this graduation requirement should be implemented or rejected? Why?
6. Would you like to add anything concerning the regulation for non-English majors?

The questions were developed on the basis of issues that were being discussed among departmental directors and teachers, which were explored at the preliminary information-gathering stage, and there were parallels between questions in the questionnaires and the interviews (see Chapter 4). The questions were piloted (one interview was held) in June 2007 to obtain information regarding the relevance and clarity of the questions, the format, and the amount of time required to answer the questions. They were revised accordingly: several questions were formulated more clearly as discussed in Chapter 4. The advisory committee members were asked to share their expectations, to identify issues that were likely to impact on universities, and to expand on the role they thought universities could play in the Ministry of Education grant project. They were also asked about the role of English in higher education, the English curriculum, and the English language proficiency achievements of students. Follow-up contacts were made with the informants, either by phone or e-mail during and after data collection, to invite them to clarify unclear points and confirm or disconfirm some findings.

In the interviews, all three advisory committee members showed great understanding of the purposes of introducing the government grant project to universities of technology. Patrick explained that:

In order to be more competitive in global markets, the government wants to develop the nation's proficiency in English, because English is an essential worldwide communication tool. The government thinks the lack of English proficiency is problematic. Therefore, the government grant project was developed.

Michael added that:

In accordance with the national development plan Challenge 2008, the Ministry of Education formulated the Enhancement of Students' Foreign Language Proficiency Project. There are several components to the project and universities can develop their organisational projects based on at least one of the components.

The committee members also commented that although universities could develop organisational projects based on at least one of the grant project components, the implementation of English proficiency tests was an intended outcome.

The Ministry of Education announced in public that 30% of university of technology students would pass English proficiency tests by 2006, 40% by 2007, and 50% by 2008. Keeping this goal in mind, I intend to approve organizational projects which include the implementation of English proficiency tests and give more grants to the university which had more students passing English proficiency tests in the previous year. Nevertheless, it is not stated clearly in official documents that implementing English proficiency tests is a required component in organisational projects. (Patrick)

Patrick's comments emphasised the prestige of English proficiency tests in the grant project. However, the Ministry of Education did not state its intentions explicitly in official documents.

When asked about the reasons why the graduation requirement was not compulsory for all universities of technology, all the committee members interviewed said that the decisions were up to the university itself. Michael explained that:

Each university has its particular concerns. For example, private universities may have difficulties attracting students to apply for the university if they state officially that non-English major students have to pass an English proficiency test as a graduation requirement. In addition, some national universities do not impose the regulation because they did not state it in any official documents while students were applying for the university. The universities worry about the fact that it would be illegal if they asked students to take the tests afterwards. Moreover, some teachers think that other foreign languages are also important. Why ask students to take English proficiency tests as a graduation requirement rather than other languages? No individual principal or director can say firmly that English is the most important foreign language, so the English graduation requirement is not enforced.

In addition to Michael's explanation, Patrick indicated that:

With the graduation threshold, students may not want to apply for and enter the universities. Some universities worry about students' low English proficiency; most of their students will not be able to pass English proficiency tests at the end of their courses. The universities also worry about low test pass rates and then merely receiving a small government grant. Therefore, the universities do not impose the regulation.

It could be seen that because of the government's liberality, universities had considerable freedom to develop their institutions and make their own policies. In addition, considering the legality of and the controversy over the graduation regulation, the Taiwanese Ministry of Education and universities of technology have not made it compulsory for all university students.

On the question of the strengths and weaknesses of imposing the English graduation requirement, the three reacted as follows:

With the English language requirement, students seem to study harder and finally learn the language. However, the English curriculum in universities has been moving in a negative direction since the regulation was imposed. The most likely consequence is that universities offer test preparation courses which were provided only by cram schools in the past. In the General English courses, some teachers spend more time on teaching to tests nowadays. Universities seem to function as cram schools and teachers' teaching has been diverted away from normality. Moreover, more students spend lots of money on cram schools to prepare for English proficiency tests, which causes a financial burden for them and boosts cram school business. It is not right and I believe it is not the original intention of setting up the graduation regulation by the universities. It is certainly not the intention of the governmental grant project! (Michael)

If we look at the universities of technology which have been implementing the regulation for six or seven years, we can see that students' test pass rates have increased and their English proficiency has improved. However, teachers are teaching to the tests. The normal pattern of the English curriculum and teaching has been changed. (Patrick)

Because most students cannot pass English proficiency tests, universities have to provide a window of opportunity. Students who have not passed any English proficiency test can take and pass a make-up test and finally graduate. Imposing the graduation regulation has not so far had the desired effect. Moreover, the priority for English teaching and learning at universities is now passing English proficiency tests. The original purpose of English education has changed. It is not right. (Sherry)

From their comments, it seemed that the advisory committee members believed that the graduation regulation motivated students to learn English and that their English ability had improved. The committee members, however, were also aware of the difficulties concerning, and problems with, implementing the graduation requirement, such as the narrowing of the English curriculum, the creation of cram-school-like teaching, the financial burden on students, the growth of private cram school business, and the ineffectiveness of the new regulation. Nevertheless, they all had, to a degree, a positive attitude towards the graduation requirement and were keen for the regulation to be carried out under certain conditions.

Michael explained that:

English is a required course in liberal arts education in technological and vocational education. I think the graduation regulation could be carried on if the General English teaching curriculum followed a normal pattern. Cram-type learning should not be encouraged.

Sherry commented that “the regulation could be carried on, but the benchmark should be set reasonably.” Patrick said that:

The regulation should be carried on, but there needs to be a set of supporting proposals. For example, universities can provide make-up tests or courses for those students who have not passed external English proficiency tests as a window of opportunity for meeting the graduation requirement.

Finally, the three interviewees gave additional comments on issues related to the implementation of the graduation requirement.

If you read annual evaluation reports numerically, the graduation requirement seems to be effective. However, students may not have attained a high level of proficiency in English, even though they have passed the tests. There are no specific criteria to assess teachers’ teaching when evaluating organisational projects, but I hope that teachers adopt a new teaching approach which is more learner-centred, select materials which motivate students to learn, and use more English in their teaching. Teachers have not been evaluated based on students’ test performances, but I hope that teachers still give the graduation requirement serious attention. (Michael)

Because of academic freedom there are no specific regulations in relation to the English requirement for English curricula and teachers in universities. Teachers are not evaluated based on students’ English proficiency test performances. In addition, some universities provide a window of opportunity for the students who fail the tests, so the implementation of the graduation requirement has not been effective thus far. Those remedial actions seem to impede the progress of policy diffusion. Moreover, the advisory committee members evaluate organisational projects mainly on the basis of universities’ self-evaluated

reports with a select few singled out for official visits. The reports could be untrue because of the pressure to meet grant project requirements. These issues should be given in-depth discussion. (Sherry)

It seems to me that students' English ability has improved. Without the graduation requirement, students lack the ability to compete in job markets and to face international competition. As regards teachers, there is no specific requirement regarding the implementation of the graduation requirement. Teachers think the regulation promotes students' motivation to learn English, although their workload has increased. (Patrick)

Michael, Patrick and Sherry all mentioned that there was no specific regulation of teachers with regard to the implementation of the English graduation requirement. They, however, expected an intended washback effect on English teaching and learning when they evaluated organisational projects, although the washback in specific areas of teaching and learning was not commented on explicitly. It was difficult to tell from the comments whether the advisory committee members thought it was necessary or possible to enforce regulations, to force university teachers to give weight to the new policy. In addition, the supporting proposal of administering make-up tests and courses and the procedure for evaluating organisational projects were all in question, suggesting the new policy was still at the initial phase of diffusion and implementation. It would take a long time to see the effectiveness or ineffectiveness of the new policy.

These findings showed how decisions about the government grant project were made at the macro level and what advisory committee members' expectations and attitudes were towards the English graduation requirement in universities of technology. However, intended washback effects, if any, were not clear in official documents and not many comments were made in the interviews about washback in the specific areas of teaching and learning.

8.3 The Taiwanese educational context at the micro level

No matter how positive the washback effect of the English graduation requirement was perceived to be at the macro level, the actual changes had to take place in the university settings and be implemented in a teaching and learning context. The following reports on the findings within the universities of technology.

8.3.1 Tertiary institutions

The new graduation regulation – passing English proficiency tests as a graduation requirement – for both English and non-English majors has been implemented in some universities of technology since September 2002. The first cohort of students affected by the regulation were the freshmen in the academic year 2002. They had to take and pass one of the following English proficiency tests, GEPT, TOEIC, TOEFL, IELTS, CSEPT or a university-developed test, by June 2006 prior to their graduation.

After the graduation regulation was announced, a series of seminars and workshops was organised to help teachers implement the new policy:

- English Teaching/Learning Resources Centre seminars and workshops. These were organised with the aim of providing teachers with information about English proficiency tests, such as the rationale and content of the tests and the test formats. Sample papers and test formats were introduced and demonstrated.
- Seminars were organised by tertiary institutions. The major aim of these seminars was to give teachers information about the organisational project, such as the content, activities and the timeline, as well as about English proficiency tests employed and the cut-off score of each test. Issues raised by teachers, such as their worries about students' low English proficiency and the success rate, were also discussed. There were also lengthy discussions among the teachers as to how teaching and learning

would be adjusted or modified according to the requirements of the organisational project or English proficiency tests.

- Textbook publishers' seminars and workshops. Some colleges and universities invited textbook publishers to provide seminars and workshops. Textbook publishers provided their understanding of different kinds of English proficiency test and showed their test-oriented textbooks. They also demonstrated to the teachers how the English teaching and learning activities and tasks in their textbooks could be carried out in the classroom.

These seminars and workshops generally attracted a large teacher audience¹. One reason for the teachers and departmental directors attending the seminars and workshops was to explore teaching support in terms of teaching content and materials support, which suggested that teachers relied on textbooks for direct support when there was a potential curriculum change. The large number of teachers and departmental directors attending seminars and workshops also demonstrated a possible change in the departmental English curriculum and teaching practices that might be brought about by the implementation of English proficiency tests in universities.

8.3.2 Department administration

A series of interviews was conducted with departmental directors in the department of Applied Languages, Liberal Arts Education or Language Centres which was in charge of General English teaching in universities of technology. The interviews aimed to find out how university administrators reacted to the government grant project, how the university made decisions about the organisational project and the new regulation, and how the departmental directors rearranged, if at all, General English curricula at an overall level. The interviews focused on the expectations and opinions of stakeholders at the micro level of the Taiwanese educational context.

Semi-structured face-to-face interviews were conducted with three departmental directors in three different Taiwanese universities of technology between April and June 2008. The information about this project was outlined to the directors by phone in advance. Oral consent was obtained from all of the participants prior to the interviews. To maintain confidentiality, the informants were given the pseudonyms Bill, Tony, and Cindy. Each director was interviewed in Mandarin and was encouraged to elaborate on their answers as the aim was to generate in-depth information. Each interview lasted for about 30 minutes. Notes were taken at the interviews and all of the interviews were audio-recorded. The data were transcribed in full in the original language and then translated into English, and finally analysed into major themes. The interview schedule can be summarised as follows (see Appendix 4.3.3):

1. What do you think are the major reasons why the Ministry of Education introduced the project on The Enhancement of Students' Foreign Language Proficiency in universities of technology?
2. Why did/didn't your university decide to bring in passing English proficiency tests as a graduation requirement for non-English majors?
3. (If not implementing) Does the department plan to bring in passing English proficiency tests as a graduation requirement for non-English majors in the future? Why?
4. (If implementing) Has your university introduced any innovation because of the implementation of English proficiency tests in the university?
5. What do you expect teachers to do with the regulation?
6. What will happen if students fail to meet the regulation?
7. What do you think about bringing in passing English proficiency tests as a graduation requirement?
8. (If implementing) Do you think this regulation should be implemented or rejected? Why?
9. Would you like to add anything concerning the regulation for non-English majors?

The questions were again developed on the basis of issues that were being discussed among teachers, which were collected at the preliminary information-gathering stage, and they were also piloted (one interview was held) the previous year to obtain information, as before, on the relevance and clarity of the questions, the format, and the amount of time required to answer the questions. The questions were revised accordingly: several questions were formulated more clearly as discussed in Chapter 4, Section 4.4. The departmental directors were asked to share their expectations, to identify issues that were likely to impact on the department, teachers and students, and to expand on the role they thought teachers and students could play in the organisational project. They were also asked about the role of English in higher education, the English curriculum, and the English language proficiency achievements of students. Follow-up contacts were made with the departmental directors, either by phone or e-mail during and after data collection, to invite them to clarify unclear points and confirm or disconfirm some findings.

In the interviews, the three departmental directors showed slightly different understandings of the government grant project purposes.

Because of globalisation and internalisation, English is needed. Many companies assign employees to learn techniques in other countries and English is usually the language used to communicate. In many industries, English is also needed if the company wants to do business with companies in other countries. The government is aware of the necessity for English, which is why it developed the grant project. (Bill)

Of course, the main purpose of the Enhancement of Students' Foreign Language Proficiency Project is to promote students' language ability, especially English. The Ministry of Education provides project guidelines to universities. Based on its needs, each university is developing its own organisational project. (Cindy)

Compared with Bill and Cindy, Tony appeared to have less understanding of the purpose of the project.

I began to teach at this university two years ago. The departmental director at that time told me to develop the organisational project in relation to the governmental project. I didn't read any official documents. I just did what the director told me to do. So, I am not sure about the purposes of the Enhancement of Students' Foreign Language Project. But, according to the title, I guess the purpose is to improve university students' English ability. You can find the information on the website of the Ministry of Education. (Tony)

According to Bill, internationalisation was a buzzword and university graduates with high English proficiency were in desperately high demand in the Taiwanese job market in order to meet socioeconomic needs. The comment was congruent with the advisory committee members' comments and the official document on the government grant project. However, it could be seen that, although the departmental directors were aware of the grant project, not all of them understood its aims and purposes thoroughly. To develop an organisational project, a departmental director might merely follow the guidelines provided by the Ministry of Education without a great deal of thought, or merely carry on doing what had been done before.

When asked about why the department did or did not impose the graduation requirement on non-English majors and about the future plan concerning the regulation, Bill explained that:

We don't impose the regulation on all students. Since 2002, only one department has imposed the requirement on their students. Other departments encourage students to take English proficiency tests by offering them financial incentives or waiving required English courses, but it is not compulsory. The pass rate is ok so far, so we plan to impose the regulation on all students in the academic year 2009. We believe that exerting pressure on students may motivate them to learn. In the end, students' English ability is improved.

Tony said:

The reason we impose the graduation requirement on both English major and

non-English major students is simply that we want to push students to study harder.

Cindy commented that:

It is a trend among universities to impose the graduation requirement and the Ministry of Education has considered using the English proficiency pass rate as an indicator for evaluating universities' performances. By imposing the regulation, we hope that students will study harder. However, students in universities of technology usually have low English ability. It is difficult to implement the regulation and may also cause crises later on. There are debates in the university. Some teachers agree with the imposition of the regulation because they believe that students will study harder and imposing the regulation enhances the university's reputation. Others, however, don't agree with the policy because they think the regulation will make students suffer. Therefore, we didn't impose the regulation until 2006.

It can be seen that using tests as an instrument for enhancing students' learning was a deep-rooted belief among departmental directors. More and more universities of technology were implementing the graduation regulation or planned to do so in the near future. There was a general belief that the graduation requirement would make students take English courses seriously. It prodded students into studying English hard and finally improved their English ability. However, there was a concern about university of technology students having low English ability and debates about what to do among teachers, so the diffusion rate of the policy varied from university to university.

On the question of innovations in departmental English curricula, they commented:

We have some new courses and facilities. For example, we have adopted streaming. We believe that it facilitates teaching and learning when students with the same or similar level of English ability are assigned to the same class. We also offer applied English courses, after-school remedial courses and English proficiency test preparation courses. We have built new multi-media language classrooms and a new self-study centre where students can learn online. We buy software and maintain a General English Proficiency Test (GEPT) question bank for students to practise in the self-study centre. Students are required to do

self study for 48 hours per semester. Once a semester, first- and second-year students are required to take a university-designed test based on the question bank and their scores are counted as 25% of their English grades for the semester. We believe that practising mock exams is the best way to prepare students for what they will be tested on. (Bill)

We have a new self-study centre where students do online English proficiency mock exams, or practise with a question bank. As in other universities, we offer General English Proficiency Test (GEPT) listening and reading preparation courses when we have enough grants. In conjunction with the university-designed English learning passport, we offer a variety of activities, such as an English corner where students have conversations with English teachers, an English summer camp, and English speech contests. (Tony)

With the governmental grant for this project, we offer remedial courses and English proficiency test preparation courses. These courses are free for students. (Cindy)

These comments made it clear that their universities had provided a variety of courses, activities and facilities for students and they planned to do so as long as they kept receiving the governmental grant. It seemed that the strategy of student coaching by employing simulations of the test format and content was widely used in self-study centres, as well as test preparation courses. There was a belief that doing mock exams and preparing for tests were felt to be crucial to the students' success on the tests. The comments were consistent with the results of teacher and student questionnaire surveys in which both teachers and students indicated that Group 1 teachers assigned and went through mock exam papers more often than did Group 2 teachers. Furthermore, remedial courses were offered in all three universities. The directors seemed to believe that extra courses needed to be added to the present English curriculum.

When asked about the expectations of teachers regarding the implementation of the graduation requirement, Bill commented that:

I expect teachers to cooperate with the departmental policies. For example, we

have adopted streaming and there is a grade limit for each group to avoid demotivating students' learning. Students are assigned to three groups with A as the highest level group and C as the lowest level group. Because each group uses different materials and has a different English level, the highest grade for Group A students is limited to 100, for Group B students 90, and for Group C students 80. Teachers have to be strict about this requirement. In addition, I expect teachers to urge students to do more self study and to accept responsibility for their own learning.

Tony remarked that:

I notify teachers of the organisational project and requirements for students at the beginning of each academic year. I expect teachers to pass the information to students, let the students know what they are facing, what our expectations are, and how they should perform in relation to these expectations.

Cindy said that:

I expect teachers to focus their efforts on listening and reading activities. It would be better if they used test-oriented materials in classes.

All three departmental directors expected teachers to explain the regulation clearly to students and help students prepare for the tests. By comparison, the results of the teacher and student questionnaire surveys (Chapters 5 and 6) showed that teachers did not spend a large amount of time talking about the graduation regulation in classrooms and did not direct their efforts to test preparation activities. Similarly, students did not pay more attention to the graduation regulation. There was thus a gap between the departmental directors' expectations and teachers' and students' perceptions and what actually happened in classrooms.

On the question of solutions for students who were unable to achieve a minimum graduation benchmark, they commented:

We will ask students to take external and internal English proficiency tests by their third year at the university. Those students who do not pass the tests have to take an

after-school course in their third year. The course is free for them. When the students pass the course, they meet the graduation requirement. (Bill)

The first cohort of students are in their second year of study, so we have not yet come to the time when we have to confront the English graduation requirement. I hope students can pass external tests and meet the graduation requirement as early as possible. For now we offer a make-up test and a remedial course for third- and fourth-year students who do not pass external English proficiency tests. I guess we will let those students graduate when they pass the internal test or the remedial course. (Tony)

Because the first cohort of students are in their second year of study, we don't have graduates confronting the regulation, the resolutions are still under discussion. There is a discussion about whether we postpone students' graduations or provide a window of opportunity to them, such as asking students to take a make-up test or extra courses, as other universities do. (Cindy)

Although the three universities had not had graduates confronting the graduation requirement, the directors were aware of the potential problems they would have to deal with. Although the resolutions were yet to come and being debated, suggesting the complexity of policymaking, remedial assistance seemed to be the preferred strategy.

When asked about strengths and weaknesses of the graduation requirement, their responses were:

With the regulation, some students have a motivation for learning English and they do more self study. However, others do not pay much attention to the regulation, similarly with teachers. (Bill)

The good points of the regulation are that studying English becomes a habit. Based on the scores of pre- and post-tests carried out in the university, it can be seen that students' English proficiency has improved. Moreover, the university has extended the period of taking required English courses from two years to three years. We believe that increased teaching hours lead to increased proficiency gains. However, there are weak points in the new policy. The university offers a self-designed test to third-year students. If those students pass the test, they meet the graduation requirement without taking external tests. When

the fourth-year students do not pass either an internal or an external test and they do not want to take any English proficiency test once again, they can take remedial courses instead. The university provides a window of opportunity for students to meet the requirement, so some students do not pay much attention to the regulation and do not take external tests. (Tony)

I think the regulation will eventually help improve average ability and higher ability level students' English proficiency. But it is less effective in dealing with low English level students. Teachers and students are under great pressure. (Cindy)

It can be seen that the directors believed that the graduation requirement motivated some students to learn English and study harder and that it might boost test scores eventually. There was, however, a concern over demotivating low ability level students as regards both their English learning and their ability to meet the graduation requirement, which was also a great concern frequently mentioned by teachers in the interviews. Moreover, there was also a concern over the effectiveness of the policy implementation in terms of providing a window of opportunity for failing students.

On the question of whether the graduation requirement should be carried on or rejected, Bill commented that:

The regulation should be implemented because it motivates students to learn English. In addition, English has played and continues to play an important role in globalisation. The market for graduates who are proficient in English is large, and students should have good prospects. I suggest students ought to take external tests. They will have a better opportunity in job hunting either in companies or in governmental bodies with an internationally or nationally recognised English certificate.

Tony also agreed that the graduation requirement should be carried on, but gave no further comment. Cindy commented that:

Because the number of universities of technology has grown but the national birth rate has decreased, the university ends up with students who

have low academic achievement or have been rejected by other institutions. If the university wants to carry on the graduation regulation, it needs to set up entry standards strictly. Even with streaming in General English classrooms, there is a huge gap in entrants' English proficiency. There is a lengthening tail of students at the lower end. It is difficult to ask most students to pass an English proficiency test in order to meet the graduation requirement at the moment.

It seemed that all three directors had a positive attitude towards the implementation of the English graduation requirement. Nevertheless, details and problems in the diffusion and implementation of the policy, such as resolutions for failing students and the mismatch between students' English ability and the graduation benchmark, needed to be considered more carefully.

8.4 The major issues emerging from the interviews

The interviewees expressed their opinions on the graduation English requirement and suggested a number of factors that were currently having an effect on universities of technology. These factors were political, economic, and social considerations, including industrial and commercial globalisation, the ongoing globalisation of English, and the aims of the government grant project.

A certain degree of washback effect was perceived on the departmental English curriculum as a result of the introduction of the government grant project. The departmental directors' concerns related to teaching and learning excellence, leading to increased English proficiency gains, which, it was believed, could be achieved through a variety of English courses, more innovative teaching approaches, improved teaching and learning materials, the use of internationally or nationally recognised proficiency benchmarking, and the increased use of technological resources, such as self-study centres and web-based materials. The departmental directors were also concerned for the employability of graduates, which, it was

believed, could be achieved through ensuring that a high level of English proficiency was accompanied by global awareness, adaptability, creativity and the opportunity to acquire new skills.

Some negative effects, however, were reported by both the committee members and departmental directors. The phenomenon of the narrowing of the English curriculum, the creation of cram-school-like teaching, the financial burden on students, and the growth of private cram school businesses have been observed. In addition, there was general agreement that there was a wide range of ability among students. It was considered that the higher achievers were doing extremely well, but that there was a lengthening tail of students who were underperforming. There was doubt about whether a reasonable benchmark could be set that would prove that students had attained a certain level of English proficiency without demotivating students to learn English. Moreover, the departmental directors were concerned that, since the test pass rate was relevant to the university's academic reputation and ranking, as well as the amount of government grant obtained, there was pressure to make, and anxiety about making, the policy sound as though it was proving effective in the self-evaluation reports. Finally, it seemed that there was a gap between the expectations (and understanding) of stakeholders at the macro and the micro level, which has caused problems for the diffusion of the policy and the effectiveness of the implementation.

Spolsky (1994) pointed out that the washback effect usually happens irrespective of test designers' intentions and is out of their control, though the reasons might be complicated in different educational settings. It is unlikely that any changes would happen in a linear manner. As Wall and Alderson (1993) admitted:

What we did not know at the time was just how difficult it would be to determine whether washback has occurred at all, and to decide, if there was no evidence for it, whether this was because there was no such thing or because there were conditions in the educational setting that were preventing it from 'getting through ...'
(p.47-48)

Even if there were reasonably favourable conditions in the educational settings in Taiwan at both the macro and micro level, as reported, it has still proved difficult to formulate clearly the nature of the washback effect, and even more difficult to state unambiguously whether the effect was positive or negative.

This chapter has summarised the findings at one stage in Phase III of the study, which comprised the investigations conducted in the decision-making process at both the macro and micro levels in the Taiwanese educational context. The next chapter will make a comparison between the findings in all three phases of the study.

Notes

¹ According to the English Teaching/Learning Resources Centre in central Taiwan, there were 62 out of 80 teachers present at a recent teacher seminar, held on April 10th, 2009, in which issues related to international English certificates were introduced.

Chapter Nine: Discussion

9.1 Introduction

This chapter first synthesises the findings which answer the three research questions and were obtained by using different research methods. The major topics in this study were investigated through an exploratory model of investigating washback designed to fit the Taiwanese educational context as follows:

1. The Taiwanese educational context at the macro level
 - The Advisory Committee in the Department of Technological and Vocational Education within the Ministry of Education
2. The Taiwanese educational context at the micro level
 - Department administration
 - Surveys – teachers’ and students’ attitudes and opinions
 - University visits – teachers’ and students’ classroom behaviours

Whereas Group 1 were subject to the English graduation requirement, Group 2 were not. For both groups the impact of the English graduation requirement on teaching, in general, was small and occurred in the following areas:

- washback on teaching materials;
- washback on teachers’ attitudes towards the teaching and learning of English;
- washback on students’ attitudes towards the teaching and learning of English;
- washback on teachers’ classroom behaviours.

The following section of this chapter (9.2) revisits the major issues and tries to integrate the findings from the three phases of the study.

Lastly, the research results suggest that current theory could not fully capture the gamut of washback mechanisms I documented. A review of the washback working model in the

Taiwanese educational context of the study is discussed in Section 9.3.

9.2 Synthesis of the findings

9.2.1 Research question 1: What strategies did the Taiwanese Ministry of Education and university departments use to implement the Enhancement of Students' Foreign Language Proficiency Project?

Phase III of the study consisted of an investigation into the decision-making stage of implementing English proficiency tests in the Taiwanese educational context at the macro and micro levels. The research methods employed in this phase were interviews and document studies. This phase of the study answered the first research question: what strategies did the Taiwanese Ministry of Education (MOE) and university departments use to implement the project on The Enhancement of Students' Foreign Language Proficiency? The question was split into sub-questions about the rationale and strategies used both by the MOE and universities, and about reactions of the advisory committee members and departmental directors in relation to the imposition of the English graduation regulation (see Chapter 8).

In this study, the Taiwanese educational context at the macro level referred to the International Cooperation and Exchange Advisory Committee in the Department of Technological and Vocational Education (TVE) within the Ministry of Education. The Department of TVE was in charge of national TVE and had supervisory control of colleges and universities of technology. The advisory committee members' brief was to develop and formulate policies and to advise and evaluate projects in a systematic and coherent manner (see Section 8.2).

The findings from the interviews with the advisory committee members suggested that Taiwanese education organisations tried to improve university students' English proficiency through diverse activities within a government-funded project. As one of those diverse

activities, the advisory committee members intended to promote the use of English proficiency tests in universities of technology. There was a general belief that the imposition of such tests would encourage students to study harder and consequently improve their English proficiency. Due to academic freedom and the fear of violation of the relevant law, the government was not able to make the English graduation requirement compulsory for all universities of technology. The committee members, however, had a fairly positive attitude towards the requirement (see Section 8.2.1).

The situation at the micro level, the actual teaching situation within the university setting, was not perceived quite as favourably as things were at the macro level. Although the departmental directors strongly supported the implementation of English proficiency tests as a graduation requirement, some concerns were raised. There was general agreement that there was a wide range of ability among students. It was considered that the higher achievers were doing extremely well, but that there was a lengthening tail of students who were underperforming. The departmental directors were worried about the potential pitfalls of an extension to students' study due solely to the English certification requirement. Some universities had decided to provide failing students with a window of opportunity which allowed them to take university-administered mock English proficiency tests, or remedial courses as a make-up, whereas others were still seeking solutions. This compensation plan of mock tests or remedial teaching, however, impeded the effectiveness of the graduation requirement. In addition, a range of phenomena, generally negative, such as the narrowing of the English curricula, the cram-school-like teaching, the financial burden on students, and the growth of private cram school business had also occurred (see Section 8.3.2). Nevertheless, despite these problems, the departmental directors had a positive attitude towards the English graduation regulation. They believed that (1) the regulation made students study more, (2) employers preferred certificate-holding applicants, (3) the regulation made the department feel more confident of its graduates' English abilities, (4) the regulation gave the university a good reputation, and (4) the regulation made the department more competitive in getting governmental grants.

Since the Taiwanese Ministry of Education has started to use English proficiency pass rates as an indicator of university performance, the English graduation requirement will have a pivotal impact on more and more universities of technology's policies. It was found that the government and the universities of technology had made considerable efforts to improve students' English proficiency and to assist students passing English proficiency tests to obtain their first degree. However, there was a discrepancy between the way teachers, students and policymakers viewed the effects of the English graduation requirement, which had impeded policy diffusion and the effectiveness of the implementation. The gap was mostly evident in the fact that, unlike teachers and students, the authorities portrayed a much more positive picture of the testing event and expressed satisfaction with the way English proficiency tests were being administered within the educational system. The policymakers interviewed certainly believed that English proficiency tests were effective tools for controlling the educational system and therefore used them repeatedly in prescribing the behaviour of teachers and students.

9.2.2 Research question 2: What was the nature and scope of the washback effect on teachers' and students' perceptions of aspects of teaching in relation to English proficiency tests?

The major issues regarding washback on teaching explored in both the teacher and student questionnaire surveys were as follows.

1. Teachers' reactions to, and perceptions of, the new policy:
 - reactions to the new policy;
 - the reasons behind imposing the new policy;
 - alterations to teaching methods due to the new policy.
2. Teaching materials:
 - textbook arrangements related to teaching materials;
 - teaching and learning resources.

3. Teachers' classroom behaviours:

- teaching arrangement;
- medium of instruction;
- teacher talk;
- teaching activities.

4. Assessment and evaluation:

- factors that influenced teaching;
- the use of mock exams.

5. Teachers' attitudes towards aspects of learning:

- learning aims;
- learning strategies.

The teacher questionnaire results showed that Group 1 teachers had a generally positive reaction to the new policy, which required non-English majors to take and pass at least one English proficiency test as a graduation requirement. There was a significantly higher proportion of Group 1 teachers (51.9 per cent) who welcomed the policy than Group 2 teachers (34.6 per cent), and a correspondingly lower proportion of Group 1 teachers (29.6 per cent) who were sceptical about the policy than Group 2 teachers (30.8 per cent).

In terms of teachers' perceptions of the new policy, it was clear that Group 1 teachers' perceptions of the reasons behind imposing the graduation regulation were consistent with the underlying theories behind the new policy. The teachers were aware that the main purposes of imposing the graduation requirement were to meet the government regulation, to prepare students for their future career, and to cope with the present decline in English standards. The agreement between teachers' perceptions and those of the policymakers suggested a positive attitude towards the implementation of the English graduation requirement.

The teachers' positive attitude towards the graduation regulation may have had an influence on students' perceptions of reasons to learn English at university. The factor, 'to satisfy university regulations', was perceived as a more important reason to learn English by Group 1 students than by Group 2 students, and obtaining a good job was perceived as the

most important reason by both groups. The results suggested that Group 1 students were instrumentally motivated to learn English (see Section 6.4.1). As observed in the classrooms, the students' motivation may have been influenced by their teachers, who offered test-relevant information and persuaded students to prepare for English proficiency tests in class (see Chapter 7).

When teachers were asked what changes they had made to their teaching over the previous five years, however, Group 1 teachers' perceptions were relatively similar to those of Group 2 teachers. Of the twelve items explored in the area of teaching, the only one that differed significantly between the two groups was 'do more mock exam papers', suggesting a superficial impact of the new policy in terms of teaching methods. This result was consistent with the student questionnaire's question about students' attitudes towards teaching activities inside and outside class (SQ9). 'Going through mock exam papers' was reported more often for the Group 1 classrooms; as this was an activity closely related to the English proficiency tests to be taken, it appeared to indicate a clear washback effect (see Sections 5.3.3 and 6.3.2).

There was no significant difference between the groups on the question of teaching activities. The most frequent activities that teachers carried out in class were explaining specific language items and explaining the meaning of the text. There was, however, a discrepancy between teachers' attitudes towards the relative importance of classroom activities and employing new teaching methods. Both groups of teachers gave a higher weighting to the activities of 'encouraging more student participation in class', 'putting more emphasis on reading', and 'putting more emphasis on listening'. However, they considered employing new teaching methods to further communicative competence as a change that would be difficult to implement and which they were thus not likely to make (see Section 5.3.3). These results were supported by the student questionnaire, in which Group 1 students reported that they carried out the learning activities such as speaking, pair work, group discussions, and language games less often than did Group 2, which indicated fewer communicative practice opportunities for Group 1 students in class (see Section 6.3.3). The findings suggested that although Group 1 teachers had a positive attitude towards imposing

English proficiency tests as a graduation requirement, they felt a degree of reluctance to make changes in certain aspects of the teaching that they were expected to carry out.

Teachers' reluctance to employ more communicative practices seemed to have an influence on the students' preferred learning strategies. In the student questionnaire, the results for seven out of the twelve items regarding students' learning strategies differed significantly between the groups, though the preference rates were *lower* for Group 1 than for Group 2. The least used learning strategy was expressing opinions in English in class, suggesting there was little or no washback effect on oral communicative skills, even though the authorities expected students to develop those skills (see Section 6.4.2).

With regard to whether teachers employed test-oriented textbooks and what teaching and learning resources were used, it appeared that the majority of the teachers in both groups did not employ test-oriented textbooks and they relied mainly on the content and organisation of General English textbooks as the source of their lesson plans. General English textbooks provided the syllabus, in terms of what to teach. Neither the initial assumption made by the researcher that the implementation of English proficiency tests in universities would require the use of test-oriented materials and a variety of teaching and learning resources, nor the expectation of increasing oral components in class were borne out by the results. Given the important role that textbooks have played in university of technology English education, it is not surprising that the washback effect on classroom teaching appears to have been minimal, since the teachers for the most part did not employ test-oriented materials (see Section 5.4).

In terms of classroom teaching behaviours, there was no difference in the language used by the two groups of teachers. The majority of the teachers used half English and half Chinese as the medium of instruction due, according to the teachers, to the low level of their students' language proficiency. Consistent with the results of the teacher questionnaire, Group 1 students reported that their teachers used half English and half Chinese predominantly as the medium of instruction (see Sections 5.5.1 and 6.2). Furthermore, although Group 1 teachers talked slightly more often to groups and to individuals than did Group 2 teachers, suggesting that Group 1 students received more practice opportunities, talking to the whole class was still

the main activity in the classroom. The results of both teacher and student questionnaires showed that the pattern of teachers' talk in both groups was similar, in that teachers talked to the whole class the most, followed respectively by talking to groups, and individual students. Group 1 teachers still dominated and controlled the classroom talk for most of the lesson time, though the new policy did appear to encourage teachers to be more learner-centred. In short, there seemed to be a washback effect of the new policy in the area of teacher talk, but it was not as great as had been expected (see Sections 5.5.2 and 6.3.1).

Regarding the factors that influenced teachers' teaching, two factors stood out as differentiating Group 1 and Group 2: learners' expectations and the need to obtain satisfaction from teaching. The results indicated that using English proficiency tests as an agent for change might perhaps have led the Group 1 teachers to pay more attention to learners' expectations. In the interviews, teachers claimed that students' learning attitudes were associated with their learning interests. Accordingly, promoting students' learning interests may facilitate positive learning attitudes and hence promote learning results, including their performance on English proficiency tests. Moreover, the most influential factor on teachers' teaching at an overall level was claimed to be teaching experience and beliefs, and the least influential factor was textbooks (see Section 5.6.1). There was, however, a discrepancy in that textbooks were reported by the teachers as providing the most direct teaching support, yet teachers regarded them as relatively less important than the other factors in influencing teaching.

Finally, there was a washback effect from the English proficiency tests on teachers' perceptions of aspects of learning (see Section 5.7). As for strategies, Group 1 teachers had recommended that their students put the emphasis on listening more than Group 2 teachers had, suggesting a possible change in Group 1 teachers' attitudes towards how learners should study in the context of the new regulation (see Section 5.7.2).

Besides washback on teaching, major issues regarding washback on learning explored in this study were as follows.

1. Students' learning contexts:

- study hours;
 - perception of the existence of the graduation regulation;
 - test-oriented courses taken.
2. Students' attitudes towards teaching and learning activities inside and outside class:
- their own learning activities in class;
 - their use of English in class;
 - their use of English outside class.
3. Students' attitudes towards aspects of learning:
- aspects affected by English proficiency tests.
4. Students' attitudes towards English examinations.

With respect to the students' learning contexts, students who had to sit for English proficiency tests did not spend more time than other students studying English and did not attend test-oriented classes, suggesting a minimal impact of the graduation regulation on motivating students to do self-study in English. There was no more attention being given by them to examination preparation. The majority of the students spent less than two hours a week studying English (see Section 6.2). The policymakers may have wanted the students to put more effort into learning English because of the English graduation requirement, but for the students, being required to take English proficiency tests did not necessarily lead to action. How to bridge the gap between the students' English learning and the extremely inadequate effort expended should serve as food for thought for instructors and education authorities.

In terms of students' use of English both inside and outside class, there was a lower frequency in carrying out pair work and group discussions in Group 1 than in Group 2, which showed no washback effect on learning activities carried out in General English lessons. Although Group 1 students talked to their teachers in English slightly more than did Group 2 students, Group 1 students still seldom used English outside class (see Section 6.3). Although the implementation of English proficiency tests as a graduation requirement did have a minimal washback effect on the classroom activities, it did not encourage more English-medium activities outside class.

With regard to students' perceptions of the impact of English proficiency tests, although the majority of students in both groups reported not having been affected, those who said they had been affected reported a negative reaction and effect. Finally, it was quite clear that both groups of students showed mixed feelings about examinations. Group 1 students generally expressed a more negative attitude towards English examinations. Although students really did not like examinations, examinations seemed to be one of their key motivations for learning English, indicating the important role of examinations in students' lives (see Section 6.5). Imposing the new regulation might have changed certain aspects of the students' attitudes towards teaching and learning, but the regulation did not seem to have changed their attitudes towards the function of examinations.

In summary, both teachers and students generally viewed the English graduation requirement in a positive light. They recognised the prevalence of English in the world, and shared ideas about why English is important for students as individuals and in terms of employment opportunities. The primacy of English globally was understood by both teachers and students as 'a trend', something to be accepted and exploited, rather than resisted. English was understood by the students as a ticket to opportunities and a link to the world beyond Taiwan. The students also revealed an acute consciousness of the fact that English, as the language of the world powers, is a language from which they cannot escape.

Nevertheless, the impact of the implementation of English proficiency tests in universities was seen to be superficial rather than substantial. Although teachers might have a positive attitude and/or might have altered their classroom activities at the surface level in accordance with the imposition of the new policy, substantial changes in their teaching had not occurred. With regard to students' learning, students who had to comply with the graduation regulation seemed not to be making a concerted effort to prepare for English proficiency tests. As a means of motivating students to engage in more English language learning, the graduation requirement was not proving effective. What was not apparent from the study was whether there existed an actual washback effect of the English graduation requirement, or a perceived effect (Cheng, 1999: 255). What did result, however, was a

definite change in the content of teaching, indicating the power of the existence of examinations, perceived or otherwise. It is important to point out that the short period of time the new policy had been in operation – one to two years in the participating universities – and the short period of research time could explain the finding that only superficial changes were observed.

9.2.3 Research question 3: What was the nature and scope of the washback effect on teachers' behaviours as a result of implementing the English proficiency tests?

The study explored whether the implementation of English proficiency tests as a graduation requirement had any washback effect on classroom teaching in terms of teacher and student behaviours. The research question answered was 'What was the nature and scope of the washback effect on teachers' behaviours as a result of implementing the English proficiency tests'?

The main expectations of the classroom observations were that teachers whose students had to comply with the new regulation would:

- frequently refer to the English graduation requirement,
- assign relatively more practice opportunities to students,
- assign relatively more class time to student activities such as pair work and group discussions,
- talk less,
- let the students talk more,
- use more test-oriented materials.

Unlike the Group 2 teachers' lessons, the Group 1 teachers' lessons observed included no activities involving group work, or even individual students interacting with the whole class, even though the new policy might be expected to have encouraged a more learner-centred classroom. The results suggested that there was no washback effect of the English proficiency

tests on the participant organisation of classroom teaching. Group 1 teachers still dominated the classrooms (see Section 7.4.1). In addition, there was not much difference between the groups as regards the nature of teacher talk, as a percentage of class time. The teachers talked during the majority of lesson time, suggesting that there was no more student talk in Group 1 classrooms than in Group 2 ones and that there was no washback effect on the aspect of teacher talk. One possible reason for the dominance of teacher talk was students' low English proficiency, which was evidenced in the teacher interviews. Brenda, for example, explained that as her students' English was at pre-intermediate level, she had to explain more and give detailed explanations, in order to make sure that students understood the contents and could later meet the English proficiency test requirements (see Section 7.5).

The purpose of looking at classroom activity type was to explore what kinds of teaching and learning were realised through various activities. By investigating the content of the activities carried out in the classroom, the researcher could explore the subject matter of the activities – what the teachers and the students were talking, reading, or writing about, or what they were listening to. Activity types were grouped into teacher activities and student activities. The results showed that the teacher was still the predominant focus of the classes in the two groups. Teachers spent the majority of lesson time on explaining vocabulary, reading texts and grammar work within the various topics which were included in the textbooks (see Section 7.4.2). Although the four teachers taught at different universities, they used the same type of teaching materials, namely pedagogical General English textbooks, with the one exception that Athena added a commercially-written and test-oriented textbook to her teaching. The textbooks used were either university-developed or commercially-written, and all were accompanied occasionally by audio-visual materials, specifically CDs for listening. As observed in the classrooms, the teachers relied mainly on the textbooks for their teaching. Therefore, only Athena taught to English proficiency tests, suggesting in her case an impact of English proficiency tests on teaching materials and teaching contents; the others did not do so (see Section 7.4.3).

A cross-comparison of the percentage of time spent on student activities shows that there

were substantial differences; the figure was 29% for Athena, 5% for Brenda, 20% for Cathy, and 41% for Dana. However, despite this difference, lessons were roughly similar across the four teachers. The activities mainly emphasised reading skills, vocabulary building, language points and sentence structures. The students carried out those activities individually. Group work was seen only in Dana's lessons and oral work was seen only in Cathy's lessons. The types and contents of the activities limited students to producing isolated sentences, which were assessed for grammatical accuracy. Overall, Group1 students did not receive more practice opportunities, as I had assumed they would, and the practices that did occur were not interactive (see Section 7.4.2).

A review of the questionnaire survey results from both teachers and students and at the teacher interviews showed that the teachers had a superficial perception of what was expected by their universities and by the Taiwanese Ministry of Education, in terms of implementing English proficiency tests (see Chapters 5, 6 and 8 and Section 7.5). They were thus unclear about what to change in their teaching to make it fit the graduation requirement. The results of the student questionnaire survey also showed that teachers tended to carry out activities that were related to English proficiency tests only at the surface level, and this was consistent with the results of the teacher questionnaire survey. A further careful look at the four teachers' classroom teaching and their interviews revealed that the changes were superficial even where the teachers agreed with the English graduation requirement. In short, the university's policies and the teachers' beliefs about teaching were factors that led to a variety of washback effects of the English proficiency tests on teaching. The graduation regulation involving the tests had not changed the teachers' fundamental beliefs and attitudes about how teaching and learning should be carried out.

Teachers are key players in determining whether educational change is viable. According to Rogers (2003), why innovations do or do not have their intended effects depends primarily on the ability of the change agents to overcome the initial resistance of organisational members to change. In school settings, what teachers would like to change is not necessarily the same as what they would actually do in their classroom (Cheng, 2005). The intention to

cause positive washback simply by imposing a new assessment system on the education system does not necessarily bring the expected outcome. Other factors that constrain implementation should be taken into consideration before an innovative assessment system is introduced. As Cheng (1999) pointed out,

If one expects practitioners to change themselves and their students, an environment conducive to such change must be fostered. The teaching context, school environment, messages from the administration, and expectations of other teachers facilitate or detract from the possibility of change. (p. 269)

9.3 Review of the washback model in the Taiwanese educational context

The results of this study suggest that neither Alderson and Wall's (1993) Washback Hypothesis, nor Bailey's (1996) basic model of washback can fully capture the washback mechanisms documented. In line with the conclusions of other studies (e.g. Alderson & Hamp-Lyons, 1996; Wall, 1996; Wall & Alderson, 1993), the findings of this study showed that the Washback Hypothesis (Alderson & Wall, 1993: 120-121) is oversimplified. Of its fifteen propositions (Alderson & Wall, 1993: 121), only the fifteenth can be corroborated in the present circumstances: "tests will have washback effects for some learners and some teachers, but not for others" (Alderson & Wall, 1993: 121). The remaining fourteen propositions can only be applied in some cases. In addition, Bailey's (1996) basic model of washback, based on Hughes' (1993) trichotomy, fails to show components under the label of 'processes', and other critical components under the labels of 'participants' and 'products', such as families and students' future careers. Because none of the available theories and research findings can delineate appropriately the mechanisms of washback for this study, a working model is proposed elaborating on and describing how the washback effect worked in terms of participants, processes and products.

The working model is adapted from Bailey's (1996: 264), Markee's (1997: 42-47) and

Stern's (1989: 210) models. It is also based on the initial washback exploratory model within the Taiwanese educational context, which was described in Chapter 3, Section 3.8.2. The model represents the mechanisms by which the effects of washback from English proficiency tests as a graduation requirement impact on classroom teaching. Washback is seen to involve participants, processes and products within the Taiwanese educational context.

According to Hughes (1993), the participants include all those who are involved: students, teachers, administrators, materials developers and publishers. They also include those who are related to the outcomes of teaching and learning, such as users of the examinations, parents and the community – the 'who' in Markee's curricular innovation model. All of their perceptions and attitudes towards their work may be affected by a test (see Bailey, 1996: 262). In this study, three levels of participant were studied: those who developed, intervened in and implemented the governmental project on The Enhancement of Students' Foreign Language Proficiency. These parties included mainly the advisory committee members within the Taiwanese Ministry of Education (responsible for developing the governmental project and evaluating organisational projects), departmental directors and teachers (responsible for interpreting the governmental project and bridging the change process by facility and methodological support), and teachers and students who faced the English graduation requirement.

'Processes' refer to "any actions taken by the participants which may contribute to the process of learning and teaching" (Hughes, 1993, cited in Bailey, 1996: 262). Hughes (1993) suggested that "such processes include materials development, syllabus design, changes in attitudes towards teaching and learning, changes to teaching methodology, the use of test-taking strategies, etc." (cited in Bailey, 1996: 262). The processes should also include who (participants) adopts (process) what (the innovation), where (the context), when (the time duration), why (the rationale) and how (different approaches in managing innovation) – the elements of Markee's (1997) curricular innovation. Accordingly, this study explored what happened at the three different levels of participant involved with the English graduation requirement. It then looked at the process of the English graduation requirement being

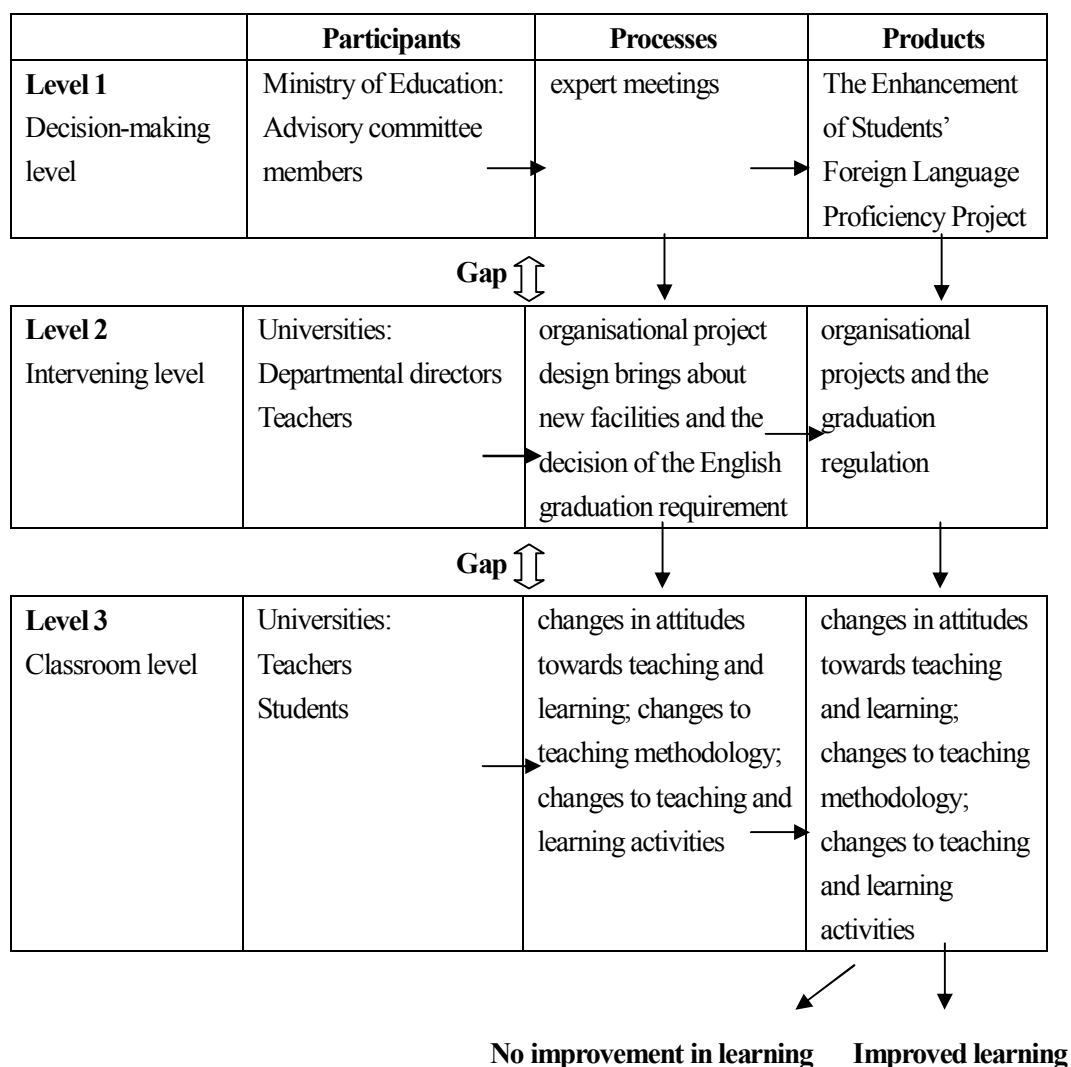
delivered in classrooms. The study focused on the processes of the washback effect not only at the behavioural level, through classroom observations to investigate the actual classroom teaching and learning, but also at both the policy level and the theoretical level, through the use of teacher and student questionnaire surveys and informant interviews (following Stern, 1989: 210).

Hughes (1993) defined the product as “what is learned (facts, skills, etc.) and the quality of the learning such as fluency, etc.” (p. 2). In this study, however, ‘product’ has a different meaning from Hughes’ definition. The washback effect being studied was seen as a process as well as a product of the processes involved. The main focus was to investigate how the participants and processes interacted within the Taiwanese education system, rather than on student learning outcomes, although the latter by themselves could be a useful focus of a washback study. Therefore, the changes in attitudes towards teaching and learning, the changes of teaching materials and methodology, and the changes to teaching and learning activities were treated as processes as well as products.

The nature and the scope of the English graduation requirement first affected the perceptions and attitudes of the participants. These perceptions and attitudes then affected what participants did in carrying out their work. The process could affect the learning outcomes, and might eventually lead to the product of teaching and learning – improved learning or no improvement in learning at all. It is important to note that during the process of diffusion and implementation gaps can occur, which is evidenced by the findings of this study. The participants interpret the policy according to their beliefs and past experiences. Because they all have different beliefs and experiences, the degree to which they feel the relative advantage, compatibility, observability, trialability, and complexity of an innovation will vary (Rogers, 2003). A certain gap exists between decision-making agencies (such as the government authorities) and intervening agencies (such as departmental directors and teachers in the universities of technology) regarding how the curriculum and the introduction of English proficiency tests are interpreted. In addition, differences in how intervening agencies and implementing agencies (such as departmental directors, teachers and students) perceive

the policy of implementing English proficiency tests as a graduation requirement forms another gap. These gaps due to interpretation and expectation differences may lead to the failure of an innovation, and then lead to the undesirable product of teaching and learning – no improvement in learning. Figure 9.1 illustrates the gaps between different levels of educational parties in Taiwan regarding how the higher education general English curriculum and the policy of introducing English proficiency tests might be interpreted.

Figure 9.1 Model of the effect of washback within the Taiwanese educational context



* → means the direction of influence.

To reiterate, washback is a complex educational phenomenon and should be investigated with reference to many “contextual variables of society’s goals and values, the educational system in which the test is used, and the potential outcomes of its use” (Bachman & Palmer, 1996: 35). It is difficult to illustrate the phenomenon thoroughly within the single model presented in Figure 9.1. Although, because of the time limitation, the model could not draw a comprehensive picture of the washback mechanisms with respect to the effectiveness of implementing English proficiency tests as a graduation requirement and how the products influence participants, processes and tests as suggested by Bailey (1996), this model does describe in detail the complexity of test washback on different levels of participant and process. In addition, the model can serve as a guide for future washback studies. The significance of this study lies in its contribution to the methodological considerations of identifying the mechanism of the washback effect, and to the identification of areas of and the extent to which classroom teaching is affected by the tests.

Chapter Ten: Conclusion

10.1 General aims of the study

As in many other countries, English is considered an essential tool by which individuals achieve success in life in Taiwan. Acknowledging the reality that English has become a lingua franca in international business, virtually all Taiwanese people are committed to learning English. This study has explored the influence of English proficiency tests on a sample of key stakeholders, particularly university teachers and students in the classroom setting. The English proficiency tests, namely GEPT, CSEPT, TOEFL, IELTS, TOEIC, BULATS, Cambridge Main Suite Exams and university self-developed tests, are used as gatekeeping tests by Taiwanese universities of technology to screen graduates for language ability. Performance in the tests may have serious implications for the life chances of test takers. Therefore, the English graduation requirement, imposed by universities and encouraged and funded by the Taiwanese Ministry of Education, was expected to exert a strong influence on teacher and student behaviour. This study accordingly set out to investigate the interaction between teachers and students in moderating the washback effects of the English graduation requirement, as well as investigating the flow of the policy diffusion and implementation.

The study included a limited number of teachers and students on a restricted sample of courses preparing non-English major students for General English study at Taiwanese universities of technology, as well as departmental directors and advisory committee members within the Ministry of Education. Non-English major students differ from their English major counterparts in that the former are not required to study English regularly for their academic studies. Indeed, non-English major students receive only two or three hours of English instruction per week, but they are required to comply with the graduation regulation as their counterparts do. The focus thus has been on the General English teaching and on the

influence of the English graduation threshold (GT) on non-English majors. Clearly, there are limits to how far the results can be expected to generalise beyond this context. Further research is needed into how teachers and learners in other contexts may experience washback due to GT. Although the GT is not compulsory in all universities of technology at the moment, when the advisory committee within the Ministry of Education was established to examine the effectiveness of its language enhancement grants, and it became immediately apparent that there was a possibility of repercussions in the form of reduced or removed grants, university directors felt obligated to support the government's call for the introduction of the GT. As a result, more and more universities of technology began to implement the GT.

Evidence from both the teachers and the students showed superficial washback effects of the English graduation requirement regarding classroom teaching activities and both teachers' and students' expectations of the regulation. The effects were to some degree negative, in the sense that the teachers and the students were narrowly focused on practice of test-related content and tasks, rather than on the development of English proficiency in a broader sense. While there was little evidence to suggest that the students played a direct role in shaping washback to the teacher, the data suggested that the teachers might have been influential in shaping washback to the learners.

Evidence from interviews with departmental directors and advisory committee members within the Ministry of Education showed a positive attitude towards the English graduation requirement. There was a general belief that the regulation encouraged students to take English language learning seriously and consequently to improve their English proficiency to a college level.

The findings of this study have led to a clearer understanding of what key stakeholders think about current English language testing and assessment in Taiwan and their effects on teaching and assessment practices. The research supports the claim that washback is neither simple nor direct, but circuitous and complicated (Wall & Alderson, 1993). It does seem that the imposition of the English graduation requirement evokes varied and unpredictable responses. Policymakers may find themselves confounded in a culture where the target

language is not necessarily highly valued by test takers; where it may be viewed less as a means of communication than as an academic exercise; where test preparation is therefore associated with practising past papers and is viewed as a short-term activity; where learners show little evidence of independence in their approach to the learning activities, whether as a result of their educational experience, or for other reasons.

The above situation suggests that tests might not be a good lever for change, if this lever is used to set the educational machinery in motion. The machinery, in the sense of the educational systems or teaching practices, will not easily lend itself to control by tests. Under such circumstances, the challenge confronting the testing experts in Taiwan is to work out a sound testing policy based on rigorous validation research, which will help maximise the effectiveness of English testing for non-native English speakers. It is hoped that language testing research can ensure that this powerful instrument does not victimise the innocent, but serves its purpose by enhancing the prosperity of the public and society.

10.2 Implications of the study

The recent expansion of tertiary education in Taiwan has inevitably pulled down the average English performance of graduates (Ko, 2002). At the same time, the Taiwanese government's commitment to developing its nation's economic competitive edge in the international market has necessitated an increase in demand for university graduates with high English proficiency (Chang, 2003). In view of graduates' English deficiency and society's rapidly mounting dissatisfaction with the deficiency, it has become a matter of urgency for tertiary institutions to take measures to rectify the situation. Imposing a control on graduation is probably one of the major measures institutions need to take.

In many ways the most expedient method of assessing graduating students' English ability is by means of a test. Tests are often used to generate top-down educational change. Indeed, Shohamy (1992) pointed out that "external tests are currently used to force students to

study, teachers to teach, and principals to modify the curriculum” (p.513). The results of my study, however, indicated that using English proficiency tests as a graduation requirement is problematic for a number of reasons. First, no single policy can guarantee success without any side effects, especially where high-stakes testing is involved. The implementation of the English graduation requirement involves a number of thorny issues, such as administrative and financial problems, and inevitable resistance. Though universities of technology can be confident that the public welcomes their introduction of graduation standards control, it can be anticipated that the greatest resistance to their implementation of controls may come internally from the affected parties in the institutions.

Second, using a single test or set of tests for multiple purposes can diminish the value of that test for any single purpose. To exercise graduation regulation, it is essential to ensure the validity and reliability of the tests employed. Validity refers to the extent to which a test measures what it is intended to measure and not what it is not designed to measure. If what is to be measured is already controversial, the validity of the test is unlikely to be agreed on. Reliability is essentially concerned with how consistently the test does what it is supposed to do. A common test error is associated with the cut-off score. The indication of mastery of certain language abilities rests on reaching the cut-off score. Where the cut-off score of a standard control instrument should be set, or how to determine the cut-off score accurately, is often subjective (Davies, 1990). Taking the IELTS and the TOEIC as an example, the academic module of the IELTS measures English proficiency levels in order to evaluate candidates’ readiness for academic studies, whereas the TOEIC measures individuals’ English communication levels in a business context in order to evaluate candidates’ suitability for professional employment. The constructs of the two tests are distinct from each other and they test different areas of knowledge. The academic module of the IELTS consists of receptive (reading and listening) and productive (writing and speaking) subtests, whereas the TOEIC consists of listening and reading subtests – these being the only components of the test. To claim that educational institutes can adopt any of them as an entrance, placement or graduation test, or that employers can use any of them to recruit employees conversant with

English is overselling the tests. In addition, the two tests' score reporting systems are also different. Claiming from a table of score equivalences that obtaining 550 in the TOEIC equals a 4 in the academic module of the IELTS makes the validity of the table of equivalences questionable, since the two tests test different knowledge with different purposes (see Appendix 1). With the table of score equivalences in question and without any guidance on applying the table, the validity of the graduation requirement is questionable. The policymakers seem to be oversimplifying the applicability of English proficiency tests.

As long as concerns over the deterioration of English proficiency still exist in the business sector, which is highly influential as regards government policies in Taiwan, and as long as there is insufficient proof to convince the government that university students' English language proficiency has considerably improved, the chances are that an English graduation test will eventually become mandatory for all graduating students. Given this likelihood, the following suggestions are offered as being conducive to positive washback on teaching and learning and to the effectiveness of policy implementation.

10.2.1 Implications for making an ideal graduation test

Up-to-date information shows that universities of technology in Taiwan use a variety of standardised tests to set standards of English across a variety of professions. These tests are international tests such as IELTS, TOEFL, TOEIC, BULATS and Cambridge Main Suite Exams, and locally developed tests such as GEPT and CSEPT (see Chapter 2). Graduating students can take whichever test is most relevant for their prospective profession and meets the graduation requirement. The locally developed tests are widely available throughout Taiwan and less costly for test takers than the international tests, but they do not grant international recognition. With the international tests, on the other hand, students know that their test scores would receive international recognition and those wishing to go on to study abroad can also do so without having to take a further language qualification. These tests,

however, are designed for a different context and purpose and hence they may not be appropriate for the Taiwanese tertiary context.

Different tests also have different purposes and targeted test takers. For example, the TOEIC and BULATS are workplace-oriented tests, likely to be biased in favour of students studying business, and against others such as those in the sciences. The test scores are highly relevant as evidence of English language proficiency for future employment. On the other hand, the academic module of the IELTS and the TOEFL are designed to assess English language proficiency for academic purposes. Generally speaking, none of their components are specifically workplace-oriented. People whose English proficiency can cope with academic studies do not necessarily perform well in professional workplace communication because the two types of communication involve different sets of constructs, strategies and registers. There are few grounds for believing that both the IELTS and the TOEFL can effectively serve the purpose of benchmarking university graduates for professional workplace communication.

In addition to using a graduation test as a means of improving students' English proficiency, the Taiwanese Ministry of Education also intended to use the graduation test to facilitate students' job hunting or their pursuit of future studies (Ministry of Education, 2005b). It is therefore important that the development of the graduation test is valid for both purposes, that is, for employment and for academic studies. However, as discussed above, none of the tests can claim with confidence to serve both purposes at the same time.

Some of the current graduation tests adopted by universities of technology, such as IELTS, TOEFL and GEPT, appear to be more comprehensive than the others, as they contain components covering all four commonly recognised language skills – listening, speaking, reading and writing. This comprehensive format may make the tests look more credible to the general public, especially to some potential test users. In contrast, the TOEIC contains only listening and reading and the CSEPT contains only listening and lexical knowledge, which, to some test users, may sound insufficient in terms of skill coverage.

Deciding what a test should assess is one of the most controversial aspects of testing. A

testing instrument may assess knowledge of a language system, lexical knowledge of the language, and a wide variety of other language-related knowledge. As mentioned above, what should really be tested for graduation control purposes is graduates' ability to use language for effective communication. Basically, the community expects university graduates to be able to communicate effectively. It does not expect every graduate to be an expert in language. Universities of technology which implement or intend to implement graduation controls should keep this in mind and focus on their intended goal.

It also seems unnecessary to demand unification of the English graduation requirement within a university. The public and employers may not expect the same English proficiency from graduates of different disciplines. For example, the expectation of the language ability of a graduate of economics is usually higher than for a graduate of engineering. Given the differences in expectations and the fact that universities of technology value academic freedom, universities should be allowed to adopt different measures for graduation standards control. The authorities, however, should provide detailed information about the components tested and the score equivalence, if it is adopted, of each of the English proficiency tests, enabling students to make informed choices as to which test would best serve both their individual learner preferences and their professional needs.

10.2.2 Implications for teaching and learning

The need to develop flexible manpower for economic globalisation can provide an opportunity to remodel university curricula. The implementation of graduation regulation requires changes in existing systems, such as the structures and requirements of study programmes and assessment methods. It would be beneficial to university of technology students if English courses could be extended over four years, or test-preparation courses could be integrated into compulsory English courses, or else offered as elective courses to third and fourth year students. It is generally recognised that students need an intensive and

usually extended period of study to achieve any substantial increase in their score on a proficiency test. In the interviews, Athena mentioned that her Group 1 students did not know how to prepare for the English proficiency tests (see Chapter 7, Section 7.5). Cindy, one of the departmental directors interviewed, mentioned that two years of English language learning in university are not enough to facilitate students' English learning and hence boost their ability to meet the graduation requirement (see Chapter 8, Section 8.3.2). The current situation is that students often stop doing English-related activities in their third year at university. By the time they take English proficiency tests, their English ability has declined or is not at an adequate level to pass tests. Therefore, the university authorities could usefully extend non-English major students' English learning hours and offer test-preparation courses to enhance students' learning and maximise the washback.

In addition, universities should provide a variety of additional English learning resources and the authorities and instructors should explicitly inform students of those resources. The universities participating in this study all had self-study centres equipped with mock exams and test-oriented resources, which students could draw on for their test preparation, and these resources can be beneficial. Nevertheless, the fact remains that the main purpose of establishing the English graduation requirement was to improve students' English proficiency, rather than merely having them pass English proficiency tests. English learning resources other than test-oriented ones are a valuable contribution to students' learning. To assist students' learning and test preparation, universities should enhance their self-study centre services and disseminate the information regarding the university's resources to each student, even though, only a few may utilise the resources, for all sorts of reasons.

To help lower-proficiency students become more strategic and independent learners, teachers should consider providing them with additional opportunities and assistance. Multimedia equipment is one of the possible options. In the student questionnaire, 'watching TV or movies', 'listening to the radio or CDs' and 'reading newspapers and magazines' were the activities that the students reported carrying out most frequently in English outside the class (see Section 6.4.5). Of these, 'watching TV or movies in English' was the most preferred

learning strategy (see Section 6.4.2). As suggested by Shih's (2007) study, students are likely to be more interested in learning English if they learn in pleasant surroundings using the activities and strategies they like. Therefore, I would recommend that teachers try to include students' preferred learning strategies in their lesson plans. In addition, Taiwanese universities have a mandate to make the Internet accessible to their students. Teachers should therefore take advantage of the Internet and e-learning environment to provide lower-proficiency students with additional assistance, such as perhaps answering their questions through a blog.

Some factors such as group dynamics and peer power should also be taken into consideration. Lower-proficiency students seem to need help in becoming more strategy conscious because, according to O'Malley and Chamot (1996: 133), the so-called 'effective learners' used strategies more successfully than the less effective ones. Van Lier (2004) suggested that scaffolding for language learning could involve "not only an expert-novice relationship, but also an equal peer one, a peer to lower-level peer one, and (even) a self-access one" (p. 162). As long as learners were situated in an environment with "interlocutors of different kinds" (van Lier, 2004: 162), this would be an appropriate context for scaffolding to take place. Accordingly, it is recommended that, in the Taiwan context, university teachers should provide more pair work and group discussions in order to provide scaffolding to reach the 'zone of proximal development' (ZPD).

According to the organisational projects of the universities participating in this study and the interviews with departmental directors, a wide range of extracurricular activities is used by various universities. Examples include offering counselling services, holding a variety of contests, establishing an English corner, giving speeches and subsidising study tours abroad (Ministry of Education, 2005b). Nevertheless, the effectiveness of implementing those extracurricular activities is in question. As mentioned in the teacher and departmental director interviews, it seems that most of the students who participated in those activities were ones with higher English proficiency. In order to encourage lower-proficiency students to participate in extracurricular activities to increase their English practice opportunities, allowing assessment to include a portfolio of work reflecting their ability to use English in a

variety of contexts might well be a useful instrument. Although the compilation of portfolios is time-consuming and they can be difficult to interpret and rate in a standardised way, such a method of assessment would nevertheless be more oriented towards learner needs (Berry & Lewkowicz, 2000). The advantages of using portfolios to enhance as well as evaluate learning have been described in the literature (see Daiker et al., 1996; Hamp-Lyons & Condon, 2000; Moya & O'Malley, 1994). Creating a portfolio, for example, necessitates reflection on students' own performance and on their improvement, encouraging continued effort throughout their studies and it can reflect authentic contexts both within and outside the classroom. In addition, in the Taiwanese university context developing a portfolio would give students something to show future employers concerning their language experiences.

Moreover, universities should keep teachers and students fully informed of the graduation requirement and the available options. In the student questionnaire survey, it was clear that students were confused about whether they had to comply with the graduation regulation or not and what options they had. Although teachers were aware of the purpose of imposing the graduation requirement, they seemed not to clearly understand what actions they should take with respect to their teaching. They also seldom provided students with test information or clarification because they themselves did not thoroughly understand the departmental policies and expectations, or the purposes, test formats and test contents of each English proficiency test allowable as a graduation test. Due to the inefficient diffusion of options and departmental expectations, neither the students nor the teachers were able to thoroughly evaluate the options or appreciate the merits of what were, to them, unfamiliar assessment procedures. It is clear, therefore, that much more information should be disseminated to all stakeholders regarding any options under consideration. Both the government and university authorities should clearly state their expectations and action plans to relevant stakeholders and allow sufficient time for them to be ready for the new policy.

In reality, the university authority should divide a long-term goal into several manageable short-term aims, as a long-term goal is sometimes too abstract to put into action. For example, if students are not able to pass the elementary level of GEPT, but are required to

pass the intermediate level three years later, without any ongoing help from teachers or available resources, students may think the goal is quixotic and either wait until the last minute to take an English proficiency test or adopt a wait-and-see approach, hoping, perhaps, for a solution to come along at some point in the future. To engineer positive washback, it is important to understand students' levels of proficiency and set classroom tasks at an appropriate level, so that the tasks may be reasonably challenging, but not too challenging, to motivate students at an appropriate level (Watanabe, 2000). To this end, it is important to monitor students' development on a regular basis. If universities can offer a consistent improvement in their English programmes and conduct both pre- and post-tests each year to monitor students' learning, students may consider the goal more attainable and therefore be more willing to prepare for the test.

As Markee (1997) noted, a top-down approach "does not promote long-lasting, self-sustaining innovation effectively" (p. 64) and the findings from this study serve to confirm his point. Teacher beliefs and experiences are both important for successfully implementing a new policy, as evidenced in the literature review and in the teacher interviews (see Chapters 3 and 7). A bottom-up, university-based approach to curriculum renewal is a desirable alternative to the top-down educational change strategy if the aim of improving university students' English proficiency is to be achieved. In the Taiwan context, the education authorities could usefully develop partnerships and respect the expertise of teachers who are responsible for implementation. Understanding the practical wisdom held in universities requires that policymakers engage with teachers who have been successful in achieving the desired objectives to develop an understanding of the conditions underpinning success. Additional understandings can be obtained from those who have experienced difficulties in achieving such objectives, by asking questions such as, 'What maintains the *status quo* and what might be the consequences of disturbing it for those responsible for implementation?' Ongoing collaboration and information exchange is the key and these can be achieved by methods like an annual meeting and/or a web log. The design of communication networks with teacher involvement can not only contribute to the

sustainability of a project, but also clarify the political and administrative dimensions of a project. As a result, teachers stand more chance of seeing the implementation of the new policy as internal and controllable, rather than external and uncontrollable. In addition, the organisational project itself will end up being planned in a more contextually-appropriate way and be refined to ensure a better fit between its environment and the authorities' expectations.

Although most of the instructors in universities of technology are academically qualified, being proficient in English or having experience of studying abroad does not allow them to teach efficiently in all circumstances. The results of this study suggest that the impacts of English proficiency tests on classroom teaching remain superficial. It seems that the degree of teachers' familiarity with a range of teaching methods was a factor mediating the process of producing washback (see Chapters 5 and 7). How teachers perceived teaching and learning might differ, but not how they carried out the teaching and learning activities. The graduation requirement has given teachers information on *what*, but not *how*. Simply importing English proficiency tests into the education system will not fulfil the intended goal of promoting communicative instructional practices and hence of improving students' English proficiency. Professional development must be involved in the process of implementing the new policy. Given such concerns, university departments should provide teachers with in-house workshops or seminars to understand the rationale of the graduation regulation and the nature of different English proficiency tests, and how to incorporate the knowledge into their lesson plans and instructional practices. In addition, university departments should provide teachers with more training on how to use test data to critique and improve their instruction. These goals can again be achieved by workshops or seminars. Without professional development aimed at understanding and using test data, teachers might not know how to use this information to improve their instruction or help their students meet the graduation requirement.

Research studies into high-stakes testing often show that it encourages teaching to the test and negatively affects students' learning (Smith, 1991a; Noble & Smith, 1994a; Chapman & Synder, 2000). One obvious result of teaching to the test is that it often involves a

narrowing of curriculum and instruction. However, narrowing of curriculum or instruction is not automatically categorisable as appropriate or inappropriate. For example, Athena claimed that the graduation regulation helped sharpen her focus with respect to her course aims and content as well as her lesson planning (see Chapter 7, Section 7.5). It was nevertheless observed that Athena interrupted her regular classroom instruction in the weeks prior to in-house TOEIC administrations to prepare her students. If the preparation is focused on test format instead of on the underlying learning objectives, the consequence may be higher test scores without improved student learning. Therefore, such an interruption in classroom instruction was and is inappropriate. Teachers should integrate assessment into their classroom instruction in order to avoid any negative effects of teaching to the test.

However, the above-mentioned changes cannot happen overnight. The translation of global imperatives for education into a local process is subject to the constraints of local players and conditions, thereby exposing one major tension in the country. The tension concerns the difference between the government's pace and readiness and that of implementing agents. Remodelling of curricula inevitably creates a gap between reform idealists, such as the government, who press for changes, and reform implementers, including teachers and departmental directors, who are relatively conservative or even reluctant to change. The gap can be narrowed only when the former can offer more time, understanding, and support to the latter. As shown in this study, as well as suggested by Law (2004), teachers, who played a key role in curricular reform, required psychological preparation to arouse and sustain their belief in reform, as well as retraining to help them acquire new skills and re-acquaint themselves with curricula and pedagogy.

10.3 Limitations of the study

An unavoidable problem for PhD students is that there is a limited amount of time available for the planning and execution of the research project, and limited opportunity to revise

research plans in line with findings as they emerge. Shohamy et al. (1996) reported that washback can evolve over time, so a longitudinal study would perhaps be better able to capture and monitor the ebb and flow of the English graduation requirement's impact. Ideally, I would have liked to observe the same students from the same institutions on entry to and exit from their Bachelors degree study. However, this would have been impossible in view of the time limitations associated with PhD research. Even so, the findings would have been more interesting if the same number of students from different study years had been involved in the study. As it was, only first and second year students were able to participate. The data for this study were collected over just five months, so follow-up studies are indispensable in exploring the English graduation requirement's long-term washback. Since the Ministry of Education has considered the English proficiency test pass rate as one indicator of university performance, the impacts are believed to be increasing.

Another limitation was the fact that it is extremely difficult to secure the necessary support and involvement of other people in a PhD study because it is often perceived not to have the authority and status of research that is conducted by qualified researchers. For this reason, I am particularly grateful to those advisory committee members, departmental directors, teachers and students who devoted effort and valuable time to participating in this study. Even so, in spite of the efforts of over 650 participants, the difficulties of securing the involvement of participants are evident in the outcomes. Only departmental directors from institutions which have implemented the English graduation requirement for two years participated in interviews, and the majority of students who completed the questionnaires were first and second year students. The imbalance in the involvement was a question of willingness and methods of approaching participants. The findings would have been more interesting if the study had involved directors from institutions which had implemented the English graduation requirement for more than five years and from institutions which had not yet implemented the regulation. It would also have been more interesting if more third and fourth year students had been able to complete the questionnaires.

This study is one of the few washback studies that have employed both quantitative and

qualitative data to study washback. The third limitation of this study, therefore, relates to the fact that there were not many existing instruments in the area of washback which could be drawn upon. No single uniform questionnaire has emerged as being widely used to survey either teachers or students about language testing washback. Bailey (1999a) pointed out that it would be a valuable contribution to the available methodological instruments for washback study to develop a widely usable questionnaire for teachers and for students.

I tried to develop parallel questionnaires drawing upon theoretical underpinnings from classroom language teaching and learning studies and contacts with departmental directors, teachers and students at the preliminary information-gathering stage (see Chapters 3 and 4, and Appendix 4). Due to the time constraints on the study, the questions and items in both questionnaires were relatively limited. There were also problems with some of the theoretical constructs in the questionnaires; items within each question needed to form a more coherent construct of aspects of teaching and learning. This weakness prevented more advanced statistical analysis of the questionnaire data. Statistical analysis was limited to t-test comparisons. This serves as a reminder to devote as much time as possible to the development of the instrument; the questionnaires in this study could have been refined to a greater extent if it had been possible to devote more time to further trials and testing of the items within each question.

The classroom observation scheme also had some limitations. I tried to develop an appropriate instrument to capture classroom interaction with a focus on large segments of the lessons. The research problem lay in the question of the levels of classroom interaction on which washback could have an impact. I feel that the classroom observation scheme could be improved further in order to provide a closer look at classroom interaction as brief as a single turn.

While the use of questionnaires, interviews and classroom observations in this study has provided a means of investigating the washback phenomenon on different stakeholders, more sensitive instruments will probably be required to probe it further. In-depth interviews with students or the use of learning records, in conjunction with an examination of English

proficiency test scores, might provide the evidence to support more detailed accounts of such relationships.

In addition, I myself was the “primary instrument of data collection and analysis” (Merriam, 1998: 42), which implies that it was largely dependent upon my ability to elicit participant responses, and also on participants’ enthusiasm to share their perspectives and their willingness to allow me into their lives. By exploring participants (decision-making, intervening, and implementing levels) at different levels of the educational context, a more in-depth study would, to some extent, reveal more detailed information about the mechanism of the washback effect. In a way, a washback study remains a compromise, allowing limited investigation of both the macro level and the micro level, unless a team of researchers can gather together to look at all facets of teaching and learning (Saville & Hawkey, 2004).

10.4 Future research directions

Bearing in mind the above limitations, one possible direction for future research is to conduct extensive longitudinal studies on the washback effect of the English graduation requirement in order to evaluate in what way and to what extent the regulation influences stakeholders, particularly teachers and students, over a period of time. Following English language teachers over several years, and interviewing them and observing their lessons, will allow researchers to gain in-depth understanding of their perspectives on their roles as English language teachers in relation to the implementation of the English graduation requirement and might help to explain how and to what extent the regulation influences their instruction in classrooms. Similarly, following students over a period of time, interviewing them and conducting tests with them at different stages of their English education might help to explain how and to what extent washback from the regulation influences students’ English language learning and allow one to evaluate the effectiveness of the regulation. A longitudinal study would not only generate better mutual understanding between participants and researchers,

but also offer insights into how the English graduation requirement over time influences the attitudes and behaviour of university teachers and students.

A second direction would be to study the impact on all levels of the Taiwanese education system. Most of the participants in this study were from private universities in the vocational tertiary education sector and the participating universities were less experienced than those in the public sector in implementing the new policy. More studies are required to investigate the English graduation requirement's impact on national universities of technology and on more experienced universities which have implemented the policy for several years, in order to shed light on the use of the English graduation requirement by tertiary institutions with different levels of experience and discover the impact on the teaching and learning of English in different educational institutions. Moreover, this study did not discover if the location of a university contributed to explaining how washback from the English graduation requirement influenced teachers' curricular planning and instruction. One possible reason may be that most of the participants were from universities located in southern Taiwan. Therefore, departmental directors, teachers and students from central and northern Taiwan should all be included to make the results more comprehensive.

A third direction would be to study the impact on students who have completed their schooling and on people who surround students, such as parents, siblings and partners, as well as the impact from the workplace. These aspects of washback have rarely been investigated. Considering that the interviewees in my study reported that the above-mentioned people and factors do have an impact on students' learning (see Chapters 7 and 8), more relevant studies, particularly on the English language requirements of workers in workplaces and occupations, from corporations to government and quasi-government institutions, are required.

As long as English proficiency tests are used as a gatekeeping device for university graduation, further investigation into the different forms of preparation for the tests – their aims, methodology, and ultimately, their effectiveness – must be carried out. As the number of studies of the English graduation requirement increases, we will gain a better understanding of the washback effect of the regulation in different classrooms and, more generally, its impact

in this high-stakes environment. Moreover, it seems inevitable that different types of English proficiency test for non-native English speakers without common criteria for proficiency will invite the problem of incompatibility in terms of domains of language to be measured and their level of difficulty (Bachman, Davidson, Ryan, & Choi, 1995; Davidson, Turner, & Huhta, 1997). Although the Common European Framework of Reference for Languages (CEFR) is employed in the current graduation requirement situation in Taiwan, research studies have shown that the CEFR is in itself problematic (Alderson, 2007; Bonnet, 2007; Byrnes, 2007; Little, 2007; North, 2007; Weir, 2005; Westhoff, 2007). Given the variety of language proficiency criteria being employed, there is a need for rigorous validation research on the compatibility of standards among the tests, to develop proficiency standards for Taiwanese learners in particular.

A high proportion of the departmental directors and teachers interviewed indicated that they were concerned about their students' abilities when developing their lesson plans and when debating whether to impose the English graduation regulation. Given its purpose to promote English learning, the graduation requirement is supposed to stimulate students of low rather than high English proficiency, increasing their motivation and eliciting greater effort. It is therefore necessary to compare the motivation to learn English of low-proficiency and high-proficiency students in relation to the English graduation requirement. It would also be valuable to make a direct comparison of the test scores in terms of students' English proficiency gains and universities' relative effectiveness in preparing students for meeting the graduation requirement. Further studies on proficiency that involve a range of different universities would provide policymakers with useful comparative data. This study has demonstrated that students' and teachers' reflections deserve to be heard by other groups of stakeholders, including EFL educators, education policymakers, exam developers and university administrations, in order to improve practices and results of English language testing and assessment in Taiwan. It is also abundantly clear that, as major stakeholders in the educational system, students must be consulted on the development of a reporting mechanism at graduation, if it is to have any chance of succeeding when implemented.

Finally, the findings of this study have raised a number of issues regarding English language instructors that would be worthy of further research. These include issues relating to the qualifications and training of tertiary-level teachers of English and their experience of curriculum planning and course design. I believe that what is needed is not only examination practice, such as coaching, but a sound approach to developing English language proficiency. This study has shown that imposing the English graduation requirement does not change the degree of emphasis on examinations, nor does it necessarily redirect teachers' methods of teaching in any fundamental way, if teacher education and professional development is not involved (see Chapters 5, 7 and 8). Imposing the graduation regulation changed the content of teaching rather than the methods teachers used to teach. Efforts should be made to investigate factors in the educational systems to seek better means to promote communicative teaching in Taiwanese universities of technology. Factors that are thought to hinder the operation of washback include: teachers' beliefs about language teaching and test preparation; teachers' educational background; teaching styles; teaching experience; inadequate communication between relevant stakeholders (Alderson & Hamp-Lyons, 1996; Cheng, 2005; Wall & Alderson, 1993; Watanabe, 2004). In short, this study has shown that washback works for certain teachers but not others. In order to understand the 'why' and 'how' of the above questions, a closer analysis could be undertaken. Investigations into these factors and their interactions might deepen our understanding of the function of washback and suggest ways of improving this "lever" for positive, lasting, and widespread change.

Appendices

Appendix 1 Comparative graduation English language proficiency benchmarks from various institutions

National Taipei University of Technology (2005)

Type of test	Required benchmarks
GEPT	Intermediate – stage I
TOEFL (PBT)	457 and above
TOEFL (CBT)	137 and above
TOEFL (iBT)	47 and above
IELTS	4 and above
Cambridge Main Suite Exams	Preliminary English Test
Business Language Testing Service (BULATS)	Level 2
TOEIC	550 and above
CSEPT	230 and above

National Formosa University (2005)

Common European Framework	Type of test	Required benchmarks
B1	GEPT	Intermediate – stage I
	TOEIC	670 and above
	TOEFL (CBT)	200 and above
	IELTS	6 and above
	In-school online GEPT	Intermediate – stage I

Fooyin University (2005)

Common European Framework	Type of test	Required benchmarks
A2	TOEIC	350 and above
	TOEFL (CBT)	90 and above
	TOEFL (PBT)	390 and above
	IELTS	3 and above
	GEPT	Elementary stage
	CSEPT	170

Appendix 2 Common Reference Levels

(Source: Council of Europe, 2001)

Basic User	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.
	A2	Can understand sentences and frequently-used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
Independent User	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst traveling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.
	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
Proficient User	C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
	C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.

Appendix 3 Questionnaires and interview schedules for the pilot study

3.1 Teacher questionnaire

Teachers' Perceptions of English Proficiency Tests in Taiwan Technical Universities

Dear Teacher,

I am conducting a small research survey on the impact of implementing English proficiency tests in technical universities in Taiwan for my doctorate at the University of York. I would like to ask you for your opinions of English teaching and the necessary preparation you intend to carry out in order to cope with the educational change. To help me, please fill in this questionnaire based on your own experience. It should take no more than thirty minutes to complete. *All information will be treated in the strictest confidence.* Thank you very much.

If you would like a summary of the final report, please e-mail me:

Hui-Fen Hsu
University of York
hh524@york.ac.uk

PART ONE Please tick one appropriate answer or provide written answers.

- (1) Your gender: Female Male
- (2) Your age: 21-30 31-40 41-50 above 50
- (3) Your highest education level: Bachelor Masters PhD
- (4) Number of years you have been teaching English in technical universities:
 1-3 4-6 7-9 Other _____
- (5) Levels you currently teach. Tick more than one box if necessary:
 Level one Level two Level three Level four
- (6) Number of periods you teach English per week in this technical university:
 3 – 6 7 – 10 11 – 14 Other _____

PART TWO Please circle the appropriate answer or provide written answers.

(7) If your department implements proficiency tests as a graduation requirement, what do you see as the major aims?	not a major concern at all	of little concern	of great concern	main concern
To meet the government regulation	1	2	3	4
To prepare students for their future career	1	2	3	4
To cope with the present decline in English standards	1	2	3	4
To motivate students to use integrated skills	1	2	3	4
To encourage students to play an active role in their current learning	1	2	3	4
To enable students to communicate more with foreigners	1	2	3	4
To encourage better textbooks	1	2	3	4

If your department does not implement proficiency tests as a graduation requirement, can you summarize the reasons why it decided not to?

(8) What kind of extra work or pressure if any do you think the proficiency tests put on you in your teaching?	no real increase	little increase	some increase	a large increase
Having to follow a new syllabus	1	2	3	4
Doing more lesson preparation	1	2	3	4
Preparing more materials for students	1	2	3	4
Revising the existing materials	1	2	3	4
Adopting new teaching methods	1	2	3	4
Setting up new teaching objectives	1	2	3	4
Organising more exam practice	1	2	3	4

(9) To what extent are the following changes you have made in your English lessons during the past five years? I now ...	not at all	some extent	a reasonable extent	a very large extent
Employ new teaching methods	1	2	3	4
Use a more communicative approach in teaching	1	2	3	4
Put more stress on group discussion	1	2	3	4
Put more emphasis on listening	1	2	3	4
Put more emphasis on speaking	1	2	3	4
Put more emphasis on reading	1	2	3	4
Put more emphasis on writing	1	2	3	4
Put more emphasis on the integration of skills	1	2	3	4
Encourage more student participation in class	1	2	3	4
Employ more real life language tasks	1	2	3	4
Do more mock exam papers	1	2	3	4
Encourage students to participate in extracurricular activities	1	2	3	4
Teach according to the proficiency test format	1	2	3	4

(10) To what extent have you recommended the following English learning strategies to your students during the past five years? I recommend students..	not at all	some extent	a reasonable extent	a very large extent
To take notes	1	2	3	4
To expose themselves to a range of English media	1	2	3	4
To express their opinions in English in class	1	2	3	4
To put more emphasis on listening	1	2	3	4
To put more emphasis on speaking	1	2	3	4
To put more emphasis on reading	1	2	3	4
To put more emphasis on writing	1	2	3	4
To initiate questions in English	1	2	3	4
To be more active in classroom participation	1	2	3	4
To use English more in their daily life	1	2	3	4

(11) If your colleagues or you use mock tests, what are your main purposes in doing so?	not a major concern at all	of little concern	of great concern	main concern
To give feedback to teachers	1	2	3	4
To assess students' learning difficulties	1	2	3	4
To motivate students	1	2	3	4
To direct students' learning	1	2	3	4
To prepare students for proficiency tests	1	2	3	4
To identify areas for re-teaching	1	2	3	4

(12) What do you think are students' reasons for learning English in Taiwan?	not a major concern at all	of little concern	of great concern	main concern
To pursue further studies	1	2	3	4
To pass examinations	1	2	3	4
To obtain jobs	1	2	3	4
To satisfy school requirements	1	2	3	4
To satisfy parents' requirements	1	2	3	4
To acquire basic knowledge and forms of English	1	2	3	4
To be able to communicate with foreigners	1	2	3	4
To have more and better opportunity in the future	1	2	3	4
To meet the requirements of the society	1	2	3	4

(13) How often is your teaching assessed by your university and what is involved?	never	seldom	sometimes	always	How often?
Your own reflections on teaching	1	2	3	4	_____
The performance of your students in final exams	1	2	3	4	_____
The overall inspection of your students' work by your university	1	2	3	4	_____
Anonymous student evaluation of teaching	1	2	3	4	_____
Evaluation by colleagues	1	2	3	4	_____
Evaluation by the head of department	1	2	3	4	_____
Other _____	1	2	3	4	_____

(14) To what extent do the following factors influence your teaching?	not at all	some extent	a reasonable extent	a very large extent
Professional training	1	2	3	4
Academic seminars or workshops	1	2	3	4
Teaching experience and belief	1	2	3	4
Teaching syllabus	1	2	3	4
Past experience as a language learner	1	2	3	4
The need to obtain satisfaction from teaching	1	2	3	4
Textbooks	1	2	3	4
Proficiency tests	1	2	3	4
Learners' expectations	1	2	3	4
Peers' expectations	1	2	3	4
The head of department's expectations	1	2	3	4

PART THREE Please tick one appropriate answer.

(15) What is your *current* reaction to implementing proficiency tests as a graduation requirement?

- I am sceptical about the policy because _____
- neutral because _____
- I welcome the policy because _____

(16) What is the medium of instruction you use when you teach English in the classroom?

- English only
- English supplemented with occasional Chinese explanation
- Half English and half Chinese
- Mainly Chinese

(17) Who generally makes the decision about lesson content and structure?

- The head of department
- English teachers together
- Yourself
- Other _____

(18) How do you arrange your teaching in your university? Tick more than one box if necessary:

- Following the textbook arrangement
- Following the department scheme of work
- Divided into language activities/tasks
- Topic-based themes

(19) Who makes the major decision on the choice of textbooks?

- The head of department
- English teachers together
- Yourself
- Other _____

Please circle the appropriate answer.

(20) How often do you do the following in class?	almost never happens	uncommon	common	almost whole lesson
Talk to the whole class	1	2	3	4
Talk to groups of students	1	2	3	4
Talk to individual students	1	2	3	4

(21) How often do you do the following activities in class?	seldom or never	sometimes	most classes	every class
Tell the students the objectives of each lesson	1	2	3	4
Demonstrate how to do particular language activities	1	2	3	4
Explain the meaning of the text	1	2	3	4
Explain specific language items such as words or sentences	1	2	3	4
Go through textbook exercises	1	2	3	4
Go through mock exam papers	1	2	3	4
Organise language games	1	2	3	4
Organise pair work or group discussions	1	2	3	4
Organise integrated language skill tasks	1	2	3	4
Talk about proficiency tests	1	2	3	4

(22) How often do you use the following resources in your teaching?	seldom or never	sometimes	most classes	every class
Textbooks	1	2	3	4
Supplementary materials	1	2	3	4
Television	1	2	3	4
Radio	1	2	3	4
Newspapers	1	2	3	4
Magazines	1	2	3	4
Language laboratory	1	2	3	4
Pictures and/or cards	1	2	3	4
Examination syllabus	1	2	3	4
Overall lesson plan	1	2	3	4

- End of Questionnaire -

Thank you very much for your help

3.2 Student questionnaire

A. English version

Students' Perceptions of English Proficiency Tests in Taiwan Technical Universities

Dear students,

I am conducting a small research survey for my doctorate at the University of York. I would like to ask you for your opinions of English learning in the university and also your opinions of English examinations. To help me, please fill in this questionnaire based on your own experience. It should take no more than twenty minutes to complete. **All information you provide here will be treated in the strictest confidence.** Thank you for your valuable time.

Hui-Fen Hsu
University of York
hh524@york.ac.uk

PART ONE Please tick the appropriate answer or provide written answers.

- (1) Your gender: Female Male
- (2) Your grade: First Second Third Fourth
- (3) Your English level: Level one Level two Level three Level four
- (4) What is the medium of instruction your teacher usually uses to teach you English?
- English only
- English supplemented with occasional Chinese explanation
- Half English and half Chinese
- Mainly Chinese
- (5) Besides your English classes, how many hours do you study English per week?
- _____ hour(s)
- (6) How many hours of private tutorials do you undertake for the preparation of English proficiency tests?
- _____ hour(s)

PART TWO Please circle the appropriate answer.

(7) How often does your teacher do the following in a typical English lesson?	almost never happens	uncommon	common	almost whole lesson
Talk to the whole class	1	2	3	4
Talk to groups of students	1	2	3	4
Talk to individual students	1	2	3	4

(8) How often does your teacher do the following activities in your English lessons?	seldom or never	sometimes	most classes	every class
Tell your class the objectives of each lesson	1	2	3	4
Demonstrate how to do particular language activities	1	2	3	4
Explain the meaning of the text	1	2	3	4
Explain specific language items such as words or sentence structures	1	2	3	4
Go through textbook exercises	1	2	3	4
Go through mock exam papers	1	2	3	4
Organise language games	1	2	3	4
Organise group work or discussion	1	2	3	4
Talk about proficiency tests	1	2	3	4

(9) How often does your teacher use the following resources in your English lessons?	seldom or never	sometimes	most classes	every class
Textbooks	1	2	3	4
Supplementary materials	1	2	3	4
Television	1	2	3	4
Radio	1	2	3	4
Newspapers	1	2	3	4
Magazines	1	2	3	4
Language laboratory	1	2	3	4
Pictures and/or cards	1	2	3	4

(10) How often do you do the following <i>in English in class?</i>	seldom or never	sometimes	most classes	every class
Doing Pair work	1	2	3	4
Doing Group discussion	1	2	3	4
Expressing your own ideas	1	2	3	4
Asking questions	1	2	3	4

(11) How often do you do the following <i>in English outside class?</i>	never	uncommon	common	always
Talk to your teachers	1	2	3	4
Talk to your classmates	1	2	3	4
Talk to other people outside university	1	2	3	4
Watch television	1	2	3	4
Listen to the radio	1	2	3	4
Read newspapers and magazines	1	2	3	4

(12) How often do you do the following activities in your English lessons?	seldom or never	sometimes	most classes	every class
Reading	1	2	3	4
Listening	1	2	3	4
Writing	1	2	3	4
Speaking	1	2	3	4
Practising grammar items	1	2	3	4
Learning vocabulary	1	2	3	4
Carrying out group discussion	1	2	3	4
Doing language games	1	2	3	4
Doing mock exam papers	1	2	3	4

Please tick the appropriate answer:

(13) Do you agree with the following reasons for learning English?	<input type="checkbox"/> agree	<input type="checkbox"/> disagree
To acquire basic knowledge and forms of English	<input type="checkbox"/> agree	<input type="checkbox"/> disagree
To get a better job	<input type="checkbox"/> agree	<input type="checkbox"/> disagree
To be able to go into higher education	<input type="checkbox"/> agree	<input type="checkbox"/> disagree
To be able to communicate with foreigners	<input type="checkbox"/> agree	<input type="checkbox"/> disagree
To be able to watch English movies and listen to English programmes	<input type="checkbox"/> agree	<input type="checkbox"/> disagree
To meet the requirements of the society	<input type="checkbox"/> agree	<input type="checkbox"/> disagree
To have more and better opportunity in the future	<input type="checkbox"/> agree	<input type="checkbox"/> disagree
To pass examinations	<input type="checkbox"/> agree	<input type="checkbox"/> disagree
To fulfill my parents' expectations	<input type="checkbox"/> agree	<input type="checkbox"/> disagree
To satisfy school requirements	<input type="checkbox"/> agree	<input type="checkbox"/> disagree

(14) Do you like to use the following strategies for learning English?

- | | | |
|--|-------------------------------|----------------------------------|
| Reading newspapers and magazines in English | <input type="checkbox"/> like | <input type="checkbox"/> dislike |
| Listening to radio programs in English | <input type="checkbox"/> like | <input type="checkbox"/> dislike |
| Watching TV or videos in English | <input type="checkbox"/> like | <input type="checkbox"/> dislike |
| Taking part in group activities in class | <input type="checkbox"/> like | <input type="checkbox"/> dislike |
| Expressing opinions in English in class | <input type="checkbox"/> like | <input type="checkbox"/> dislike |
| Learning by games | <input type="checkbox"/> like | <input type="checkbox"/> dislike |
| Learning grammar rules | <input type="checkbox"/> like | <input type="checkbox"/> dislike |
| Putting more emphasis on learning vocabulary | <input type="checkbox"/> like | <input type="checkbox"/> dislike |
| Doing exercises and homework | <input type="checkbox"/> like | <input type="checkbox"/> dislike |
| Taking notes | <input type="checkbox"/> like | <input type="checkbox"/> dislike |
| Participating in extracurricular activities | <input type="checkbox"/> like | <input type="checkbox"/> dislike |

(15) Have you ever been affected by your exam scores? yes no

(16) In what aspects have you been affected by exam scores?

- | | | |
|----------------------------------|------------------------------|-----------------------------|
| Self-image | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| Motivation to learn | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| Teacher and student relationship | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| Anxiety and emotional tension | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| Future job opportunity | <input type="checkbox"/> yes | <input type="checkbox"/> no |

(17) Do you agree with the following opinions?

- | | | |
|---|--------------------------------|-----------------------------------|
| Students dislike examinations. | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| Students' learning is improved by practising mock exam papers. | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| Taking examinations is a valuable learning experience. | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| Examinations force students to study harder. | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| Examinations have an important effect on student self-image. | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| A student's score on an examination is a good indication of how well she or he has learned the material. | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| A student's performance on an examination is a good indication of how well she or he will be able to apply what has been learned. | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| Examinations should NOT be used as the sole determiner of student grades. | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| Students perform better in an exam situation than in normal teaching situation. | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| Mock examinations are important ways to learn. | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| All students work hard to achieve their best in the proficiency tests. | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| Examination is one of the motivations for students' learning. | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |

Thank you very much for your help

B. Chinese version

台灣科技大學學生對實施英語能力檢定測驗之看法

親愛的學生：

我正爲了我在英國約克大學的博士論文實施研究調查。此問卷在於了解你在學校的英語學習情況和你對英語考試的看法。請根據你本身的經驗來回答。完成這份問卷將不超過二十分鐘。你在這裡所提供的一切資訊均會得到保密。感謝你撥空填寫。

許慧芬
英國約克大學
hh524@york.ac.uk

第一部分 請選擇恰當的答案打勾或填寫適當的答案。

- (1) 你的性別 男 女
- (2) 你的年級 一 二 三 四 其他 _____
- (3) 你的英語課程級數 第一級 第二級 第三級 第四級
- (4) 你的英語老師用哪種語言上課？
 只用英語
 大部分用英語，有時用中文解釋
 一半英語，一半中文
 主要用中文
- (5) 除了學校的英語課，你每週花幾小時讀英語？
每週 _____ 小時
- (6) 爲參加英語能力檢定測驗，你每週上幾小時學校課程外的英語輔導課？
每週 _____ 小時

第二部分 請圈選適當的答案。

(7) 你的老師上英語課時如何面對你們講話？

	幾乎從未發生	不常	經常	幾乎整節課
面對全班同學講話	1	2	3	4
面對部分同學講話	1	2	3	4
面對個別同學講話	1	2	3	4

(8) 你的老師上英語課時多常做以下的活動？

	從不或很少	有時候	大多數的課	每一節課
講述每節課的目的	1	2	3	4
示範如何進行語言活動	1	2	3	4
講解課文涵義	1	2	3	4
講解語言現象，如單詞、句型	1	2	3	4
講解課本習題	1	2	3	4
講解模擬考題	1	2	3	4
組織語言遊戲活動	1	2	3	4
組織小組討論	1	2	3	4
提及英語能力檢定測驗	1	2	3	4

(9) 你的老師多常在英語課中使用以下資源？

	從不或很少	有時候	大多數的課	每一節課
教科書	1	2	3	4
補充教材	1	2	3	4
電視	1	2	3	4
收音機	1	2	3	4
報紙	1	2	3	4
雜誌	1	2	3	4
語言教室	1	2	3	4
圖片或卡片	1	2	3	4

(10) 你多常上課時用英語做以下的活動？

	從不或很少	有時候	大多數的課	每一節課
兩人練習	1	2	3	4
小組討論	1	2	3	4
舉手發表自己的觀點	1	2	3	4
在課堂上提出問題	1	2	3	4

(11) 你多常在課堂外用英語做以下的活動？	從不	不常	經常	總是
和老師用英語講話	1	2	3	4
和同學用英語講話	1	2	3	4
和校外人員用英語講話	1	2	3	4
看英語電視節目	1	2	3	4
聽英語電台節目	1	2	3	4
讀英文報紙或雜誌	1	2	3	4

(12) 你在英語課時多常從事以下的活動？	從不或很少	有時候	大多數的課	每一節課
閱讀	1	2	3	4
聽力	1	2	3	4
寫作	1	2	3	4
會話	1	2	3	4
語法練習	1	2	3	4
詞彙學習	1	2	3	4
小組討論	1	2	3	4
做語言遊戲	1	2	3	4
做模擬考題	1	2	3	4

請選擇適當的答案打勾。

(13) 你同意以下對學習英語的動機看法嗎？	<input type="checkbox"/> 同意	<input type="checkbox"/> 不同意
爲了學習語言的基本知識	<input type="checkbox"/> 同意	<input type="checkbox"/> 不同意
爲了找一份好工作	<input type="checkbox"/> 同意	<input type="checkbox"/> 不同意
爲了能上研究所	<input type="checkbox"/> 同意	<input type="checkbox"/> 不同意
爲了能和外國人用英語交談	<input type="checkbox"/> 同意	<input type="checkbox"/> 不同意
爲了能夠看懂英語電影和聽懂英語電台節目	<input type="checkbox"/> 同意	<input type="checkbox"/> 不同意
爲了適應社會的需求	<input type="checkbox"/> 同意	<input type="checkbox"/> 不同意
爲了將來有更多、更好的發展機會	<input type="checkbox"/> 同意	<input type="checkbox"/> 不同意
爲了通過英語考試	<input type="checkbox"/> 同意	<input type="checkbox"/> 不同意
爲了符合父母的期望	<input type="checkbox"/> 同意	<input type="checkbox"/> 不同意
爲了達到學校的規定	<input type="checkbox"/> 同意	<input type="checkbox"/> 不同意

(14) 你喜歡使用以下的英語學習方法嗎？

- | | | |
|-----------|-----------------------------|------------------------------|
| 閱讀英文報紙和雜誌 | <input type="checkbox"/> 喜歡 | <input type="checkbox"/> 不喜歡 |
| 收聽英語電台節目 | <input type="checkbox"/> 喜歡 | <input type="checkbox"/> 不喜歡 |
| 收看英語電視節目 | <input type="checkbox"/> 喜歡 | <input type="checkbox"/> 不喜歡 |
| 參加英語小組活動 | <input type="checkbox"/> 喜歡 | <input type="checkbox"/> 不喜歡 |
| 用英語交換意見 | <input type="checkbox"/> 喜歡 | <input type="checkbox"/> 不喜歡 |
| 透過遊戲學習 | <input type="checkbox"/> 喜歡 | <input type="checkbox"/> 不喜歡 |
| 學習語法規則 | <input type="checkbox"/> 喜歡 | <input type="checkbox"/> 不喜歡 |
| 將重點放在學習詞彙 | <input type="checkbox"/> 喜歡 | <input type="checkbox"/> 不喜歡 |
| 做練習和作業 | <input type="checkbox"/> 喜歡 | <input type="checkbox"/> 不喜歡 |
| 練習記筆記 | <input type="checkbox"/> 喜歡 | <input type="checkbox"/> 不喜歡 |
| 參加課外英語活動 | <input type="checkbox"/> 喜歡 | <input type="checkbox"/> 不喜歡 |

(15) 你是否受過考試成績的影響？

- 是 否

(16) 在以下哪些方面你受到英語考試成績的影響？

- | | | |
|------------|----------------------------|----------------------------|
| 個人形象 | <input type="checkbox"/> 是 | <input type="checkbox"/> 否 |
| 學習動機 | <input type="checkbox"/> 是 | <input type="checkbox"/> 否 |
| 學生和老師之間的關係 | <input type="checkbox"/> 是 | <input type="checkbox"/> 否 |
| 個人情緒 | <input type="checkbox"/> 是 | <input type="checkbox"/> 否 |
| 將來的就業機會 | <input type="checkbox"/> 是 | <input type="checkbox"/> 否 |

(17) 你同意以下的看法嗎？

- | | | |
|-------------------------|-----------------------------|------------------------------|
| 學生不喜歡考試 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 通過做模擬試題，學生的學習成績有所提高 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 參加考試是一種有益的學習經歷 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 考試迫使學生努力學習 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 考試成績對學生的形象好壞起重大作用 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 一個學生的考試成績反應他是否掌握了學過的知識 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 一個學生的考試成績反應他是否能夠運用所學的知識 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 考試成績不應做為衡量學生學習的唯一標準 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 學生在考試時比在日常學習情況下表現更好 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 模擬考試是重要的學習方法之一 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 大部分學生都力爭在考試中盡可能取得好的成績 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 考試是學生學習的主要動機之一 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |

問卷結束
衷心感謝你的幫助

3.3 Interview schedules

3.3.1 Interviews with teachers

Before lesson interview (lesson plan/teaching notes)

1. What are the objectives of today's lesson?
2. What activities are included?
3. How do you intend to introduce the activities?
4. What materials are used today?
5. Are the activities and/or materials test-oriented?
6. What are the most difficult parts? What might go wrong?

Follow-up interview (for teachers whose students sit for English proficiency tests)

1. Class-based questions (depending on observation)
2. What do you think are the major reasons why the department has made passing English proficiency tests a graduation requirement?
3. What do you think about making passing English proficiency tests a graduation requirement? (What did you expect? Do you think the result is effective? Do you think this innovation should be accepted or rejected?)
4. Does this innovation influence your teaching?
5. (If 'yes' to 4.) How does it influence you? (Prompts: Do you teach differently—method/materials/activities...? Do you teach/not teach certain items...?)
6. (If 'yes' to 4.) How do you feel about test-influence on teaching? (Prompts: Is it good/bad? Does it improve teaching/learning?)
7. Have you observed any changes in your students' behaviour since the implementation of proficiency tests?
8. Would you like to add anything concerning the influence of English proficiency tests?

Follow-up interview (for teachers whose students did not sit for English proficiency tests)

1. Class-based questions (depending on observation)
2. What do you think are the major reasons why the department did not make passing

English proficiency tests a graduation requirement?

3. What do you think about making passing English proficiency tests a graduation requirement? (Do you think it is necessary/unnecessary? Do you think this innovation would be accepted in your university?)
4. If your department decides to bring in passing English proficiency tests as a graduation requirement, will this innovation influence your teaching?
5. If 'yes' to 4.) How will it influence you? (Prompts: Do you teach differently – method/materials/activities...? Do you teach/not teach certain items...?)
6. (If 'yes' to 4.) How do you feel about test-influence on teaching? (Prompts: Is it good/bad? Does it improve teaching/learning?)
7. Do you think making passing English proficiency tests a graduation requirement would change students' behaviour?
8. Would you like to add anything concerning the influence of proficiency tests?

3.3.2 Interviews with advisory committee member in the Ministry of Education

1. Why did the Ministry of Education make the decision to develop the project on The Enhancement of Students' Foreign Language Proficiency?
2. Why did the Ministry of Education make the decision to include the implementation of English proficiency tests in the project?
3. What do you think are the reasons why some technical universities have made passing English proficiency tests as a graduation requirement while others have not?
4. What do you think about making English proficiency tests a graduation requirement? (What did you expect? Do you think the result is effective? Are you monitoring the effect/reviewing it? Have they had negative feedback? Do you think it is necessary to make English proficiency tests a graduation requirement?)
5. Would you like to add anything else concerning the implementation of English proficiency tests as a graduation requirement?

3.3.3 Interviews with departmental directors

1. What do you think are the major reasons why the Ministry of Education introduced the project on The Enhancement of Students' Foreign Language Proficiency?
2. Why did/didn't the department make the decision to make passing English proficiency tests a graduation requirement? (short term/long term plan; funding; resources)
3. (If implementing) What do you think about the effectiveness of the implementation of English proficiency tests as a graduation requirement? (What did you expect? examples of success and/or barriers; Do you think this innovation should be accepted or rejected?)
4. (If not implementing) Does the department plan to bring in passing English proficiency tests as a graduation requirement in the future? Why?
5. Would you like to add anything else concerning the implementation of English proficiency tests as a graduation requirement?

Appendix 4 Questionnaires and interview schedules for the main study

4.1 Teacher questionnaire

Teachers' Perceptions of English Proficiency Tests in Taiwan Universities of Technology

Dear Teacher,

I am conducting a small research survey for my doctorate at the University of York in the United Kingdom. I would like to ask you for your opinions of General English teaching and English proficiency tests (TOEFL, TOEIC, IELTS, GEPT, and CSEPT). To help me, please fill in this questionnaire based on your own experience. It should take no more than twenty minutes to complete. *All information will be anonymous and treated in the strictest confidence.* Thank you very much.

If you would like a summary of the final report, please e-mail me:

Hui-Fen Hsu
University of York
hh524@york.ac.uk

PART ONE Please tick one appropriate answer or provide written answers.

- (1) Your gender: Female Male
- (2) Your age: 21-30 31-40 41-50 above 50
- (3) Your highest education level: Bachelor Masters PhD/EdD
- (4) Number of years you have been teaching General English in universities of technology:
 1-3 4-6 7-9 Other _____
- (5) Number of periods a week you teach General English to non-English majors in this university:
 3 – 6 7 – 10 11 – 14 Other _____

PART TWO Please circle/tick the appropriate answer for each item or provide written answers.

(6) What do you think are university students' reasons for learning English in Taiwan?	not a major concern at all	of little concern	of great concern	main concern
To pursue further studies	1	2	3	4
To pass examinations	1	2	3	4
To obtain a 'good' job	1	2	3	4
To satisfy university requirements	1	2	3	4

(7) Is passing an English proficiency test a graduation requirement for non-English majors in this university? yes **go to Q8**
 no **go to Q9**

(8) What do you see as the major aims of the university in making this regulation?	not a major concern at all	of little concern	of great concern	main concern
To meet the government regulation	1	2	3	4
To prepare students for their future career	1	2	3	4
To cope with the present decline in English standards	1	2	3	4
To motivate students to use integrated skills	1	2	3	4
To direct students' learning	1	2	3	4
To enable students to communicate more with foreigners	1	2	3	4
To encourage better textbooks	1	2	3	4
To assess teachers through students' performance	1	2	3	4

Go to Q10

(9) Can you summarize the reasons why the university decided not to make the regulation?

(10) To what extent have you made the following changes in your General English lessons to non-English majors during the past five years?

I now ...

	not at all	some extent	a reasonable extent	a very large extent
Employ new teaching methods	1	2	3	4
Use a more communicative approach in teaching	1	2	3	4
Put more stress on group discussion	1	2	3	4
Put more emphasis on listening	1	2	3	4
Put more emphasis on speaking	1	2	3	4
Put more emphasis on reading	1	2	3	4
Put more emphasis on writing	1	2	3	4
Put more emphasis on the integration of skills	1	2	3	4
Encourage more student participation in class	1	2	3	4
Employ more real life language tasks	1	2	3	4
Do more mock exam papers	1	2	3	4
Teach according to English proficiency test format	1	2	3	4
Other comments _____				

(11) To what extent have you recommended the following English learning strategies to your non-English major students during the past five years? I recommend students...

	not at all	some extent	a reasonable extent	a very large extent
To take notes	1	2	3	4
To expose themselves to a range of English media	1	2	3	4
To express their opinions in English in class	1	2	3	4
To put more emphasis on listening	1	2	3	4
To put more emphasis on speaking	1	2	3	4
To put more emphasis on reading	1	2	3	4
To put more emphasis on writing	1	2	3	4
To use English more in their daily life	1	2	3	4

(12) To what extent do the following factors influence your General English teaching?	not at all	some extent	a reasonable extent	a very large extent
Professional training received	1	2	3	4
Academic seminars or workshops taken	1	2	3	4
Teaching experience and beliefs	1	2	3	4
Past experience as a language learner	1	2	3	4
The need to obtain satisfaction from teaching	1	2	3	4
Textbooks	1	2	3	4
English proficiency tests	1	2	3	4
Learners' expectations	1	2	3	4

(13) On average, how many English proficiency mock exams do your non-English major students take every semester, either in class or in the self-access centre?

- 0 1–3 4–6 7 or more

PART THREE Please tick one appropriate answer or provide written answers.

(14) What is your *current* reaction to passing an English proficiency test as a graduation requirement for non-English majors?

- I am sceptical about the policy because _____
- neutral because _____
- I welcome the policy because _____

(15) How do you arrange your General English teaching to non-English majors in this university? Tick more than one box if necessary:

- Following a General English textbook's arrangement
- Following English proficiency test-oriented materials
- Divided into separate skills such as reading or listening

(16) Have you chosen textbooks for your General English classes to non-English majors because of English proficiency tests?

- yes no some, please give details _____

(17) What is the medium of instruction you use when you teach General English to non-English majors?

- English only
- English supplemented with occasional Chinese explanation
- Half English and half Chinese
- Mainly Chinese

Please circle the appropriate answer for each item.

(18) How often do you do the following in your General English classes?

	almost never happens	uncommon	common	almost whole lesson
--	-----------------------------	-----------------	---------------	----------------------------

Talk to the whole class	1	2	3	4
Talk to groups of students	1	2	3	4
Talk to individual students	1	2	3	4

(19) How often do you do the following activities in your General English classes?

	seldom or never	sometimes	most classes	every class
--	------------------------	------------------	---------------------	--------------------

Explain the meaning of the text	1	2	3	4
Explain specific language items such as words or sentences	1	2	3	4
Go through textbook exercises	1	2	3	4
Go through mock exam papers	1	2	3	4
Organise pair work or group discussions	1	2	3	4
Organise integrated language skill tasks	1	2	3	4
Talk about English proficiency tests	1	2	3	4

(20) How often do you use the following resources in your General English classes?

	seldom or never	sometimes	most classes	every class
--	------------------------	------------------	---------------------	--------------------

Textbooks	1	2	3	4
English proficiency test-oriented materials	1	2	3	4
Television	1	2	3	4
Radio	1	2	3	4
Newspapers	1	2	3	4
Magazines	1	2	3	4
Language laboratory	1	2	3	4
Pictures and/or cards	1	2	3	4

Thank you very much for taking the time to complete this questionnaire

4.2 Student questionnaire

A. English version

Students' Perceptions of English Proficiency Tests in Taiwan Universities of Technology

Dear students,

I am conducting a small research survey for my doctorate at the University of York in the United Kingdom. I would like to ask you for your opinions of English learning in the university and also your opinions of English proficiency tests (TOEFL, TOEIC, IELTS, GEPT, and CSEPT). To help me, please fill in this questionnaire based on your own experience. It should take no more than fifteen minutes to complete. **All information will be anonymous and treated in the strictest confidence.** Thank you very much.

Hui-Fen Hsu
University of York
hh524@york.ac.uk

PART ONE Please tick one appropriate answer or provide written answers.

- (1) Your gender: Female Male
- (2) Year of study: First Second Third Fourth
- (3) Your major: _____
- (4) What is the medium of instruction your teacher usually uses to teach you English?
- English only
- English supplemented with occasional Chinese explanation
- Half English and half Chinese
- Mainly Chinese
- (5) Besides your required English classes, how many hours a week do you study English?
- 0 1-2 3-4 Other _____
- (6) Do you have to take an English proficiency test in order to obtain the first degree?
- yes no
- (7) Do you take a course or courses in the university or in a cram school for the preparation of English proficiency tests?
- yes no

Turn the page

PART TWO Please circle the appropriate answer for each item.

(8) How often does your teacher do the following in a typical English lesson?

	almost never happens	uncommon	common	almost whole lesson
--	-----------------------------	-----------------	---------------	----------------------------

Talk to the whole class	1	2	3	4
Talk to groups of students	1	2	3	4
Talk to individual students	1	2	3	4

(9) How often does your teacher do the following activities in your English lessons?

	seldom or never	sometimes	most classes	every class
--	------------------------	------------------	---------------------	--------------------

Explain the meaning of the text	1	2	3	4
Explain specific language items such as words or sentence structures	1	2	3	4
Go through textbook exercises	1	2	3	4
Go through mock exam papers	1	2	3	4
Talk about English proficiency tests, including TOEFL, TOEIC, IELTS, GEPT, and CSEPT	1	2	3	4

(10) How often do you do the following activities in your English lessons?

	seldom or never	sometimes	most classes	every class
--	------------------------	------------------	---------------------	--------------------

Reading	1	2	3	4
Listening	1	2	3	4
Writing	1	2	3	4
Speaking	1	2	3	4
Practising grammar items	1	2	3	4
Learning vocabulary	1	2	3	4
Carrying out pair work	1	2	3	4
Carrying out group discussions	1	2	3	4
Doing language games	1	2	3	4
Doing English proficiency mock exams in class or in the self-assess centre	1	2	3	4

(11) How often do you do the following *in English in class*?

	seldom or never	sometimes	most classes	every class
--	------------------------	------------------	---------------------	--------------------

Doing pair work	1	2	3	4
Holding group discussions	1	2	3	4
Expressing your own ideas	1	2	3	4
Asking the teacher questions	1	2	3	4
Asking a fellow student questions	1	2	3	4

Turn the page

(12) How often do you do the following <i>in English outside class?</i>	never	uncommon	common	always
Talk to your teachers	1	2	3	4
Talk to your classmates	1	2	3	4
Talk to other people outside university	1	2	3	4
Watch television or movies	1	2	3	4
Listen to the radio or CDs	1	2	3	4
Read newspapers and magazines	1	2	3	4
Do online English proficiency mock exams	1	2	3	4

Please tick the appropriate answer for each item.

(13) What are your main reasons for learning English?

- | | | |
|---|--------------------------------|-----------------------------------|
| To be able to make a success of your career | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| To obtain a 'good' job | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| To pursue further studies beyond BSc/BA | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| To be able to communicate with foreigners | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| To be able to watch English movies and listen to English radio programmes | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| To satisfy university requirements | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| To pass examinations | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| To meet the expectations of society, please give details _____ | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
-

(14) Do you use the following strategies for learning English?

- | | | |
|---|------------------------------|-----------------------------|
| Reading English newspapers and magazines | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| Listening to English radio programmes and CDs | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| Watching TV or movies in English | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| Taking part in group activities in class | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| Expressing opinions in English in class | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| Besides required English lessons, taking optional English lessons | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| Putting more emphasis on learning grammar rules | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| Putting more emphasis on learning vocabulary | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| Doing exercises and homework | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| Taking notes | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| Practising English writing | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| Participating in extracurricular activities | <input type="checkbox"/> yes | <input type="checkbox"/> no |

(15) Have you ever been affected by English proficiency tests, including TOEFL, TOEIC, IELTS, GEPT, and CSEPT?

<input type="checkbox"/> yes	go to Q16
<input type="checkbox"/> no	go to Q17

(16) How have you been affected by English proficiency tests?

- | | | |
|--|------------------------------|-----------------------------|
| Self-image, please explain _____ | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| More interested in learning English | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| Less interested in learning English | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| Clearer on learning objectives | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| Anxiety and emotional tension | <input type="checkbox"/> yes | <input type="checkbox"/> no |
| Future job opportunity, please explain _____ | <input type="checkbox"/> yes | <input type="checkbox"/> no |

(17) What are your views on English exams, including the English proficiency tests, currently being administered in the university?

- | | | |
|---|--------------------------------|-----------------------------------|
| Examinations increase students' motivation to learn English. | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| Students' learning is improved by practising mock exams. | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| Taking examinations is a valuable learning experience. | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| Examinations force students to study harder. | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| A student's score on an examination is a good indication of how well she or he has learned the material. | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| A student's performance on an examination is a good indication of how well she or he will be able to apply what has been learned. | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| Examinations should NOT be used as the sole determiner of student grades. | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| Mock examinations are important ways to learn. | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |
| Every university graduating student should pass at least one English proficiency test. Please explain _____ | <input type="checkbox"/> agree | <input type="checkbox"/> disagree |

Thank you very much for your help

B. Chinese version

台灣科技大學學生對實施英語能力檢定測驗之看法

親愛的同學：

我正爲了我在英國約克大學的博士論文進行研究調查。此問卷在於了解你在學校的英語學習情況和你對英語能力檢定測驗 [包含托福 (TOEFL)、多益 (TOEIC)、國際英語測驗 (IELTS)、全民英檢(GEPT)，大學校院英語能力測驗 (CSEPT)，和學校自辦英語檢定測驗] 的看法。請根據你本身的經驗來回答。完成這份問卷將不超過十五分鐘。你在這裡所提供的一切資訊均會得到保密。感謝你撥空填寫。

許慧芬
英國約克大學
hh524@york.ac.uk

第一部分 請選擇恰當的答案打勾或填寫適當的答案。

- (1) 你的性別 女 男
- (2) 你的年級 一 二 三 四
- (3) 你的主修科系 _____
- (4) 你的英語老師用哪種語言上課？
 - 只用英語
 - 大部分用英語，有時用中文解釋
 - 一半英語，一半中文
 - 主要用中文
- (5) 除了學校的英語課，你每週花幾小時讀英語？
 - 0 1 - 2 3 - 4 其他 _____
- (6) 你的學校有實施英語畢業門檻嗎？
 - 有 沒有
- (7) 你有爲了參加英語能力檢定測驗(以下簡稱英檢)而參加校內的準備課程或到校外補習嗎？
 - 有 沒有

請翻頁

第二部分 請為每個項目圈選適當的答案。

(8) 你的老師上英語課時如何面對你們講話？	幾乎從未發生	不常	經常	幾乎整節課
面對全班同學講話	1	2	3	4
面對部分同學講話	1	2	3	4
面對個別同學講話	1	2	3	4

(9) 你的老師上英語課時多常做以下的活動？	從不或很少	有時候	大多數的課	每一節課
講解課文涵義	1	2	3	4
講解語言現象，如單字、文法句型	1	2	3	4
講解課本習題	1	2	3	4
講解英檢模擬試題	1	2	3	4
提到英檢，包含托福 (TOEFL)、多益 (TOEIC)、國際英語測驗 (IELTS)、全民英檢(GEPT)，大學校院英語能力測驗 (CSEPT)，和學校自辦英語檢定測驗	1	2	3	4

(10) 你在英語課時多常從事以下的活動？	從不或很少	有時候	大多數的課	每一節課
閱讀	1	2	3	4
聽力	1	2	3	4
寫作	1	2	3	4
會話	1	2	3	4
文法練習	1	2	3	4
詞彙學習	1	2	3	4
兩人一組練習	1	2	3	4
小組討論	1	2	3	4
做語言遊戲	1	2	3	4
在課堂上或自學中心做英檢模擬試題	1	2	3	4

(11) 你多常上課時用英語做以下的活動？	從不或很少	有時候	大多數的課	每一節課
與同學做兩人一組練習	1	2	3	4
小組討論	1	2	3	4
舉手發表自己的觀點	1	2	3	4
對老師提出問題	1	2	3	4
對同學提出問題	1	2	3	4

請翻頁

(12) 你多常在課堂外用英語做以下的活動？	從不	不常	經常	總是
和老師講話	1	2	3	4
和同學講話或練習會話	1	2	3	4
和校外人員講話	1	2	3	4
看電視節目或電影	1	2	3	4
聽廣播或光碟	1	2	3	4
讀課外讀物，例如英語報紙、雜誌	1	2	3	4
上英檢網站練習模擬試題	1	2	3	4

請為每個項目選擇適當的答案打勾。

(13) 你為什麼要學習英語？

- | | | |
|----------------------|-----------------------------|------------------------------|
| 為了在專業領域成功 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 為了找一份好工作 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 為了能上研究所 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 為了能和外國人用英語交談 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 為了能夠看懂英語電影和聽懂英語廣播 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 為了達到學校的規定，例如英語畢業門檻 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 為了通過英語考試 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 為了適應社會的需求。請解釋： _____ | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |

(14) 你使用以下的英語學習方法嗎？

- | | | |
|------------------|----------------------------|----------------------------|
| 閱讀英語報紙或雜誌 | <input type="checkbox"/> 是 | <input type="checkbox"/> 否 |
| 收聽英語廣播或光碟 | <input type="checkbox"/> 是 | <input type="checkbox"/> 否 |
| 收看英語電視節目或電影 | <input type="checkbox"/> 是 | <input type="checkbox"/> 否 |
| 參與課堂小組活動 | <input type="checkbox"/> 是 | <input type="checkbox"/> 否 |
| 用英語表達意見 | <input type="checkbox"/> 是 | <input type="checkbox"/> 否 |
| 除了必修課程，多選修英語相關課程 | <input type="checkbox"/> 是 | <input type="checkbox"/> 否 |
| 將重點放在學習文法規則 | <input type="checkbox"/> 是 | <input type="checkbox"/> 否 |
| 將重點放在學習詞彙 | <input type="checkbox"/> 是 | <input type="checkbox"/> 否 |
| 做練習和作業 | <input type="checkbox"/> 是 | <input type="checkbox"/> 否 |
| 練習記筆記 | <input type="checkbox"/> 是 | <input type="checkbox"/> 否 |
| 練習英語寫作 | <input type="checkbox"/> 是 | <input type="checkbox"/> 否 |
| 參加課外英語活動 | <input type="checkbox"/> 是 | <input type="checkbox"/> 否 |

(15) 你是否受過英檢 [包含托福 (TOEFL)、多益 (TOEIC)、國際英語測驗 (IELTS)、全民英檢 (GEPT)，和大學校院英語能力測驗 (CSEPT)，和學校自辦英語檢定測驗] 的影響？

是 → 請到 16 題並繼續

否 → 請到 17 題

(16) 在以下哪些方面你受到英檢的影響？

- | | | |
|-------------------|----------------------------|----------------------------|
| 個人形象好壞，請解釋 _____ | <input type="checkbox"/> 是 | <input type="checkbox"/> 否 |
| 對學習英語更有興趣 | <input type="checkbox"/> 是 | <input type="checkbox"/> 否 |
| 對學習英語變得沒有興趣 | <input type="checkbox"/> 是 | <input type="checkbox"/> 否 |
| 對學習英語有更清楚的目標 | <input type="checkbox"/> 是 | <input type="checkbox"/> 否 |
| 造成情緒緊張及焦慮 | <input type="checkbox"/> 是 | <input type="checkbox"/> 否 |
| 將來的就業機會，請解釋 _____ | <input type="checkbox"/> 是 | <input type="checkbox"/> 否 |

(17) 你對目前學校中實施的英語考試（包含期中期末考及英檢）有何看法？

- | | | |
|-------------------------|-----------------------------|------------------------------|
| 考試能夠提升學生學習動機 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 通過做英語模擬試題，學生的學習成績有所提高 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 參加英語考試是一種有益的學習經歷 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 考試迫使學生努力學習 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 一個學生的考試成績反應他是否掌握了學過的知識 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 一個學生的考試成績反應他是否能夠運用所學的知識 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 考試成績不應做為衡量學生學習的唯一標準 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 模擬考試是重要的學習方法之一 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 每一個大學畢業生都應該通過至少一項英檢。 | <input type="checkbox"/> 同意 | <input type="checkbox"/> 不同意 |
| 請解釋 _____ | | |

問卷結束
衷心感謝你的幫助

4.3 Interview schedules

4.3.1 Interviews with teachers

Follow-up interview (for teachers whose students sit for English proficiency tests)

1. Class-based questions	(depending on observation)
2. What do you think are the major reasons why the department has made passing English proficiency tests a graduation requirement?	<ul style="list-style-type: none"> ● How do you obtain the information about the new policy? ● Have you attended any seminar or workshop related to the new policy before or after it is imposed? ● How do you obtain the information about the English proficiency tests? ● Have you attended any seminar or workshop related to the English proficiency tests before or after the regulation is imposed?
3. What do you think about making passing English proficiency tests a graduation requirement?	<ul style="list-style-type: none"> ● What are the good points that you see in it? ● What are the weak points that you see in it?
4. Does this new policy influence your teaching?	
5. (If 'yes' to 4.) How does it influence you?	<ul style="list-style-type: none"> ● Do you use different textbooks?(test-oriented) ● Who decided to change textbooks? ● Do you use different teaching methods? ● Do you use test format practices in teaching? ● Do you teach different activities? ● Do you teach or not teach certain items ...?
6. (If 'yes' to 4.) How do you feel about test-influence on teaching?	<ul style="list-style-type: none"> ● Is it good/bad? ● Does it improve teaching/learning? ● Does it put extra pressure on you?
7. Do you think the regulation has increased or decreased students' motivation to learn? Why? How?	
8. Do you think this regulation should be implemented or rejected? Why?	
9. Would you like to add anything concerning the influence of English proficiency tests?	

Follow-up interview (for teachers whose students did not sit for English proficiency tests)

1. Class-based questions	(depending on observation)
2. What do you think are the major reasons why the department did not make passing English proficiency tests a graduation requirement?	
3. What do you think about making passing English proficiency tests a graduation requirement?	<ul style="list-style-type: none"> ● What are the good points that you see in it? ● What are the weak points that you see in it?
4. Do you think this policy should be accepted or rejected? Why?	
5. If your department decides to make passing English proficiency tests a graduation requirement, will this innovation affect your teaching?	
6. (If 'yes' to 5.) How will it affect you?	<ul style="list-style-type: none"> ● Do you use different textbooks?(test-oriented) ● Do you use different teaching methods? ● Do you use test format practices in teaching? ● Do you teach different activities? ● Do you teach or not teach certain items ...?
7. (If 'yes' to 5.) How do you feel about test-influence on teaching?	<ul style="list-style-type: none"> ● Is it good/bad? ● Does it improve teaching/learning? ● Does it put extra pressure on you?
8. Do you think the regulation will increase or decrease students' motivation to learn? Why? How?	
9. Would you like to add anything concerning the influence of English proficiency tests?	

4.3.2 Interviews with advisory committee members within the Ministry of Education

<p>1. Why did the Ministry of Education decide to introduce the project on The Enhancement of Students' Foreign Language Proficiency in universities of technology?</p>	<ul style="list-style-type: none"> ● What did the Ministry of Education expect?
<p>2. Why did the Ministry of Education decide to include the implementation of English proficiency tests in the project?</p>	<ul style="list-style-type: none"> ● What did the Ministry of Education expect universities of technology to do? ● What did the Ministry of Education expect teachers to do?
<p>3. What do you think are the reasons why some universities of technology have brought in passing English proficiency tests as a graduation requirement for non-English majors, while others have not?</p>	<p>Based on</p> <ul style="list-style-type: none"> ● Short-term plan ● Long-term plan ● Funding ● Resources ● Teachers' expectations ● Students' expectations ● Students' English language level
<p>4. What do you think about imposing passing English proficiency tests as a graduation requirement for non-English majors?</p>	<ul style="list-style-type: none"> ● What are the good points that you see in it? ● What are the weak points that you see in it? ● Has the regulation been implemented as you expect? ● Is it necessary to make passing English proficiency tests a graduation requirement for non-English majors? ● Is the result effective so far?
<p>5. Do you think this regulation should be implemented or rejected? Why?</p>	
<p>6. Would you like to add anything concerning the regulation for non-English majors?</p>	

4.3.3 Interviews with departmental directors

<p>1. What do you think are the major reasons why the Ministry of Education introduced the project on The Enhancement of Students' Foreign Language Proficiency in universities of technology?</p>	<ul style="list-style-type: none"> ● What does the Ministry of Education expect universities of technology to do?
<p>2. Why did/didn't your university decide to bring in passing English proficiency tests as a graduation requirement for non-English majors?</p>	<p>Based on</p> <ul style="list-style-type: none"> ● Short-term plan ● Long-term plan ● Funding ● Resources ● Teachers' expectations ● Students' expectations and/or English level
<p>3. (If not implementing) Does the department plan to bring in passing English proficiency tests as a graduation requirement for non-English majors in the future? Why?</p>	
<p>4. (If implementing) Has your university introduced any innovation because of the implementation of English proficiency tests in the university?</p>	<ul style="list-style-type: none"> ● New courses (test preparation courses) ● Remedial courses ● New resources ● New teaching methods ● Seminar or workshop
<p>5. What do you expect teachers to do with the regulation?</p>	<ul style="list-style-type: none"> ● New teaching methods ● New materials
<p>6. What will happen if students fail to meet the regulation?</p>	
<p>7. What do you think about bringing in passing English proficiency tests as a graduation requirement?</p>	<ul style="list-style-type: none"> ● What are the good points that you see in it? ● What are the weak points that you see in it?
<p>8. (If implementing) Do you think this regulation should be implemented or rejected? Why?</p>	
<p>9. Would you like to add anything concerning the regulation for non-English majors?</p>	

References

- Aiken, L. R. (1996). *Rating scales and checklists*. New York: John Wiley and Sons.
- Alderson, J. C. (2007). The CEFR and the need for more research. *Modern Language Journal*, 91(4), 659-663.
- Alderson, J. C., & Banerjee, J. (2001). Language testing and assessment (Part 1). *Language Testing*, 34, 213-236.
- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: a study of washback. *Language Testing*, 13(3), 280-297.
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129.
- Allen, P., Fröhlich, M., & Spada, N. (1984). The communicative orientation of language teaching: An observation scheme. In J. Handscombe, R. A. Oren, & B. P. Taylor (Eds.), *On TESOL '83: The question of control* (pp. 231-252). Washington, DC: TESOL.
- Allwright, D. (1988). *Observation in the language classroom*. London: Longman.
- Allwright, D., & Bailey, K. (1991). *Focus on the language classroom*. Cambridge: Cambridge University Press.
- Anastas, J. W. (2005). Observation. In R. M. Grinnell Jr., & Y. A. Unrau (Eds.), *Social work research and evaluation: Quantitative and qualitative approaches* (7th ed., pp. 213-230). New York: Oxford University Press.
- Andrews, S. (1994). The washback effect of examinations: Its impact upon curriculum innovation in English language teaching. *Curriculum Forum*, 4(1), 44-58.
- Andrews, S. (2004). Washback and curriculum innovation. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 37-52). Mahwah, NJ: Lawrence Erlbaum Associates.
- Andrews, S., & Fullilove, J. (1994). Assessing spoken English in public examinations – Why and how? In J. Boyle, & P. Falvey (Eds.), *English language testing in Hong Kong* (pp. 57-86). Hong Kong: The Chinese University Press.

- Andrews, S., Fullilove, J., & Wong, Y. (2002). Targeting washback – a case-study. *System*, 30, 207-223.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., Davidson, F., Ryan, K., & Choi, I-C. (1995). *An investigation into the comparability of two tests of English as a foreign language*. Cambridge: Cambridge University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Baily, K. M. (1985). Classroom-centred research on language teaching and learning. In M. Celce-Murcia (Ed.), *Beyond basics: Issues and research in TESOL* (pp. 96-121). Rowley, MA: Newbury House.
- Bailey, K. M. (1992). The process of innovation in language teacher development: What, why and how teachers change. In J. Flowerdew, M. Brock, & S. Hsia (Eds.), *Perspectives on second language teacher education* (pp. 253-282). Hong Kong: City Polytechnic.
- 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. (1999a). *Washback in language testing*. Princeton, PA: Educational Testing Service.
- Bailey, K. M. (1999b). *What have we learned in twenty years of classroom research?* Featured speaker presentation at the March 1999 TESOL Convention, New York.
- Baker, E. L. (1991, September). *Issues in policy, assessment, and equity*. Paper presented at the national research symposium on limited English proficient students' issues: Focus on evaluation and measurement, Washington, D.C.
- Berman, P. (1981). Educational Change: An implementation paradigm. In R. Lehming, & M. Kane (Eds.), *Improving schools: Using what we know* (pp. 253-286). London: Sage.
- Berman, P., & McLaughlin, M. W. (1976). Implementation of educational innovation. *The Educational Forum*, 40, 345-370.

- Berry, V., & Lewkowicz, J. (2000). Exit-tests: Is there an alternative? *Hong Kong Journal of Applied Linguistics*, 5(1), 19-49.
- Berry, V., & McNeill, A. (2005). Raising English language standards in Hong Kong. *Language Policy*, 4, 371-394.
- Biggs, J. B. (Ed.). (1996). *Testing: To educate or to select? Education in Hong Kong at the cross-roads*. Hong Kong: Hong Kong Educational Publishing.
- Bishop, J. H., & Mane, F. (1999). *The New York state reform strategy: the incentive effects of minimum competency exams*. Philadelphia: Consortium for Policy Research in Education.
- Blumer, H. (1986). *Symbolic interactionism: Perspective and method*. Berkeley, CA: University of California Press.
- Bonnet, G. (2007). The CEFR and education policies in Europe. *Modern Language Journal*, 91(4), 669-672.
- Brannen, J. (1992). Combining qualitative and quantitative approaches: An overview. In J. Branner (Ed.), *Mixing methods: Qualitative and quantitative research* (pp. 3-37). Aldershot: Avebury.
- Brindley, G. (2008). Educational reform and language testing. In E. Shohamy, & N. H. Hornberger (Eds.), *Encyclopedia of language and education: Vol.7. Language testing and assessment* (pp. 365-378). New York: Springer.
- Brindley, G., & Hood, S. (1990). Curriculum innovation in adult ESL. In G. Brindley (Ed.), *The second language curriculum in action* (pp. 232-248). Sydney: National Centre for English Language Teaching and Research.
- British Educational Research Association (BERA).(2004). *Revised ethical guidelines for educational research*. Retrieved on October 28, 2009, from <http://www.bera.ac.uk>
- Broadfoot, P. (1996). *Education, assessment and society: a sociological analysis*. Buckingham: Open University Press.
- Brown, G. (1975). *Microteaching*. London: Methuen.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. New York:

Longman.

- Brown, J. D. (2001). *Using surveys in language programs*. Cambridge: Cambridge University Press.
- Bryman, A. (1992). Quantitative and qualitative research: Further reflections on their integration. In J. Branner (Ed.), *Mixing methods: Qualitative and quantitative research* (pp. 57-80). Aldershot: Avebury.
- Buck, G. (1988). Testing listening comprehension in Japanese university entrance examinations. *JALT Journal*, 10, 15-42.
- Burrows, C. (2004). Washback in classroom-based assessment: A study of the washback effect in the Australian adult migrant English program. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 113-128). Mahwah, NJ: Lawrence Erlbaum Associates.
- Byrnes, H. (2007). Developing national language education policies: Reflections on the CEFR. *Modern Language Journal*, 91(4), 679-685.
- Calder, P. (1997). *Impact of Alberta achievement tests on the teaching-learning process*. Admonition, Alberta, Canada: Alberta Teacher Association.
- Cannell, J. J. (1987). Nationally-normed elementary achievement testing in America's public schools: How all 50 states are above the national average. *Educational Measurement: Issues and Practice*, 7(4), 12-15.
- Carless, D. R. (1999). Large-scale curriculum change in Hong Kong. In C. Kennedy, P. Doyle, & C. Goh (Eds.), *Exploring change in English language teaching* (pp. 19-29). Oxford: Macmillan Heinemann English Language Teaching.
- Chang, Y-P. (2003, January 6). Education in decline, says report.. *Taipei Times*, p. 2.
- Chapman, D. W., & Snyder, C. W. (2000). Can high stakes national testing improve instruction: reexamining conventional wisdom. *International Journal of Educational Development*, 20, 457-474.
- Chaudron, C. (1988). *Second language classrooms*. Cambridge: Cambridge University Press.
- Chaudron, C. (2000). Contrasting approaches to classroom research: Qualitative and

- quantitative analysis of language use and learning. *Second Language Studies*, 19(1), 1-56.
- Chen, L. (2002). *Washback of a public exam on English teaching*. East Lansing, MI: National Center for Research on Teacher Learning. (ERIC Document Reproduction Service No. ED472167)
- Chen, M., & Johnson, D. (2004). Graduation English language proficiency benchmark in Taiwan – Issues and problems. *Proceedings of the 2004 International Conference and Workshops on TEFL & Applied Linguistics* (pp. 136-147). Taipei: Minchuan University.
- Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, 11(1), 38-54.
- Cheng, L. (1998). Impact of a public English examination change on students' perceptions and attitudes toward their English learning. *Studies in Educational Evaluation*, 24, 279-301.
- Cheng, L. (1999). Changing assessment: Washback on teacher perspectives and actions. *Teaching and Teacher Education*, 15, 253-271.
- Cheng, L. (2004). The washback effect of a public examination change on teachers' perceptions toward their classroom teaching. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: research contexts and methods* (pp. 147-170). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. Cambridge: Cambridge University Press.
- Cheng, L. (2008). Washback, impact and consequences. In E. Shohamy, & N. H. Hornberger (Eds.), *Encyclopedia of language and education: Vol. 7. Language testing and assessment* (pp. 349-364). New York: Springer.
- Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 3-18). Mahwah, NJ: Lawrence Erlbaum Associates.

- Cheng, L., Watanabe, Y., & Curtis, A. (Eds.). (2004). *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Chern, C. L. (2002). English language teaching in Taiwan today. *Asia Pacific Journal of Education*, 22(2), 97-105.
- Cimbricz, S. (2002). State-mandated testing and teachers' beliefs and practice. *Education Policy Analysis Archives*, 10(2). Retrieved February 22, 2008, from <http://epaa.asu.edu/epaa/v10n2.html>
- Clark, N. (2002). Education in Taiwan. *World Education News and Reviews*, 15(6). Retrieved February 22, 2008, from <http://www.wes.org/ewenr/02nov/Practical.htm>
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). London: Routledge.
- Coleman, J. A. (2006). English-medium teaching in European higher education. *Language Teaching*, 39, 1-14.
- Cooper, R. (1989). *Language planning and social change*. Cambridge: Cambridge University Press.
- Corbet, H. D., & Wilson, B. (1991). *Testing, rebellion and reform*. Norwood, NJ: Ablex.
- Coulthard, M. (1977). *An introduction to discourse analysis*. London: Longman.
- Council for Economic Planning and Development, Taiwan. (2002). *Challenge 2008 – National development plan: English summary*. Retrieved November 22, 2006, from Taipei: Executive Yuan <http://www.ey.gov.tw/mp?mp=1>
- Council on Education Reform, Taiwan. (1996). *The overall report on education reform consultation*. Retrieved July 19, 2009, from <http://www.edu.tw>
- Council of Europe. (2001). *Common European framework of reference: Learning, teaching and assessment*. Cambridge: Cambridge University Press.
- Council of Europe. (2006). *Reference level descriptions (RLD) for national and regional languages*. Retrieved July 20, 2009, from http://www.coe.int/t/dg4/linguistic/dnr_EN.asp
- Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks: Sage.

- Crombie, W. (2006). Taiwan and New Zealand: Some critical issues in language policy and planning. In Department of English Wenzao Ursuline College of Languages (Ed.), *The 23rd international conference on English teaching and learning in the Republic of China – Changes, chances, and challenges in English teaching & learning* (vol.1, pp.1-11). Taipei: Kaun Tang International.
- Crump, S. (1999). E-ducation: Electronic, emotionless and efficient? *Journal of Education Policy*, 14(6), 631-637.
- Curtis, A. (2000). A problem-solving approach to the management of change in language education. *Korea TESOL Journal*, 3(1), 1-12.
- Daiker, D., Sommers, J., & Stygall, G. (1996). The pedagogical implications of a college placement portfolio. In E. White, W. Lutz, & S. Kamusikiri (Eds.), *Assessment of writing* (pp. 257-270). New York: The Modern Language Association of America.
- Data Protection Act* (1998). The Office of Public Sector Information (OPSI). Retrieved on October 28, 2009, from <http://www.opsi.gov.uk>
- Davidson, F., Turner, C. E., & Huhta, A. (1997). Language testing standards. In C. M. Clapham, & D. Corson (Vol. Eds.), *Encyclopedia of language and education: Vol. 7. Language testing and assessment* (pp. 303-311). Dordrecht: Kluwer Academic.
- Davies, A. (1985). Follow my leader: Is that what language tests do? In Y. P. Lee, A. C. Y. Y. Fok, R. Lord, & G. Low (Eds.), *New directions in language testing* (pp. 3-12). Oxford: Pergamon Press.
- Davies, A. (1990). *Principles of language testing*. Oxford: Basil Blackwell.
- Denscombe, M. (2007). *The good research guide for small-scale social research projects* (3rd ed.). Maidenhead: Open University Press.
- Denzin, N. K. (1970). *The research act in sociology: A theoretical introduction to sociological methods*. London: Butterworth.
- Denzin, N. K. (1997). Triangulation in educational research. In J. P. Keeves (Ed.), *Educational research, methodology and measurement: An international handbook* (2nd ed., pp. 318-322). Oxford: Elsevier Science.

- De Vaus, D. (2002). *Surveys in social research*. 5th edition. London: Routledge.
- Dörnyei, Z. (2003). *Questionnaires in second language research*. Mahwah: Lawrence Erlbaum Associates.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.
- Dorr-Bremme, D., & Herman, J. (1983). Use of testing in schools: A national profile. In W. Hathaway (Ed.), *Testing in the schools: New directions for testing and measurement* (pp. 3-17). San Francisco, CA: Jossey Bass.
- Dunkin, M. J., & Biddle, B. J. (1974). *The study of teaching*. New York: Holt, Rinehart and Winston.
- Easen, P. (1985). *Making school-centred INSET work*. London: Croom Helm.
- Eckstein, M. A., & Noah, H. J. (Eds.). (1992). *Examinations: Comparative and international studies*. Oxford: Pergamon Press.
- Eckstein, M. A., & Noah, H. J. (1993a). *Secondary school examinations: International perspectives on policies and practice*. New Haven, MA: Yale University Press.
- Eckstein, M. A., & Noah, H. J. (1993b). The politics of examinations: Issues and conflicts. In M. A. Eckstein, & H. J. Noah (Eds.), *Secondary school examinations: International perspectives on policies and practice* (pp. 191-216). New Haven, MA: Yale University Press.
- Elder, C., & Wigglesworth, G. (Eds.). (1996). *The language testing cycle: From inception to washback*. Australia: Applied Linguistic Association of Australia.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- Elton, L., & Laurillard, D. (1979). Trends in student learning. *Studies in Higher Education*, 4, 87-102.
- ESRC (Economic and Social Research Council). (2005). *ESRC research ethics framework*. Retrieved on October 28, 2009, from <http://www.esrcsocietytoday.ac.uk>
- Executive Yuan, Taiwan. (2003). *Challenge 2008: National development plan (2002-2007)*. Retrieved July 19, 2009, from Taipei: Executive Yuan <http://www.ey.gov.tw/mp?mp=1>

- Fairbank, J. K., & Goldman, M. (1998). *China: a new history*. Cambridge, MA: Belknap Press of Harvard University Press.
- Farrell, J. P. (2000). Why is educational reform so difficult? Similar descriptions, different prescriptions, failed explanations. *Curriculum Inquiry*, 30(1), 83-103.
- Ferman, I. (2004). The washback of an EFL national oral matriculation test to teaching and learning. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp.191-210). Mahwah, NJ: Lawrence Erlbaum Associates.
- Flanders, N. A. (1970). *Analysing teaching behaviour*. Reading, MA: Addison-Wesley.
- Frederiksen, J. R., & Collins, A. (1989). A system approach to educational testing. *Educational Researcher*, 18(9), 27-32.
- Freeman, D. (1989). Teacher training, development, and decision making: A model of teaching and related strategies for language teacher education. *TESOL Quarterly*, 23, 27-45.
- Fullan, M. (2007). *The new meaning of educational change*. New York: Teachers College Press.
- Fullan, M., & Park, P. (1981). *Curriculum implementation: A resource booklet*. Ontario: Ontario Ministry of Education.
- Fullilove, J. (1992). The tail that wags. *Institute of Language in Education Journal*, 9, 131-147.
- Gee, J. P., Hull, G., & Lankshear, C. (1996). *The new work order: Behind the language of the new capitalism*. Boulder, CO: Westview Press.
- Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment*, 6, 304-312.
- Gibbons, P. (2006). *Bridging discourses in the ESL classroom*. London: Continuum.
- Gillham, B. (2005). *Researching interviewing: the range of techniques*. Maidenhead: Open University Press.

- Gillham, B. (2007). *Developing a questionnaire* (2nd ed.). London: Continuum.
- Gipps, C. V. (1994). *Beyond testing: towards a theory of educational assessment*. London: The Falmer Press.
- Gochros, H. L. (2005). Interviewing. In R. M. Grinnell Jr., & Y. A. Unrau (Eds.), *Social work research and evaluation: Quantitative and qualitative approaches* (7th ed., pp. 245-270). New York: Oxford University Press.
- Goh, C. (1999). Nationwide curriculum innovation: how do we manage change? In C. Kennedy, P. Doyle, & C. Goh (Eds.), *Exploring change in English language teaching* (pp. 5-19). Oxford: Macmillan Heinemann English Language Teaching.
- Gorard, S., & Taylor, C. (2004). *Combining methods in educational and social research*. Maidenhead: Open University Press.
- Government Information Office, Taiwan (2008). *The Republic of China Yearbook 2008*. Retrieved July 18, 2009, from <http://www.gio.gov.tw>
- Graddol, D. (2006). *English next: Why global English may mean the end of "English as a foreign language"*. London: British Council.
- Grbich, C. (1999). *Qualitative research in health*. London: SAGE.
- Green, A., & Hawkey, R. (2004). Test washback and impact. *Modern English Teacher*, 13, 66-70.
- Grinyer, A. (2002). *The anonymity of research participants: Assumptions, ethics and practicalities* (Social Research Update No. 36). Guildford: University of Surrey.
- Gullickson, A. R. (1984). Teacher perspectives of their instructional use of test. *Journal of Educational Research*, 77, 244-248.
- Hadley, G. (1999). Innovative curricula in tertiary ELT: a Japanese case study. *The Japanese Learner*, 19, 2-11.
- Haertel, E. H. (1999). Performance assessment and education reform. *Phi Delta Kappan*, 80, 662-666.
- Haladyna, T. M., Nolen, S. B., & Haas, N. S. (1991). Raising standardised achievement test scores and the origins of test score pollution. *Educational Researcher*, 20(5), 2-7.

- Hamp-Lyons, L. (1997). Washback, impact and validity: Ethical concerns. *Language Testing*, 14(3), 295-303.
- Hamp-Lyons, L., & Brown, A. (2007). *The effect of changes in the new TOEFL format on the teaching and learning of EFL/ESL: Stage 2 (2003-5): Entering innovation*. Educational Testing Service: TOEFL Research Committee.
- Hamp-Lyons, L., & Condon, W. (2000). *Assessing the portfolio: Principles for practice, theory, research*. Cresskill, NJ: Hampton Press.
- Harrison, I. (1996). Look who's talking now: Listening to voices in curriculum renewal. In K. M. Bailey, & D. Nunan (Eds.), *Voices from the language classroom: Qualitative research in second language education* (pp. 283-303). Cambridge: Cambridge University Press.
- Hayes, B., & Read, J. (2004). IELTS test preparation in New Zealand: Preparing students for the IELTS academic module. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 97-112). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hays, P. A. (2004). Case study research. In K. deMarrais, & S. D. Lapkin (Eds.), *Foundations for research: Methods of inquiry in education and the social sciences* (pp. 217-234). Mahwah, NJ: Lawrence Erlbaum.
- He, A. E. (1996). *Teaching English at China's tertiary universities: Policy, practice and future perspectives*. Unpublished doctoral dissertation, Monash University, Melbourne, Australia.
- Her, J. H. (2007). *The globalization of English: Its impact on English language education in the tertiary education sector in Taiwan*. Unpublished doctoral dissertation, University of Waikato, Hamilton, New Zealand.
- Henrichsen, L. E. (1989). *Diffusion of innovation in English language teaching: the ELEC effort in Japan, 1956-1968*. New York: Greenwood Press.
- Herman, J. L., & Golan, S. (1991). *Effects of standardized testing on teachers and learning – Another look* (CSE Technical Report 334). Los Angeles: University of California,

CRESST.

- Heyneman, S. P. (1987). Use of examinations in developing countries: Selection, research, and education sector management. *International Journal of Education Development*, 7, 251-263.
- Heyneman, S. P., & Ransom, A. W. (1990). Using examinations and testing to improve educational quality. *Educational Policy*, 4, 177-192.
- Hogan, T. P. (1986). *The relationship between the curriculum and test development: Some considerations for the future*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.
- Huang, S. Y. (1997). *Working towards a comprehensive Freshman English curriculum (Part II): A description of the Freshman English programs for non-English majors in Taiwan and a study of the differences between expectations and reality* (NSC 85-2418-H-029-003). NSC Project Report.
- Huberman, M. (1988). Teacher careers and school improvement. *Journal of Curriculum Studies*, 20(2), 119-132.
- Huberman, M., & Miles, M. (1984). *Innovation up close*. New York: Plenum.
- Hughes, A. (1988). Introducing a needs-based test of English language proficiency into an English-medium university in Turkey. In A. Hughes (Ed.), *Testing English for university study* (pp. 134-153). London: Modern English Publications.
- Hughes, A. (1993). *Backwash and TOEFL 2000*. Unpublished manuscript, University of Reading, England.
- Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.
- Johnson, D. (2004). *Proficiency benchmarking and the overarching framework for Languages in the New Zealand Curriculum* (Commissioned Research Reports by New Zealand Ministry of Education). Wellington: N. Z. Ministry of Education.
- Kachru, B. B., & Nelson, C. L. (1996). World Englishes. In S. L. McKay, N. H. Hornberger, & J. C. Richards (Eds.), *Sociolinguistics and language teaching* (pp. 71-102). New York:

Cambridge University Press.

- Karavas-Doukas, E. (1996). Using attitude scales to investigate teachers' attitudes to the communicative approach. *ELT Journal*, 50(3), 187-198.
- Kennedy, C. (1990). Evaluation of the management of change in ELT projects. *Applied Linguistics*, 9(4), 329-342.
- Ko, E. (2002, February 7). Taiwan's higher education crisis. *Taipei Times*, p. 8.
- Kuba, A. (2002). *Strategies-based instruction in Japanese high schools*. Unpublished master's thesis, Department of English, San Francisco State University.
- Lam, H. P. (1994). Methodology washback – An insider's view. In D. Nunan, V. Berry, & R. Berry (Eds.), *Bringing about change in language education* (pp. 83-99). Hong Kong: University of Hong Kong, Department of Curriculum Studies.
- Language Teaching and Learning Centre (LTTC), Taiwan. (2008). *General English Proficiency Test (GEPT)*. Retrieved September 12, 2009, from <http://www.ltcc.ntu.edu.tw>
- Langerfeld, K., Thurlow, M., & Scott, D. (1997). *High stakes testing for students: Unanswered questions and implications for students with disabilities* (Synthesis Report 26). Minneapolis: University of Minnesota, National Center for Educational Outcomes.
- Law, W. W. (2004). Translating globalization and democratization into local policy: Educational reform in Hong Kong and Taiwan. *International Review of Education*, 50 (5-6), 497-524.
- Li, D. (2001). Teachers' perceived difficulties in introducing the communicative approach in South Korea. In D. Hall, & A. Hewings (Eds.), *Innovation in English language teaching* (pp. 149-169). London: Routledge.
- Li, X. J. (1990). How powerful can a language test be? The MET in China. *Journal of Multilingual and Multicultural Development*, 11, 393-404.
- Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4-16.
- Linn, R. L., Grave, M. E., & Sanders, N. M. (1989). *Comparing state and district test results to national norms: Interpretations of scoring "above the national average"* (CSE

- Technical Report 308). Boulder: University of Colorado, CRESST.
- Little, D. (2007). The Common European Framework of Reference for Languages: Perspectives on the making supranational language education policy. *Modern Language Journal*, 91(4), 645-655.
- Long, M. H. (1980). Inside the 'black box': Methodological issues in classroom research on language learning. *Language Learning*, 30, 1-42.
- Luijten, J. M. A. (Ed.). (1991). *Issues in public examinations – A selection of the proceedings of the 1990 IAEA Conference*. Utrecht, Netherlands: Lemma.
- Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum Associates.
- 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.
- Malamah-Thomas, A. (1987). *Classroom interaction*. Oxford: Oxford University Press.
- Markee, N. (1993). The diffusion of innovation in language teaching. *Annual Review of Applied Linguistics*, 13, 229-243.
- Markee, N. (1997). *Managing curricular innovation*. Cambridge: Cambridge University Press.
- Markee, N. (2001). The diffusion of innovation in language teaching. In D. Hall, & A. Hewings (Eds.), *Innovation in English language teaching* (pp. 118-126). London: Routledge.
- Marshall, C., & Rossman, G. B. (2006). *Designing Qualitative Research*. 4th ed. Thousand Oaks: Sage.
- McDonnell, L. M., Burstein, L., Ormseth, T., Catterall, J. M., & Moody, D. (1990, June). *Discovering what schools really teach: Designing improved coursework indicators*. Prepared for the Office of Education Research and Improvement, US Department of Education.
- McNamara, T. (1996). *Measuring second language performance*. London: Longman.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.

- McNamara, T. (2004). Language testing. In A. Davies, & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 763-783). Oxford: Blackwell.
- Merriam, S. B. (1998). *Qualitative research and case study applications in education*. San Francisco, CA: Jossey-Bass.
- Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (pp. 13-103). New York: Macmillan.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241-256.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.
- Ministry of Education, Taiwan. (2004). 公佈國民小學英語教學向下延伸至三年級之課程綱要 [English language curriculum guidelines from grade 3 of elementary education]. Retrieved February 23, 2008, from <http://teach.eje.edu.tw>
- Ministry of Education, Taiwan. (2005a). *Challenge 2008: National development plan*. Retrieved November 20, 2006, from <http://english.moe.gov.tw/ct.asp?xItem=7043&ctNode=784&mp=1>
- Ministry of Education, Taiwan (2005b). *MOE grant project on the Enhancement of Students' Foreign Language Proficiency network*. Retrieved November 20, 2006, from <http://ice.lhu.edu.tw/eng/English/index.htm>
- Ministry of Education, Taiwan (2008). *Education in Taiwan 2008*. Retrieved July 19, 2009, from <http://english.moe.gov.tw/ct.asp?xItem=9895&ctNode=814&mp=1>
- Mitchell, R., Parkinson, B., & Johnstone, R. (1981). *The foreign language classroom: An observational study*. Stirling Educational Monographs 9. Stirling: University of Stirling, Department of Education.
- Mitchell, R., & Myles, F. (2004). *Second language learning theories*. London: Arnold.
- Morse, J. M., & Richards, L. (2002). *Readme first for a user's guide to qualitative methods*. Thousand Oak: Sage.
- Moya, S., & O'Malley, J. M. (1994). A portfolio assessment model for ESL. *The Journal of*

- Educational Issues for Minority Students*, 13, 13-36.
- National Research Council. (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington, DC: National Academy Press.
- Nichollos, A. (1983). *Managing educational innovation*. London: Allen and Unwin.
- Noble, A. J., & Smith, M. L. (1994a). *Measurement-driven reform: Research on policy, practice, repercussion* (CSE Technical Report 381). Tempe: Arizona State University, CSE.
- Noble, A. J., & Smith, M. L. (1994b). *Old and new beliefs about measurement-driven reform: 'The more things change, the more they stay the same'* (CSE Technical Report 373). Tempe: Arizona State University, CSE.
- North, B. (2000). *Development of a common framework scale of language proficiency*. New York: Peter Lang Publishers.
- North, B. (2007). The CEFR illustrative descriptor scales. *Modern Language Journal*, 91(4), 656-659.
- Nunan, D. (1989). *Designing Tasks for the Communicative Classroom*. Cambridge: Cambridge University Press.
- Nunan, D. (2005). Classroom research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 225-240). Mahwah, NJ: Lawrence Erlbaum Associates.
- Oladejo, J. A. (2005). Language education policy in Taiwan: The perceptions of English teachers. In Department of Applied English, Ming Chuan University (Ed.), *The proceedings of 2005 international conference and workshop on TEFL & applied linguistics* (pp. 306-319). Taipei: Crane Publishing.
- O'Malley, J. M., & Chamot, A. (1996). *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press.
- Patton, M. Q. (2002). *Qualitative evaluation and research methods* (3rd ed.). Newbury Park, CA: Sage.
- Pearson, I. (1988). Tests as levers for change. In D. Chamberlain, & R.J. Baumgardner (Eds.),

- ESP in the classroom: Practice and evaluation* (pp. 98-107). Great Britain: Modern English Publication.
- Pennington, M. C. (1995). The teacher change cycle. *TESOL Quarterly*, 29(4), 705-731.
- Petrie, H. G. (1987). Introduction to "evaluation and testing". *Educational Policy*, 1, 175-180.
- Pinar, K. (1999). Introducing new course materials. In C. Kennedy, P. Doyle, & C. Goh (Eds.), *Exploring change in English language teaching* (pp. 66-74). Oxford: Macmillan Heinemann English Language Teaching.
- Popham, W. J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappa*, 68, 679-682.
- Popham, W. J. (1993). Measurement-driven instruction as a 'quick-fix' reform strategy. *Measurement and Evaluation in Counseling and Development*, 26, 31-34.
- Qi, L. (2004). Has a high-stakes test produced the intended changes? In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 171-190). Mahwah, NJ: Lawrence Erlbaum Associates.
- Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing*, 22(2), 142-173.
- Ratnavadivel, N. (1999). Teacher education: Interface between practices and policies, the Malaysian experience 1979-1997. *Teaching and Teacher Education*, 15(2), 193-213.
- Read, J., & Hayes, B. (2003). The impact of IELTS on preparation for academic study in New Zealand. In R. Tulloh (Ed.), *International English language testing system research reports 2003* (Vol. 4, pp. 153-206). Canberra: IELTS Australia.
- Rea-Dickins, P., & Scott, C. (2007). Washback from language tests on teaching, learning and policy: evidence from diverse settings. *Assessment in Education: Principles, Policy & Practice*, 14(1), 1-7.
- Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford, & M. C. O'Connor (Eds.), *Changing assessment: Alternative views of aptitude, achievement and instruction* (pp. 37-85). London: Kluwer Academic Publishers.

- Richards, J. C., & Rodgers, T. S. (2001). *Approaches and methods in language teaching* (2nd ed.). New York: Cambridge University Press.
- Robson, C. (2002). *Real world research*. Oxford: Blackwell.
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York: Free Press.
- Saville, N., & Hawkey, R. (2004). The IELTS impact study: Investigating washback on teaching materials. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 73-96). Mahwah, NJ: Lawrence Erlbaum Associates.
- Seedhouse, P. (2004). *The international architecture of the language classroom: A conversation analysis perspective*. Oxford: Blackwell.
- Shannon, F. (2006, August 8). EFL pedagogy in Taiwan. *Taipei Times*, p. 8.
- Shepard, L. A. (1990). Inflated test score gains: Is the problem old norms or teaching the test? *Educational Measurement: Issues and Practice*, 9, 15-22.
- Shepard, L. A. (1993). The place of testing reform in educational reform: A reply to Cizek. *Educational Researcher*, 22(4), 10-14.
- Shih, C. (2007). A new washback model of students' learning. *The Canadian Modern Language Review*, 64(1), 135-162.
- Shih, Y. H. (2000). *General English education at the university level in Taiwan*. Paper presented at JACET 39th Annual Convention: English Education in the East Asia for the 21st Century. Japan: Okinawa.
- Shohamy, E. (1992). Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning. *The Modern Language Journal*, 76(4), 513-521.
- Shohamy, E. (1993). *The power of tests: The impact of language tests on teaching and learning*. Washington, DC: National Foreign Language Centre Occasional Papers.
- Shohamy, E. (1998). Critical language testing and beyond. *Studies in Educational Evaluation*, 24(4), 331-345.
- Shohamy, E. (2000). Using language tests for upgrading knowledge: The phenomenon, source and consequences. *Hong Kong Journal of Applied Linguistics*, 5(1), 1-18.

- Shohamy, E. (2001). *The power of tests: A critical view of the uses of language tests*. Essex: Longman Pearson.
- Shohamy, E. (2007). Language tests as language policy tools. *Assessment in Education*, 14(1), 117-130.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: washback effect over time. *Language Testing*, 13(3), 298-317.
- Sieber, J. E. (2009). Planning ethically responsible research. In L. Bickman, & D. J. Rog (Eds.), *The SAGE handbook of applied social research methods* (2nd ed., pp. 106-142). Thousand Oaks, CA: SAGE.
- SIL International (1999). *The American Council for the Teaching of Foreign Languages (ACTFL) guidelines*. Retrieved July 20, 2009, from <http://www.sil.org>
- Sinclair, J. McH., & Coulthard, R. M. (1975). *Toward an analysis of discourse*. London: Oxford University Press.
- Singer, E. (1993). Informed consent and survey response: A summary of the empirical literature. *Journal of Official Statistics*, 9(2), 361-375.
- Smith, M. L. (1991a). Put to the test: the effects of external testing on teachers. *Educational Research*, 20(5), 8-11.
- Smith, M. L. (1991b). Meanings of test preparation. *American Educational Research Journal*, 28, 521-542.
- Smith, M. L., Noble, A. J., Cabay, M., Heinecke, W., Junker, M. S., & Saffron, Y. (1994). *What happens when the test mandates change? Results of a multiple case study* (Technical Report 380). Tempe: Arizona State University, CRESST.
- Spada, N., & Fröhlich, M. (1995). *The Communicative Orientation of Language Teaching Observation Scheme (COLT)*. Sydney, Australia: Macquarie University, The National Centre for English Language Teaching and Research (NCELTR).
- Spolsky, B. (1994). The examination of classroom backwash cycle: Some historical cases. In D. Nunan, V. Berry, & R. Berry (Eds.), *Bringing about change in language education* (pp. 55-66). Hong Kong: University of Hong Kong, Dept. of Curriculum Studies.

- Spolsky, B. (1995). *Measured words*. Oxford: Oxford University Press.
- Spolsky, B. (1997). The ethics of gatekeeping tests: What have we learned in a hundred years? *Language Testing*, 14(3), 242-247.
- Stern, H. H. (1989). Seeing the wood AND the trees: Some thoughts on language teaching analysis. In R. K. Johnson (Ed.), *The second language curriculum* (pp. 207-221). Cambridge: Cambridge University Press.
- Stern, H. H. (1992). *Issues and options in language testing*. Oxford: Oxford University Press.
- Stiggins, R., & Faires-Conklin, N. (1992). *In teachers' hands*. Albany, NY: State University of New York Press.
- Stoller, F. L. (1994). The diffusion of innovations in intensive ESL programs. *Applied Linguistics*, 15(3), 300-327.
- Stoyhoff, S. (1989). Successfully implementing educational change and innovation. *ORTESOL Journal*, 10, 17-33.
- Swain, M. (1985). Large-scale communicative testing: a case study. In Y. P. Lee, A. C. Y. Fok, R. Lord, & G. Low (Eds.), *New directions in language testing* (pp. 34-46). Oxford: Pergamon Press.
- Tang, C., & Biggs, J. B. (1996). How Hong Kong students cope with assessment. In D. A. Watkins, & J. B. Biggs (Eds.), *The Chinese learner: Cultural, psychological and contextual influences* (pp. 159-182). Hong Kong: Center for Comparative Research in Education.
- Taylor, L. (2005). Washback and impact. *ELT Journal*, 59(2), 154-155.
- Technological and Vocational Education, Taiwan. (2008). *Course Network*. Retrieved July 19, 2009, from <http://course.tvc.ntnu.edu.tw>
- Tsao, F. F. (2001). Teaching English from elementary school in an Asian context: A language-planning perspective. *The Language Teacher Online*, 25(6), 1-5.
- Tsui, A. B. M. (1995). *Introducing classroom interaction*. London: Penguin Books.
- Van Lier, L. (1988). *The classroom and the language learner: Ethnography and second language classroom research*. London: Longman.

- Van Lier, L. (1989a). Classroom research in second language acquisition. *Annual Review of Applied Linguistics*, 10, 173-186.
- Van Lier, L. (1989b). Ethnography: Bandid, bandwagon, or contraband? In C. Brumfit, & R. Mitchell (Eds.), *Research in the language classroom* (pp. 33-53). Great Britain: Modern English Publications in association with the British Council.
- Van Lier, L. (2004). *The ecology and semiotics of language learning: A sociocultural perspective*. London: Kluwer Academic.
- Wagner, J. (1988). Innovation in foreign language teaching. *AILA Review*, 5, 99-117.
- Wall, D. (1996). Introducing new tests into traditional systems: Insights from general education and from innovation theory. *Language Testing*, 13, 334-354.
- Wall, D. (1997). Impact and washback in language testing. In C. Clapham, & D. Corson (Eds.), *Encyclopedia of language and education: Vol. 7. Language testing and assessment* (pp. 291-302). Dordrecht: Kluwer Academic.
- Wall, D. (2000). The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled? *System*, 28, 499-509.
- Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory*. Cambridge: Cambridge University Press.
- Wall, D., & Alderson, J. C. (1993). Examining washback: the Sri Lankan impact study. *Language Testing*, 10, 41-70.
- Wall, D., Kalnberzina, V., Mazuoliene, Z., & Truus, K. (1996). The Baltic States year 12 examination project. *Language Testing Update*, 19, 15-27.
- Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, 13(3), 318-333.
- Watanabe, Y. (2000). Washback effects of the English section of Japanese entrance examinations on instruction in pre-college level EFL. *Language Testing Update*, 27, 42-47.
- Watanabe, Y. (2004). Teachers factors mediating washback. In L. Cheng, Y. Watanabe, & A.

- Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 129-146). Mahwah, NJ: Lawrence Erlbaum Associates.
- Weir, C. J. (2005). Limitations of Common European Framework for developing comparable examinations and tests. *Language Testing*, 22(3), 281-300.
- Weir, C. J., & Milanovic, M. (2003). *Continuity and innovation: revising the Cambridge Proficiency in English Examination 1913-2002*. Cambridge: UCLES/Cambridge University Press.
- Westhoff, G. (2007). Challenges and opportunities of the CEFR for reimagining foreign language pedagogy. *Modern Language Journal*, 91(4), 676-679.
- White, R. V. (1988). *The ELT curriculum change: Design, innovation and management*. Oxford: Blackwell.
- White, R. V. (1991). Managing curriculum development and innovation. In R. V. White, M. Martin, M. Stimson, & R. Hodge (Eds.), *Management in English language teaching* (pp. 166-195). Cambridge: Cambridge University Press.
- White, R. V. (1993). Innovation in curriculum planning and program development. *Annual Review of Applied Linguistics*, 13, 244-259.
- Wu, W. (2004). Does direct grammar instruction improve students' performance on grammar-based tests of English as a foreign language? *Journal of Maori and Pacific Development*, 5(2), 48-103.
- Yung, C. S., & Hsu, M. J. (2005). Development of Taiwan's education reform over the past decade. *National Policy Foundation: Taiwan Development Perspectives 2005*, 69-82. Retrieved July 19, 2009, from <http://old.npf.org.tw/english/Publication/TDP2005.pdf>