

國立政治大學亞太研究英語碩士學位學程

International Master's Program in Asia-Pacific Studies

College of Social Sciences

National Chengchi University

碩士論文

Master's Thesis

台灣公務人員採用英語標準化測驗為評量機制之研究：
從 2002 到 2010 年

**English Language Proficiency Testing Policy
in Taiwan's Civil Service from 2002 to 2010**

Student: Timothy Hogan 霍力

Adviser: Dr. Judy Yu Hsueh-ying 尤雪瑛 博士

二〇一一年七月

July 2011

Abstract

The concern over test consequence has inspired research into the wider impact of language tests and testing policies, but few studies have examined this subject in the context of Taiwan. With the goal of enhancing Taiwan's global competitiveness by upgrading manpower quality, the central government implemented a 2002 policy to develop the English proficiency of civil servants by recognizing passing marks on approved English language proficiency tests as a promotion criterion. This thesis reports on a research study that adopted a multi-method approach to assess the testing policy's impact on test-takers and analyze the rationale and consequences of revisions to the policy that were implemented between 2002 and the 2011. A survey of 282 civil servants working in the banking, economics, and finance sectors yielded data about the participants' self-assessment of their English proficiency and workplace need for the language, English study and test-taking experience, impressions of English proficiency tests, and assessment of the effectiveness of the testing policy. Statistical analysis of the test impression and policy effectiveness data revealed significant correlation between positive assessments of the policy's impact and the perceived fairness of the testing policy, the policy's influence on motivation to study English, and the participants' intrinsic interest in improving their English. Interviews with officials involved in formulating and implementing the testing policy and a review of government documents related to the policy provided data that were incorporated into the Geelhoed-Schouwstra policy analysis framework and facilitated the identification of factors that influenced the outcomes of the testing policy. The results of this study of an English language testing policy help to clarify who the test-takers and test users are, how and why tests are being used, and what the consequences of test use are.

Table of Contents

Abstract	2
List of Tables	7
List of Figures	7
Acknowledgement	8
Chapter 1: Introduction	9
Context and Purpose of the Study	9
Research Questions and Hypotheses	10
Research Methodology	12
Significance of the Research	12
Thesis Organization	13
Chapter 2: Taiwan Research Context and Background	15
Context	15
Critical language testing and ethical testing	15
The English language in Taiwan	19
Civil service in Taiwan	24
Background	29
Challenge 2008	29
Tests recognized by the government	34
TOEFL	35
TOEIC	37
IELTS	39
BULATS	41
FLPT	43
GEPT	44
Conclusion	46
Chapter 3: Literature Review	48
History of Civil Service Exams	49
Language Policies and Testing	52
Policy analysis framework	52
Language policies	53

CEFR	57
Language testing policies in Asia and Taiwan	62
Test Use and Consequence	64
Language test use	64
Consequential validity	66
Assessment use arguments	69
Washback and impact	71
Testing and Motivation for Learning	73
Lifelong learning and adult education	73
Motivation: Learning and testing	76
Motivation in English learning and testing in Taiwan	81
Conclusion	83
Chapter 4: Methodology	85
Questionnaire	87
Stages of questionnaire research	87
Development	87
Handling	89
Statistics	90
Content	90
Demographics	90
English language	92
Self-assessment of English ability	92
Importance of English for work	92
Importance of English for work: Qualitative comments	93
Testing experience	93
English study	94
Test impression	94
Policy effectiveness	95
Interviews	96
Online government publications and internal documents	99
Extended Geelhoed-Schouwstra (G-S) Policy Analysis Framework	100

Conclusion	105
Chapter 5: Results and Discussion	107
Questionnaire	107
Demographics	107
English language	110
Self-assessed English ability	111
Importance of English	112
Overall averages: Self-assessed English and importance of English	114
Test experience	115
English study	122
Test impression	123
Qualitative results	126
Policy effectiveness	129
Analysis of correlation among test impression and policy effectiveness variables	132
Policy Analysis	135
Interview results	135
Research, Development, and Evaluation Commission	135
Examination Yuan	136
Central Personnel Administration	137
Ministry of Education	138
Central Bank of China (Taiwan), Ministry of Economic Affairs, and Ministry of Finance	139
Stages of the policy cycle	141
Goals	141
Objectives	142
Methods/Instruments	143
Activities	144
Performances	145
Evaluation	147
Extended G-S policy analysis framework	149

Conceptual framework	149
Institutional framework	150
Political setting	152
Social setting	153
Economic setting	154
Institutional setting	155
Conclusion	157
Chapter 6: Conclusion	158
General Aims of the Study	158
Implications of the Study	158
Limitations of the Study	163
Future Research Directions	165
References	167
Appendices	
Appendix A: Common European Frame of Reference, Common Reference Levels: Global Scale	173
Appendix B: Civil Service English Examination Scoring Table	174
Appendix C: Questionnaire: Chinese Version	175
Appendix D: Questionnaire: English Version	179
Appendix E: Interview Questions	183
Appendix F: Questionnaire Results	186
Appendix G: Pearson Correlation Among Test Impression and Policy Effectiveness Variables	198
Appendix H: Ministry of Economic Affairs, December 2010 English Proficiency Test Data	199

List of Tables

Table 1: Taiwan Civil Servants Manpower Statistics	27
Table 2: Interviews: Subjects and Schedule	98
Table 3: Participant Demographics	109
Table 4: Test Impression Results	124
Table 5: Summary of Participants' Comments on English Testing Policy	127
Table 6: Policy Effectiveness Results	130
Table 7: Pairs of Variables with Significant Pearson (r) Correlation $>.300$	133

List of Figures

Figure 1: The relationship among agencies responsible for managing the civil service	26
Figure 2: Taiwanese civil service English proficiency testing policy timeline	31
Figure 3: Stages of questionnaire research	87
Figure 4: The extended Geelhoed-Schouwstra framework of policy analysis	101
Figure 5: Comparing self-assessed English with importance of English by language skill	111
Figure 6: Overall averages: Self-assessed English and Importance of English	114
Figure 7: Tests taken by participants and grouped by test developer	117
Figure 8: Participants' CEFR levels based on self-reported test score	121
Figure 9: Participants' study of English	123

Acknowledgement

This thesis was made possible by the assistance provided to me by a wide range of people. Without the help that I received from each of them, I would not have been able to complete my research

First of all, I would like to express my deepest gratitude to the employees of the Central Bank of China (Taiwan), the Ministry of Finance, and the Ministry of Economic Affairs who not only completed the questionnaire but also agreed to distribute it among their colleagues.

My appreciation also goes out to the representatives of the Research, Development, and Evaluation Commission; the Examination Yuan; the Central Personnel Administration; the Ministry of Education; the Central Bank of China (Taiwan); the Ministry of Finance; and the Ministry of Economic Affairs for allowing me to interview them about the English testing policy.

I am also grateful to the Office of President Ma Ying-jeou for facilitating access to officials familiar with the English proficiency testing policy for civil servants.

Furthermore, I am thankful for the support and advice offered to me by my colleagues at the Language Training and Testing Center, particularly Professor Kao Tien-en, Executive Director, and Jessica R.W. Wu PhD, Program Director, Research & Development Office.

Much thanks is also due to the members of my thesis committee, including Professor Vincent Chang Wu-chang of National Taiwan Normal University and Professor Hintat Cheung of National Taiwan University for their detailed comments and insightful suggestions on my research framework and thesis. My deepest gratitude is reserved for my thesis advisor at National Chengchi University, Professor Judy Yu Hsueh-ying, for generously offering her time, expertise, and guidance during all stages of my research.

Finally, I would like to thank all the members of my family on both sides of the Pacific, and especially my wife Wang Hsueh-chiao, for their inspiration, encouragement, love, support, and understanding.

Chapter 1

Introduction

Context and Purpose of the Study

Globalization challenges nations to increase their competitiveness in order to attract investment and workers with highly valued skills. One means by which Taiwan has sought to do this is by developing the English ability of its citizens. Improved English proficiency is seen to enable Taiwanese individuals and institutions to communicate more effectively with the rest of the world. The efforts undertaken by Taiwan's government to enhance the nation's English proficiency have affected the education system, private enterprise, and the public sector. To achieve the goal of improving the quality of Taiwan's workforce, which was laid out in the Challenge 2008 national development plan, the central government undertook an initiative to improve the English of its staff by encouraging them to improve their English proficiency. A central feature of this policy was the use of English tests to motivate civil servants to develop their ability to use English. Tests have long played an important role as gate-keepers to education and professional resources in Asian societies, and they are seen by many as selection instruments that provide everyone with a fair chance at opportunity based on their merit. While tests have performed this function for centuries, stretching all the way back to the Chinese imperial examination system, they are known to result in unintended, often detrimental, consequences.

The use of language tests as a method for encouraging Taiwanese civil servants to improve their English ability was launched in 2002. In that year, the government set about devising a policy that would recognize passing scores on English proficiency tests as a criterion for promotion scoring. In order to pursue this plan, an English proficiency scale needed to be

adopted, qualified tests recognized, and an incentive system implemented. These steps took place over the next several years, but they were not without controversy. Questions were raised about the appropriateness of the language proficiency scale that was selected, the comparability of English tests that were recognized, and the impact of the incentives on the motivation of the government employees. In response to the difficulties that were encountered, the government modified specific policy measures and continued to promote the objectives and goals that it had originally laid out. In 2009, the Plan for Enhancing National English Proficiency was promulgated, and while it maintained the goal of encouraging civil servants to improve their English, proficiency testing played no explicit role in the efforts that it called for. However, proficiency testing continues to be a feature of the government's efforts to improve the English ability of civil servants. Although it would seem that there is no central policy that recognizes proficiency levels or specific tests, or calls for the use of standard promotion scoring values, the measures that were developed between 2002 and 2005 remain in effect within individual agencies. The consequences of the process of instituting and revising the English testing policy for civil servants are not well understood outside of the government, and it is this situation that the current research aims to redress. The consequences of test use are closely related to both the context in which tests are used and their intended uses.

Research Questions and Hypotheses

This research is designed to explore three research questions with regard to the impact of the testing policy and an analysis of the factors that may have an influence on the policy's effectiveness.

RQ1: What are the impacts of the policy encouraging civil servants to pass an English language proficiency test in order to qualify for promotion?

RQ2: Does this policy achieve the aims for which it was created?

RQ3: What factors have a bearing on the achievement of the SELPT policy aims?

The first question is designed to identify the consequences of the testing policy. The use of a test to motivate individuals to improve their English proficiency is predicated on the assumption that the use of the test will have an impact on the test-takers. If the test is high-stakes, test-takers will be motivated to devote time and energy to studying in order to pass it. This intended consequence is not the only likely impact of a high-stakes test, however. It is hypothesized in this thesis that a policy that calls for the use of an English language proficiency test as a promotion scoring criterion will influence the test-takers' English proficiency, study of English, use of English at work, interest in and motivation for improving their English, impression of English tests, and evaluation of the policy of the use of English tests.

The second research question aims to assess how effective the government's testing policy is at achieving its aims. It is hypothesized that the use of English language proficiency tests as a promotion scoring criterion will increase the percentage of civil servants with a minimum level of English proficiency. Those government employees who possessed English proficiency prior to the implementation of the policy but had not taken an English test will likely be motivated to certify their English proficiency by passing a recognized test, and may continue to improve their ability in order to accrue additional promotion scoring points. Similarly, a portion of those civil servants whose English proficiency does not meet the minimum level will be motivated to improve their English sufficiently to enable them to perform well enough on an English proficiency test to obtain promotion scoring points associated with the proficiency level they achieve. There will likely be another portion of civil servants who will not be motivated to improve their English ability enough to qualify for promotion scoring points. The factors that

might prevent civil servants from earning high marks on an English language proficiency test, or influence them to elect not to take a test are the focus of the third research question.

The final question seeks to learn what specific features of the testing policy, or the context of the test use, have an influence on the outcome of the policy. The objectives associated with the testing policy were mainly concerned with the percentage of civil servants who had taken a recognized English proficiency test and received a score that was at least equivalent to a basic level of proficiency. Using this measure as the basis for assessing the effectiveness of the testing policy, it is hypothesized that the achievement of the policy's aims will be influenced by the demographics of civil servants, the relative need for English in different agencies and positions, the tests that are recognized, civil servants' impressions of English tests and their interest in learning English, and the incentives that are employed.

Research Methodology

The research methodology in this study employed three key approaches. First, a questionnaire surveyed employees of three government agencies with regard to their English proficiency, need for English in their work, experience taking English test, impressions of English tests, and evaluation of the testing policy. Next, a series of interviews with government officials working in agencies that were directly involved in formulating and implementing the testing policy were carried out. Finally, government documents with relevance to the testing policy were reviewed to trace the development of the policy from 2002 to 2011.

Significance of the Research

The impact of language tests is a subject into which much research has been carried out in recent decades. Test impact studies produce evidence of the consequences of test use that can be used to support claims of the validity of the use of test scores. Such evidence is valuable for

test developers, test users, language test researchers, and ultimately for the individuals who take high-stakes tests. Policy analysis studies examine the rationale behind policies; the process of policy formulation, implementation, and revision; and policy outcomes. The results of such studies can inform policy makers about the relative benefits of specific policy measures, influence the creation of future policies, and educate stakeholders about the consequences of policies.

While there has been research into the impact of testing policies in the academic domain in Taiwan, the use of English language tests as a promotion scoring criterion for Taiwanese civil servants has received little attention. The results of this study will offer a unique perspective on the uses of English language proficiency tests in a society that places great emphasis on the role of testing in education and as a means of professional advancement.

Thesis Organization

In Chapter 2, the background and context to this research is reviewed, including the theoretical basis for the study, the role of English in Taiwan's development, the structure of Taiwan's government and civil service system, a timeline of the testing policy, and the English language proficiency tests recognized by the government. In Chapter 3, a review of the relevant literature is offered, addressing the history of civil service testing and the link to language testing, policy analysis and language testing policies, test use and validation, and lifelong learning and motivation. In Chapter 4, the research methods used in this thesis are described, including a questionnaire for civil servants, interviews with government officials, and a policy analysis framework. In Chapter 5, the results of the questionnaire, the interviews, and the policy analysis are presented and their significance is discussed. Chapter 6 concludes the thesis, discussing the

implications of the research findings, outlining the limitations of the research, and suggesting further directions for research.



Chapter 2

Taiwan Research Context and Background

In this chapter, the Taiwanese research context and the background for this thesis are introduced. First, the research context is described, establishing the need for research into the use of language tests from the perspective of critical language testing and ethical testing. Next, the historical factors influencing the status of English in Taiwan, its place in the education system, and its use in society are discussed. A review of the organization of Taiwan's central government follows, identifying its key structural features and the role of agencies with particular relevance to this study. The civil service system is then described, focusing on its demographic features, as well as the examination and promotion system. Then the English language testing policy is outlined, identifying the significant alterations that took place over a nine-year period and reviewing the performance indicators that served as a measure of the policy's effectiveness. Finally, the major English proficiency exams that were recognized by the government on the basis of their alignment with a common language proficiency standard are introduced.

Context

Critical language testing and ethical testing. Critical language testing is a field of inquiry with the goal of bringing greater transparency to language testing by enhancing accountability of those involved in the process. It seeks to consider the assumptions upon which language testing is based and determine their impact on language tests and stakeholders. Rather than simply working to collect evidence in support of the validity of inferences about test scores, scholars utilizing this approach question the methods and practices that make up large-scale, high-stakes testing and explore the potential benefits and feasibility of alternatives. Shohamy

(2000, 2001, 2008) is closely associated with critical testing and offers a set of questions that explore the role and responsibility of test stakeholders. Addressing the power of tests, she proposes asking not just about test-takers and test users, but also about the identity of testers and their agenda. She suggests considering the context in which test-takers operate and the context of the topic being tested as important factors, recognizing that tests are given for specific reasons and that test results have a powerful influence on decision-makers. Shohamy asks who will benefit from the test, as well as what messages about students, teachers and society that the test assumes. It is her recommendation that critical testing address both the intended and unintended uses of a test.

Ethical language testing is concerned with ensuring that the rights and interests of test-takers are respected and protected. This concern extends to establishing standards or codes of professional behavior among test developers. A key component of this concept is the notion that testers share responsibility for the consequences of the uses of their tests. From this perspective, testers may not ignore situations in which their tests are put to unintended uses, and have an ethical responsibility to see that the rights of individuals are not sacrificed in the interests of the society. Shohamy distinguishes between traditional testing, a scientific field concerned with the creation of quality tests that can accurately measure the knowledge of those tested through the use of objective items, and use-oriented testing, which “views testing as embedded in educational, social and political contexts” and is concerned with the rationales for testing, the effects on test-takers and the consequences on wider society (2001, pp. 3-4). In her analysis of the power of testing, Shohamy notes that tests may have detrimental effects on test-takers and are used as disciplinary tools. She identifies key assumptions that played a role in the emergence of test as power tools: they would grant opportunities to all, be objective, scientific, and use

objective item types. The power of tests is seen as a result of multiple factors, including 1) the public's perception of tests as authoritative, 2) the fact that tests allow for flexible cut scores, 3) the ability of tests to control and redefine knowledge, 4) tests' strong appeal to the public, 5) their usefulness for delivering objective proofs, 6) the fact that they allow for cost-effective and efficient policy making, and 7) the ability of tests to provide those in authority with visibility and evidence of action (Shohamy, 2001, pp. 37-41). The result of these factors is that tests are well-suited to acting as instruments of policy for those in authority. The recognition that tests are powerful is one of the motivations for undertaking this study. The author of this study believes that cultivating a greater understanding of the uses of English language proficiency tests and the consequences of their use is an ethical responsibility of those involved in test development. In this regard, Shohamy's promotion of use-oriented testing is seen as beneficial to test stakeholders, including policy makers and test-takers alike.

Kunnan (2005) characterizes language assessment as a "field that is primarily concerned with the psychometric qualities of tests and one in which test developers / researchers ignore the socioeconomic-political issues that are critically part of testing and testing practices" (Kunnan, p. 779). He calls for research into language assessment to consider the wider context, including political and economic factors; the educational, cultural, and social setting; technology and infrastructure; and legal and ethical issues. Kunnan identifies the way forward as asking how ethical test development and test use can be promoted.

Fulcher & Davidson argue that despite the potential for the misuse of tests, "ethical and democratic approaches to testing provide opportunities and access to education and employment" (Fulcher & Davidson, 2007, p. 138). They propose that the core element of an ethical approach to language testing is the concept of professionalism, which provides a contract-

-a code of ethics--that safeguards the interests of test developers, test takers, and society. In their view, such a code can be the basis for creating a community of practitioners who take part in discussion and debate, enabling a collective understanding among its members about what is felt to be true and right. The key elements of such an ethical community are the “constant exercise of self-questioning and open debate” and the serious consideration of contrary views and evidence (Fulcher & Davidson, p. 140). This view of professionalism within the field of language testing is democratic in the sense that it conceptualizes the greatest good as that which enables “the highest development of both society and the individual” (Fulcher & Davidson, p. 141). Fulcher & Davidson (2007) note that Messick (1989, p. 86) similarly described positive consequence as an outcome of distributive justice that provides access to conditions that benefit individual well-being, “conceived to include psychological, physiological, economic, and social aspects.” This view recognizes the political nature of testing as an exercise of power that may result in benefits equally accruing to both individual test-takers and the greater society. McNamara and Roever assert that language testing research must move beyond its practical activity and recognize its role in the “articulation and perpetuation of social relations” (McNamara & Roever, 2006, p. 40).

The ideas introduced in this section are fundamental to the research approach of this thesis. The power of tests is substantial, and an analysis of the uses of tests is of value to stakeholders. The knowledge of testing has traditionally been divided, with test developers, test users, and test-takers all maintaining their separate understanding of tests. This research attempts to go some way toward sharing testing knowledge among those with different perspectives on the subject but a common interest in understanding the implications of the use of language tests. While this thesis looks specifically at the use of English language proficiency

tests as a promotion scoring criterion for civil servants in Taiwan, the value of the results that it produces has a wider significance that may be of use to testers, test users, and test-takers in other parts of the world and dealing in languages other than English.

The English language in Taiwan. The role of the English language in Taiwan's society is related to the nation's political, educational, and economic relations with the English-speaking world, particularly the United States. U.S.-Taiwan relations were strong for decades in the mid-twentieth century, from the 1950s until the U.S. recognition of the P.R.C. in 1976, when official relations between the Washington and Taipei were severed. Because of their close relations during that period, Taiwan sent both government officials and students to the U.S. to receive training and education, and these individuals needed to be able to communicate in English. In 1951, the U.S. Government Aid Agency and the Executive Yuan's Council on US AID jointly established the English Training Center in Taipei, the forerunner of the Language Training and Testing Center (LTTC), to train participants in technical assistance programs in English language skills before their departure (Kunnan and Wu, 2010, p. 77). While the English Training Center initially trained about 100 students per year, the number quickly grew.

In addition to the civil servants that Taipei sent to the U.S., university students also traveled to the U.S. to take advantage of the educational opportunities there. In 1950, a total of 3,637 Taiwanese were studying at American universities and colleges, and that number grew steadily each year, reaching a peak of 37,580 in 1994 before gradually decreasing to the 28,065 recorded in 2009 (MOE, n.d.). While most of those students chose to remain in the U.S. when their education was complete, many that did return accepted positions as professors at Taiwan's most prestigious universities.

Taiwan's economic development also played a role in the use of English in Taiwan as the nation's export-oriented economy was heavily dependent on the U.S. market for its goods. It was important for Taiwan firms to have some employees with English proficiency to carry on communication with their American counterparts, few of whom could speak Chinese. These political, educational, and economic factors had an impact on the development of the use of English in Taiwan that continues today.

Due to the reasons cited above, and the equally influential role of English as an international language that facilitates communication among nations for which English is not a native language, English is a feature of Taiwan's education system at all levels. While English has long been taught as a subject in Taiwanese universities, both to students majoring in English and those who study other subjects, it has recently increased in importance as universities and colleges have adopted policies that require students to demonstrate English proficiency as either an admission or graduate requirement (Hsu, 2009). English education also has a role in Taiwan's secondary schools, where students attend English classes and prepare to take the General Scholastic Aptitude Test and the Department Required Test, the competitive national exams developed and administered by College Entrance Exam Commission that feature a section that tests students' knowledge of English vocabulary and grammar. English is even a required subject in elementary schools throughout Taiwan, and the lowering of the age at which students begin studying English has been a controversial development over recent decades.

Alongside the growing importance of English in Taiwan's elementary schools, junior and senior high schools, and colleges and universities, English education is also a feature of the supplemental school system. Preschool age children in kindergartens study English, often from native-speaking teachers, leading some parents and scholars to question the consequences of

teaching children English before they are able to speak their mother tongue, whether Mandarin Chinese, Taiwanese (Southern Min), Hakka, or one of Taiwan's indigenous Austronesian languages. With English being taught in elementary and junior and senior high schools, attending after-school lessons in English has also become commonplace for Taiwanese students. In such courses, students are drilled in English in order to prepare them to take the high school entrance exam, which also contains an English section, and the university entrance exams. While such test preparation courses may prepare these students to earn good scores on their English exams, relatively few students develop the ability to communicate effectively in English through these approaches.

Adults also study English in Taiwan, particularly in its larger cities, where English ability is seen as a useful skill for academic success and career development. University graduates who hope to study for advanced degrees in the U.S. or other English-speaking countries attend test preparation schools. Courses at these schools train students to obtain high marks on English proficiency tests, such as the Test of English as a Foreign Language (TOEFL) and the International English Language Testing System (IELTS), which are used as admission criteria by foreign universities. In addition to the test preparation centers, some English language schools also train adult students to develop their communicative ability in the language. Because of the heavy emphasis on learning English in order to pass tests, and the heavy workload that students encountered in their other courses, many college and university graduates, excepting those who majored in English, have difficulty carrying on a conversation or writing a simple letter in the language. If they hope to use English in their work, whether for a foreign company or a local firm that does business internationally, many adults find it necessary to pursue additional study of English in order to develop practical communication skills in English. Adults in Taiwan may

also study English to use when they are traveling overseas for work or pleasure, and to satisfy their curiosity about foreign cultures and customs.

The testing of English in Taiwan occurs along two main axes, achievement testing and proficiency testing. The English exams taken by students enrolled in English courses in the school system are achievement tests. Achievement exams are used to measure the extent to which the test-taker has mastered material included in the curriculum. Such exams include classroom-based assessment instruments created by teachers to measure their students' learning progress and large-scale exams that are used by high schools, colleges, and universities for decisions related to admission. Entrance exams such as those taken by Taiwanese students seeking admission to high school and universities are high-stakes tests because they function as gate-keeping devices that control access to educational resources that could have a major influence on the test-taker's future success. Their content is based on the curriculum of junior high and senior high school English courses and focus on knowledge of grammar and vocabulary to the exclusion of language skills such as listening, speaking, and writing. Many Taiwanese students who come from families with sufficient financial means attend supplemental schools that provide instruction intended to improve the students' chances of passing entrance exams. Such educational practices are narrowly focused on acquiring the knowledge required to earn a high score on an achievement test, and are not meant to develop a student's interest in the subject for its own sake or their ability to use the language for communication.

English proficiency exams, on the other hand, may be high- or low-stakes depending on their use. When such exams are used as gate-keepers to limit access to universities, employment, promotion, overseas posting, etc., they are considered to be high-stakes because the consequences of earning a high or low score are significant to the test-takers. In Taiwan, such

high-stake English proficiency exams are associated with test preparation practices similar to those employed in preparation for taking high-stakes achievement tests. Low-stakes proficiency exams include those that are used to assess a learner's progress in learning a language but do not have a direct bearing on educational or workplace opportunities. Those English proficiency tests that are commonly taken by Taiwanese English learners in high-stakes contexts will be introduced later in this chapter.

As demonstrated in the above section, English in Taiwan is an important academic subject for students because of its inclusion on high-stakes achievement and proficiency exams. The use of English outside of classrooms is much more limited and concentrated in specific contexts such as workplaces that employ foreign employees, and in private enterprises or government agencies that deal with English-speakers. Many native English speakers in Taiwan are employed as English teachers or editors based on their English ability. Their Taiwanese colleagues often use English to speak to them, since many of these native-English speakers possess only limited proficiency in Chinese. In the electronics industry, English speakers with professional skills or experience may also be employed, necessitating the use of English if the foreign employees lack sufficient Chinese proficiency. English may also be used by Taiwanese families that employ foreign domestic help from the Philippines or Southeast Asia, since both groups may have some proficiency in English. In terms of business or government agencies that deal with foreign customers or clients on a regular basis, English may be spoken by staff in the tourism industry, including travel agencies and hotels, the banking industry, particularly foreign exchange departments, the foreign affairs departments of police departments, especially in Taiwan's larger cities where foreign residents are more concentrated, and in immigration bureaus that process visa applications and similar matters. English is also used in the academic context by

participants at international conferences conducted in Taiwan. Likewise, some of Taiwan's universities maintain international degree programs, attended by both international and local students, in which English is the language of instruction, requiring local students to attend lectures in English and write their papers in the language as well. As for the media in Taiwan, English-language movies, television programs, newspapers, and radio stations are present, but except for movies and television programs (which usually include Chinese subtitles), their impact on Taiwanese society is rather limited. While English newspapers are undoubtedly read by Taiwanese, these media outlets seem to be targeted at foreign residents and tourists based on the advertisements they carry and the content that they publish.

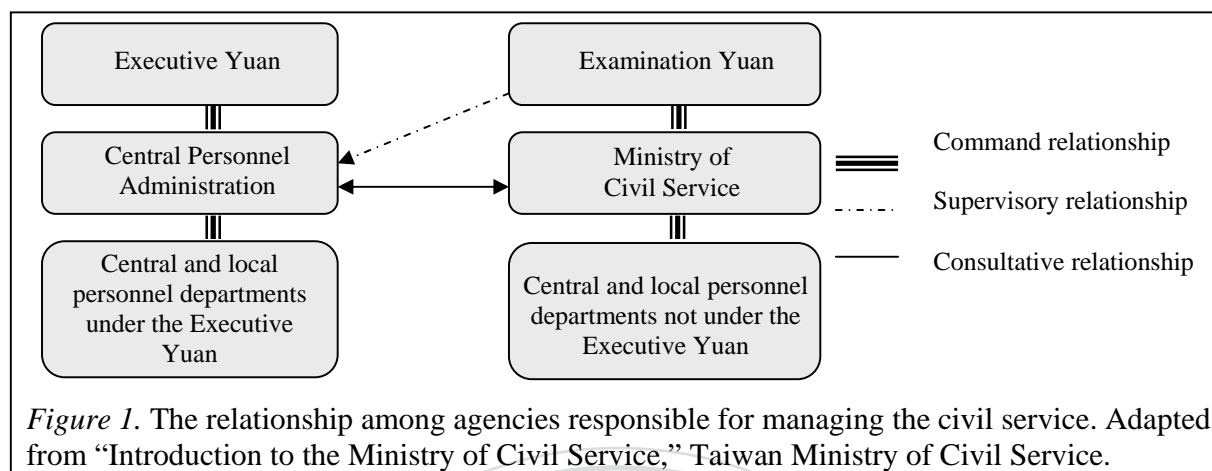
In the preceding section, the role of English in Taiwan has been introduced, focusing on its use in the context of education, both as a subject of instruction and as an examination subject, and also in the workplace, where it may be used to communicate with English-speaking customers or counterparts. In the next section, the structure of Taiwan's government and its civil service system are introduced.

Civil service in Taiwan. Taiwan's central government is headed by the president and consists of the Office of the President and five branches, known as *yuan*: the Executive Yuan, the Legislative Yuan, the Judicial Yuan, the Examination Yuan, and the Control Yuan. The Executive Yuan is led by the premier, who is appointed by the president, and is responsible for overseeing the administration of the eight ministries and twenty-nine other cabinet-level agencies that make up the executive branch of the government.¹ The Legislative Yuan is the law-making branch of the central government, and is comprised of 113 legislators who are led by the president of the Legislative Yuan, also known as the legislative speaker. The Judicial Yuan is

¹ The Executive Yuan will be reorganized in 2011 and will consist of just twenty-nine cabinet-level agencies, down from the current thirty-seven.

headed by its president and is responsible for the administration of the court system, including district courts, high courts, and the Supreme Court. The main responsibility of the Examination Yuan is to manage the civil service system, the main purpose of which is to “ensure equality of opportunity among candidates for government employment and to set uniform standards, salaries and benefits throughout the central government as well as local governments” (GIO, 2010, p.66). The Control Yuan, made up of twenty-nine members, is under the leadership of its president, and its responsibilities involve investigating unethical or criminal acts committed by government employees and agencies.

Taiwan's civil service system is organized and administered according to the laws governing the management of government employees, specifically the Civil Service Employment Act (Ministry of Civil Service, 2008). This system includes civil servants' classification, selection, hiring, screening, pay, performance evaluation, promotion and transfer, training, awards and commendation, benefits, retirement, and other aspects. Management of the civil service system is the responsibility of the Examination Yuan, and it exercises this obligation through the offices of the Ministry of Civil Service (MOCS), Examination Yuan; the Central Personnel Administration (CPA), Executive Yuan; and central and local personnel departments. The MOCS has authority over employment and discharge, performance evaluation, pay grading, promotion and transfer, and associated duties. The CPA is responsible for personnel administration in government agencies under the Executive Yuan, while the Examination Yuan supervises policies and practices in consultation with the MOCS. Figure 1 presents the relationship of the agencies responsible for managing Taiwan's civil service system.



The Examination Yuan and MOCS establish the titles, ranks, grades, and the ratio of various ranks within the civil service system. Rank is defined in the CSEA as “distinction of appointment level and basic conditions of qualification,” while grade is defined as “distinction of level of responsibilities and conditions of qualification.” Rank is classified as elementary (Grades 1 through 5), junior (Grades 6 through 9), and senior (Grades 10 through 14). Included within the civil service system are employees of administrative agencies, military public servants, and employees of state-run enterprises. Civil servants are hired on the basis of rank, grade, and series (positions with similar duties and required education level) after having passed a civil service recruitment examination. The Civil Service Examinations are divided into three categories: the Senior, Junior, and Elementary Examinations; the Special Examinations; and the Rank Promotion & Qualifications Upgrade Examinations. Access to these examinations is based on an examinee’s education level: the Senior Exam is open to those with at least a college education; the Junior Exam requires a high school diploma or above; and the Elementary Exam has no education requirement. University professors and experts in various fields are involved in planning, writing questions, grading, and conducting oral examinations.

According to the CPA in 2009, there were 226,393 civil servants, excluding teachers and soldiers, employed by the government in the elementary, junior, or senior ranks. Table 1 presents selected demographic information on Taiwan's civil servants. It may be seen that civil servants at the junior rank make up the majority of the total number, while males outnumber females at all three ranks, particularly at the senior rank. The education levels of civil servants vary, with those with undergraduate and graduate degrees making up the largest portion at the junior and senior ranks. The median age of civil servants also increases at the progressively higher ranks.

Table 1

Taiwan Civil Service Manpower Statistics

Rank	Total #	Gender	%	Education level	%	Median age
Elementary	91,728	Male	55	High school / vocational school	27	43.42 years
		Female	45	5-year college	34	
				Undergraduate	34	
				Graduate	5	
Junior	125,678	Male	57	High school / vocational school	7	45.65 years
		Female	43	5-year college	26	
				Undergraduate	45	
				Graduate	22	
Senior	8,987	Male	80	High school / vocational school	1	52.28 years
		Female	20	5-year college	8	
				Undergraduate	38	
				Graduate	53	

Note. Adapted from “Statistics: Analysis of Public Servants Manpower (Executive Yuan and Subordinate Administrative Agencies and Schools, 2009 2nd Quarter),” by Taiwan Central Personnel Administration, 2009.

Promotion of government employees is carried out in accordance with the Civil Service Promotion Act in relation to the needs of the individual agencies. The guiding principles in this process are matching individuals to their work, emphasizing equally both ability and

achievement, promoting from inside and recruiting from outside an agency, and employing “transparent, just, and impartial methods” (MOCS, n.d.). Promotion rating standards are utilized to rank employees, and points are awarded in various categories, including education level, examination scores, years of service, merits and demerits, training, professional skills, leadership, and ethics. For some criteria, such as education and examination results, points are awarded for achievement, while other criteria, such as years of service, accrue additional points annually. When one position opens, a list of qualified personnel is drawn up. Based on their promotion ratings, the names and qualifications of the top three candidates, or twice the number of positions when more than one position is open, are submitted to the agency head for selection.

This discussion of the context for this study has included a rationale for research into the uses of language tests from the perspective of critical language testing and ethical testing, a discussion of the importance of English in Taiwan, and an outline of the Taiwanese civil service system. These three subjects are fundamental to the research approach adopted in this study. An examination of the uses of language tests is carried out in order to clarify both why and how English proficiency tests are used. The use of English in Taiwan is linked to geographic, cultural, and social factors that are necessarily unique, although there are certainly parallels with conditions in other Asian nations. Taiwan's civil service system, relying heavily on the use of competitive examinations for the recruitment and promotion of employees as a consequence of its historical link with the Chinese imperial examination system, also has an essential influence on the use of English testing. Without a description of these various factors, an analysis of the use of English proficiency tests within Taiwan's civil system would be seriously disadvantaged.

Background

Challenge 2008. National development in Taiwan is guided by government planning, with strategic goals outlined in six-year plans. Challenge 2008, the national development plan covering the years 2002 to 2008, was formally approved in May 2002 and called for reforms in three areas (government, banking, and finance) and investment in four fields (cultivating talent; research, development, and innovation; international logistics; and a high-quality living environment). Ten major targets were identified, with one of those, “Cultivate talent for the E-generation,” emphasizing the development of foreign language ability, particularly English. Its aim was to strengthen the nation’s ability to meet the challenges of globalization by improving the quality of its workforce, and was linked with a call to “designate English as a quasi-official language” and encourage the use of English in daily life (GIO, 2002).

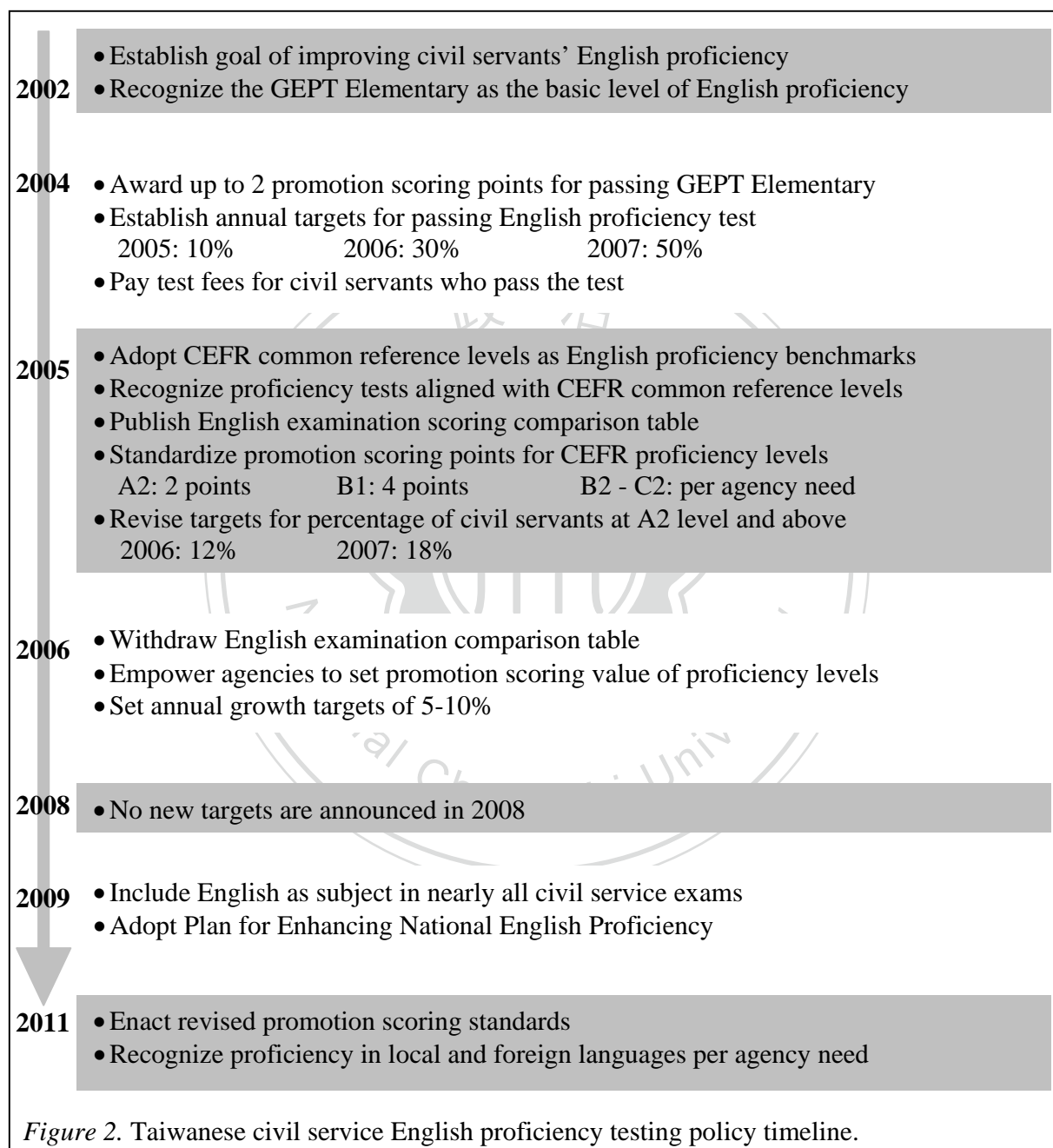
To achieve the broad goals outlined in Challenge 2008, the “Action Plan for Creating an English-friendly Living Environment” was enacted in November 2002. The rationale for this plan was to “enhance our citizen’s global capabilities and adaptability, and allow for the development of a high quality workforce that is prepared for the digital epoch” (Executive Yuan, 2005). The Action Plan specified strategies for achieving its goals, and one of those, “Training programs for English professionals,” specifically addressed the need to promote English proficiency by advocating various measures. These included: (a) establishing criteria for assessing the English proficiency of civil servants; (b) promoting English proficiency for civil servants; (c) assisting government agencies to form English language learning groups, (d) assigning personnel with high English proficiency to positions dealing with foreign visitors and affairs; and (e) improving the English proficiency of staff in key areas, including financial services.

The agencies tasked with carrying these out included the Research, Development, and Evaluation Commission (RDEC), the Ministry of Education (MOE), the CPA, supervising agencies, related agencies, and local governments. The RDEC coordinated the efforts of the other agencies in support of the goals of the Action Plan. The MOE took responsibility for establishing English proficiency levels that would be used as benchmarks, relying on experts in English education and testing for advice and recommendations. The CPA was primarily involved with incorporating the proficiency levels identified by the MOE into the promotion scoring standards, and establishing procedures for certifying the civil servants' achievement of the proficiency levels as indicated by the English language proficiency tests that they have passed. The various administrative agencies of the central government and local governments were responsible for encouraging their employees to improve their English and to certify their achievement of a basic level of proficiency by taking an English proficiency test. The key measures associated with the policy to encourage civil servants to improve their English proficiency by taking an English proficiency test are included in Figure 2.

In 2002, the MOE selected the General English Proficiency Test (GEPT) Elementary level as the basic level of proficiency for civil servants, and encouraged civil servants and the general public to take the test. The GEPT was developed by the LTTC in 1999 with the support of the MOE. At that time, 1% to 2% of all civil servants at the Junior, Elementary, or Senior ranks were believed to have some proficiency in English (Personal communication with CPA officer, March 4, 2011).

In 2004, it was estimated that 10% of civil servants had developed English proficiency, and approximately 6% had passed the GEPT Elementary. In that year, annual targets for the

percentage of civil servants expected to have demonstrated basic English proficiency were established, setting a goal of 10% for 2005, 30% for 2006, and 50% for 2007.



In 2005, the policy measures were significantly revised. The amendments included the adoption of a language proficiency framework, recognition of multiple English language proficiency tests, the lowering of targets, and the publication of an English examination promotion scoring table. The Common Reference Levels of the Common European Framework for Reference for Languages, commonly known as the CEFR (Council of Europe, 2001), were adopted as English proficiency benchmarks in May 2005. The CEFR describes language proficiency at six different levels: A1, A2, B1, B2, C1, and C2. The Global Scale of the CEFR Common Reference Levels is provided in Appendix A. With guidance from local experts in English teaching and testing, the MOE selected the CEFR A2 level, which roughly corresponds to the proficiency level of the GEPT Elementary, as the minimum level of English proficiency for civil servants. Only those tests that had demonstrated alignment with the CEFR would henceforth be recognized as criteria for awarding promotion scoring points to civil servants for their English proficiency. In September of that same year, the CPA posted an English examination promotion scoring table on its official website. This is presented in Appendix B. The table specified the tests that were recognized, linked scores on the various tests to the CEFR proficiency levels, and indicated how many promotion scoring points a civil servant would obtain for each proficiency level (CPA, 2005). By the end of the year, only 6% of civil servants had earned a score at the A2 level on a recognized test. The targets for 2006 and 2007 were subsequently lowered, with a goal of 12% for the first year and 18% for the second.

In 2006, the English examination promotion scoring table was withdrawn. However, the policy of recognizing tests aligned with the CEFR proficiency levels remained unaltered. Under the new policy terms, administrative agencies were given the responsibility of setting the promotion scoring values for the different English proficiency levels according to their specific

needs. These agencies were urged to meet the target for 2007, and to plan for annual increases of 5% to 10% once the target had been reached. The English proficiency targets for 2006 and 2007 were met, with 12% of civil servants achieving at least A2 in the first year, and 18.9% reaching that benchmark in the second year. In 2008, no further targets were published. English proficiency remained a criterion for promotion scoring, but the promotion scoring values associated with different CEFR proficiency levels were set by the administrative agencies.

In 2009, the MOEX announced that English would be included as a subject in nearly all civil service examinations, with the exception of special examinations for the disabled, beginning on January 1, 2010. The MOEX's mid-term plans for the testing of English included increasing the weighting of English among general examination subjects and investigating the feasibility of implementing the testing of English and other foreign languages as special examinations (MOEX, 2008). In addition, 2009 saw the release of the Plan for Enhancing National English Proficiency (PENEP), authored by the RDEC. The PENEP made little direct mention of English testing and emphasized providing civil servants with opportunities to improve their English through participating in learning activities, attending training courses, and taking part in overseas study.

In 2010, a proposal for revising the promotion scoring standards for civil servants was adopted, with the new standards coming into force on March 1, 2011. The new promotion scoring standards no longer specifically recognized English proficiency as a criterion. Instead, administrative agencies were directed to award promotion scoring points for "language proficiency," defined as ability in local or foreign languages, as per the need of the agency. No promotion scoring values for specific proficiency levels were indicated on the CPA's most recent promotion scoring standards. By the end of 2010, approximately 25% of civil servants had

gained proficiency in English: 18% had achieved the A2 level; 4.5% had reached B1; and 2.1% the C1 level (Personal communication with CPA officer, March 4, 2011).

The developments described in this section are intended as an outline of the process of the modification of the policy to recognize scores on English proficiency tests as criteria for the awarding of promotion scoring points to civil servants. Further details of this policy cycle are discussed in Chapter 6.

Tests recognized by the government. In this section, the English language proficiency tests that were recognized as having been aligned with the CEFR will be introduced. The individual tests are organized according to the agency that developed them. First, the Test of English as a Foreign Language (TOEFL) and the Test of English for International Communication (TOEIC), produced by the Educational Testing Service (ETS) in the United States, are reviewed. Then two tests developed by Cambridge ESOL in the UK, the International English Language Testing Service (IELTS) and the Business Language Testing Service (BULATS) are discussed. Finally, the Foreign Language Proficiency Test (FLPT) and the General English Proficiency Test (GEPT), developed and administered by the Language Training and Testing Center in Taiwan, are presented. The Cambridge ESOL Main Suite Examinations and the LTTC's College Student English Proficiency Test (CSEPT) are not included in this discussion even though they were specified in the English examination promotion scoring table. It was expected that the number of civil servants taking these examinations would be quite limited. The TOEFL iBT (internet-based test), launched in Taiwan in 2006, is included in this review, despite the fact that it is not included in the scoring table. It is currently the only TOEFL test administered in Taiwan.

TOEFL. The Test of English as a Foreign Language (TOEFL) was developed by the Educational Testing Service (ETS) of Princeton, New Jersey, U.S.A. The TOEFL has been administered since 1964, and is annually taken by more than 750,000 people. ETS claims that the test “measures ability to use and understand English at the university level” (ETS, 2011). TOEFL scores are mainly used by universities as admission qualification for foreign students, but they are also used by immigration departments and licensing agencies for certification, as well as by individuals who want to gauge their progress in English.

The TOEFL PBT (paper-based test) was first administered in 1964. While the PBT is still administered in test centers without access to the Internet, it has not been given in Taiwan since 1998, when it was replaced by the TOEFL CBT. The TOEFL PBT contains three sections, Listening Comprehension, Structure and Written Expression, and Reading Comprehension, and takes approximately 3 and one-half hours to complete. All items in these three parts are multiple-choice questions (MCQs). Those taking the TOEFL PBT are also required to sit for the 30-minute Test of Written English (TWE). Unlike the iBT, the PBT does not include a speaking section. Score reports are mailed to institutions and test-takers approximately five weeks after the test date. Scores on the TOEFL PBT are reported on a scale from 310 to 677, with TWE scores reported separately on a scale of 0 to 6. Scores on the TOEFL PBT remain valid for two years after the test date.

The TOEFL CBT was introduced by ETS in 1998 and phased out with the introduction of the iBT in 2006. The TOEFL CBT consists of four parts, Listening, Structure, Reading, and Writing, of which the first two are computer adaptive, meaning that successive test items are chosen based on the test-taker's answer to a preceding question. In the first two sections, Listening and Structure, there is no time limit, but test-takers may not leave an item unanswered.

The number and range of test items that a test-taker must complete in these sections will vary depending on the answers provided. The reading section includes MCQ items in addition to items that make use of other methods. In the writing section, test-takers must write on a topic chosen by the computer, and have the option to write on paper or with a computer keyboard.

The TOEFL iBT was introduced in 2006 with the phase out of the TOEFL CBT (computer-based test). Unlike the CBT, the iBT is not computer adaptive, having the same range of questions in any form of the test. The test contains four sections, Listening, Reading, Speaking, and Writing, and takes approximately four to four and one-half hours to complete. The iBT employs an integrated skills approach to testing all four language skill areas, with the speaking and writing tests requiring test-takers to read and/or listen before speaking, and read and listen before writing. In the Listening section, test-takers hear lectures, classroom discussions, and conversations before answering between 34 and 51 questions. In Reading, they answer 36 to 70 questions based on academic texts. There are six tasks in the Speaking section, and test-takers are asked to express an opinion on a familiar topic and then speak based on reading and listening tasks. TOEFL claims that the test's content simulates actual classroom tasks such as comprehending a lecture or participating in a discussion and that the language used in the test "closely reflects what is used in everyday academic settings" (ETS, 2011a).

The total number of points possible in the TOEFL is 120, with marks in each of the four parts converted to a 30-point scale. Scores are available online in fifteen days, and the four skill scores and the total score are reported to test-takers and designated institutions. To ensure that TOEFL scores are used appropriately, ETS recommends that institutions consider the score profile of test-takers, and not just the total score, since individual departments may want to assign greater priority to one skill or another depending on need. It also suggests that institutions

conduct their own validation studies when setting cut scores, and conduct periodic reviews to confirm that cut scores are providing adequate information for admission decisions. The fee for taking TOEFL varies between US\$150 and US\$225 depending on the country in which the test center is located. As of 2011, the fee for those taking the TOEFL in Taiwan is US\$160 (approximately NT\$4,800).

TOEIC. The Test of English for International Communication (TOEIC) was developed by ETS and first administered in 1979. ETS claims that TOEIC “measures the everyday English skills of people working in an international environment” and suggests that the test is most appropriately used in the hiring, placement, and promotion of employees in an organization in which “workplace/everyday-life English is a required job skill,” or for measuring English proficiency levels of students or individuals over time (ETS, 2011b). In 2010, over six million candidates worldwide took TOEIC, with 180,933 taking the test in Taiwan, an 18% increase from 2009 (ETS, 2011c).

In its original form, TOEIC tested only listening and reading, with 200 MCQ items divided evenly between two sections. In 2006, along with the launch of the TOEIC Speaking and Writing Test, the Listening and Reading Test was redesigned in order to bring the test more closely into alignment with theories of communicative competence and simulate more closely the communication styles and language contexts of international business (Powers, Kim & Weng, 2008). The number of questions has not changed, nor the time necessary to complete the test, but some tasks have been modified. In the Listening section, test-takers first answer 10 questions about photographs. Then they hear thirty questions and choose the best response. Next, they hear 10 conversations and answer three questions about each of them. Finally, they hear 10 short talks and answer three questions for each. In the Reading section, test-takers answer 40 questions

about incomplete sentences. Then they answer 12 questions about text completion. In the next part, they read seven to ten individual texts and answer two to five questions for each one. Finally, test-takers read four pairs of reading texts and answer five questions per pair of texts. The marks for each section are converted to a scale of 5 to 495 and then added to produce a total score.

The TOEIC Speaking and Writing Test takes approximately 80 minutes to complete and is set in contexts that examinees might encounter in their daily life or at the workplace. The Speaking Test lasts about 20 minutes and includes six different tasks. The tasks include reading aloud, describing a picture, responding to written questions, responding to questions based on information provided, proposing a solution to a problem, and giving an opinion on a topic. In most of the tasks, examinees are provided with time to read the prompt material or prepare their response. Test-takers responses are digitally recorded and scored by multiple examiners. The recordings are evaluated according to the following criteria: pronunciation, intonation and stress, grammar, vocabulary, cohesion, relevance, and completeness. The sum of all ratings for each question is converted to a scaled score of 200. The Writing Test lasts approximately one hour and includes three task types: writing individual sentences based on pictures, responding to a written request (email or letter), and writing an opinion essay of at least 300 words. The examinees' writing is evaluated according to criteria that include grammar, relevance, quality and variety of sentences, vocabulary, organization, and coherence. The sum of all ratings for each question is converted to a scaled score of 200. In addition to receiving a scaled score for each part of the test, test-takers are assigned a proficiency level on a scale of 1 (lowest) to 8 (highest) for speaking and 1 to 9 for writing. Score reports for both the Listening and Reading Test and the Speaking and Writing Test are available in about two weeks and include separate

marks for each part of the test, a total score, a percentile rank based on scores from the previous three years, score descriptors, and abilities measured. TOEIC scores remain valid for 2 years after the test date.

IELTS. The International English Language Testing Service (IELTS) is jointly managed by the British Council, IDP: IELTS Australia, and Cambridge ESOL. (In this paper, IELTS will be referred to as a product of Cambridge ESOL for ease of reference.) IELTS was first administered in 1989 after the completion of a validation study of its predecessor, the English Language Testing Service (ELTS), led to the construction of the new test (Hyatt & Brooks, 2009). IELTS is used by educational institutions, governments, professional bodies, and businesses and is intended to aid in the recruitment of applicants with the ability to communicate effectively in English (IELTS, 2009). Over 1 million people take IELTS annually. In marketing itself to test users, IELTS cites its global recognition, convenience, and suitability for purpose based on its expert design and concern for fairness and accuracy.

The test has four parts, each of which tests a separate language skill: listening, speaking, reading, and writing. All candidates take the same listening and speaking tests, while those who are applying for admission to a university or college take the Academic Reading and Writing modules, and those taking IELTS for employment purposes take the General Training Reading and Writing components. The listening test is comprised of four sections, the first two with a conversation and a monologue set in social contexts, and the remaining two, a conversation between up to four speakers and a monologue, set in training or educational contexts. Test-takers hear a range of native-speaker accents. The Academic Reading component consists of three authentic texts that are said to be “recognizably appropriate” for test-takers planning to enroll in undergraduate or graduate programs. The General Training Reading component also

features authentic texts, one longer and several shorter, of types that test-takers might be expected to read on a daily basis in an English-speaking country. Both modules of the writing component consist of two tasks. In Task 1 of the Academic Writing module, test-takers summarize the information expressed in a graph, table, or chart. In Task 2, they must write an opinion essay in a formal style based on a viewpoint, argument, or problem. In Task 1 of the General Training Writing module, test-takers write a letter in response to a presented situation; while in Task 2, they write an opinion essay similar to that in the Academic Reading component, but in a less-formal style. Speaking tests are conducted face-to-face with examiners, a method that IELTS explains is closer to “a real-life situation” than having test-takers respond to recorded prompts. In Part 1, test-takers answer questions about a range of familiar topics. In Part 2, they are asked to speak on a topic for two minutes and to answer one or two questions related to their talk. In Part 3, the topic in Part 2 is expanded, and test-takers answer further questions.

IELTS is offered up to four times a month, and takes approximately 3 hours to complete. There is no limitation on how often IELTS can be taken. The listening, reading, and writing sections must be completed on the same day, but the speaking test may be taken up to seven days earlier or later than the other sections. Scores on each section of the test are converted to the IELTS Band Score Scale, from 1 (the lowest) to 9 (the highest), with scores being awarded for full and half bands. IELTS explains that organizations using the test should set minimum scores depending on their specific requirements and consider scores on the individual components in addition to the total score when assessing an applicant's ability. Test results are available thirteen days after the test, with one copy of the score report being mailed to the test-taker and up to five other copies sent to designated institutions. The test fee for test-takers in Taiwan is

NT\$5,100 (approximately US\$170). IELTS scores are valid for up to two years after the date of the test.

BULATS. The Business Language Testing Service (BULATS) was developed and is jointly run by Cambridge ESOL, the Alliance Française (France), Goethe-Institut (Germany) and the Universidad de Salamanca (Spain) in more than thirty countries. BULATS is designed to test the foreign language ability of employees who need to use a foreign language at work (BULATS, 2011). It is also used by participants in professional/business courses in which a foreign language is an element of the course. BULATS is a multilingual test available in English, French, German, and Spanish and tests all four language skill through three tests: the Listening and Reading Test, the Speaking Test, and the Writing Test. There are three versions of the Listening and Reading Test. The Standard Test is a paper-based test that was introduced in 1997. Two computer-based tests, one administered over the Internet and one that is run via a CDROM, were introduced in 2002. The Speaking Test and the Writing Test were also launched in 2002. Periodic revisions are made to the test content and format on all forms of BULATS at the same time. The topics included in the test include personal information, the office/business environment, relations with clients, travel, health, buying and selling, products and services, and general interest topics.

The three versions of the Listening and Reading Test assess listening and reading skills and knowledge of grammar and vocabulary. Both versions of the computer-based tests are computer adaptive and take approximately 60 minutes to complete, while the Standard Test takes 110 minutes. The test methods include multiple choice, text completion, and error identification and correction. On the Listening Test, examinees hear UK, North American, and Australasian accents. Test results are available immediately, and marks include an overall score as well as

individual scores for listening and reading and language knowledge. The scores on each section are weighted differently, with raw scores being converted to a score on a scale of 100. In addition to numerical scores, score reports also include the CEFR level achieved and ALTE (Association of Language Testers in Europe) can-do statements.

The BULATS Speaking Test is administered to test-takers individually by an examiner and takes twelve minutes. There are three parts to the test, which is digitally recorded and scored by a second examiner. Part 1 is an interview, and the test-taker answers questions about familiar topics. In Part 2, the test-taker gives a short presentation and answers one or two follow up questions from the examiner. In Part 3, the test-taker participates in a simulated situation requiring the exchange of information with the examiner, and then discusses a related topic with the examiner. The test-taker's oral proficiency is assessed with reference to six criteria: accuracy of grammar and vocabulary, range of grammar and vocabulary, fluency, pronunciation, discourse management (development and organization of ideas), and interactive communication (contribution to conversation). The test-taker's score on the Speaking Test is reported as a CEFR level with three sub-bands (-, =, and +) for each level.

The BULATS Writing Test includes two tasks and lasts 45 minutes. In Task 1, test-takers write a response of approximately 60 words in length to an email or letter. Task 2 requires test-takers to write a report of approximately 200 words in length in response to given information. Both tasks are marked by two examiners with reference to scoring criteria that include accuracy and appropriateness of grammar and vocabulary, organization of ideas, and effectiveness. Like the Speaking Test score, the Writing Test score is reported as a CEFR level with sub-bands. The BULATS test fee is NT\$1,300 for the Standard test, NT\$1,500 for the Writing Test, and NT\$1,500 for the Speaking Test.

FLPT. The Foreign Language Proficiency Test (FLPT) was developed by Taiwan's LTTC and first administered in 1965 (LTTC, 2011a). It is a norm-referenced test that is mainly used by public and private institutions to select qualified candidates for promotion or overseas training and assignment. The FLPT is also used by colleges and universities to evaluate student learning and to select qualified candidates for international exchange programs (LTTC, 2009). The FLPT is available in five languages, English, Japanese, French, German, and Spanish. Paper-based versions of the test are available in all five languages, while a computer-based form of the English version of the test has been developed and was first administered in 2010. The test is administered monthly or by arrangement with test-using institutions.

In the Listening Section, test-takers select the best answers to questions, choose a phrase with the same meaning as an utterance on the tape, and answer comprehension questions about conversations and short talks. The Usage Section assesses test-takers' knowledge of grammar and sentence structure. In the Vocabulary and Reading Section, items are designed to test knowledge of vocabulary and idioms as well as reading comprehension skills. These three sections take approximately two hours to complete. The Speaking Test takes a further 15 minutes and includes reading aloud, translation, answering questions, discussion, and picture description. The Listening, Usage, Vocabulary and Reading portions of the test employ multiple-choice questions and are machine scored. The Speaking Test is tape-mediated and marked by two examiners. Scores on the FLPT Listening, Usage, and Vocabulary and Reading Test are reported on a score scale of 120, while scores on the Speaking Test are reported on a scale of 0 to 4, with four sub-bands (O+, 1+, 2+, and 3+). The test fee for all four parts is NT\$1,750, but the Speaking Test (or the Listening, Usage, Vocabulary and Reading) may be

taken separately for NT\$1,000. Score reports are mailed to test-takers approximately three weeks after the test date.

GEPT. The General English Proficiency Test (GEPT) is a five-level criterion-referenced test of the English language ability of learners in Taiwan. The GEPT was developed by the LTTC with the support of Taiwan's Ministry of Education and is administered at testing centers throughout Taiwan. The GEPT grew out of a project by the LTTC to create a test to assess the English language ability of learners in Taiwan and encourage people of all ages to learn English. The GEPT Intermediate Level was introduced in 2000, and this was followed by the Elementary and High-Intermediate levels in 2001. The Advanced Level Test was launched in 2002, and the Superior Level came out in 2004. The GEPT tests all four language skills and is administered in two phases. Test-takers must pass the first stage, consisting of the Listening and Reading Tests, before they can register for the second phase, the Speaking and Writing Tests.

The GEPT is in wide use in Taiwan, having tested nearly four million examinees since its launch in 2000 (LTTC, 2011b). GEPT scores are used as a criterion for the promotion of civil servants, for decisions related to admissions, placement, or graduation by public and private institutions of higher learning, and for determining the English proficiency of employees by businesses. The various levels of the GEPT were originally designed to align with the expected English proficiency of students at different academic levels, but they were later also linked to CEFR levels (LTTC, 2011c). The Elementary Level corresponds to the expected English proficiency level of a junior high school graduate in Taiwan and also to CEFR A2. The Intermediate Level approximates the proficiency level of a high school graduate in Taiwan and CEFR B1. The High-Intermediate Level aligns with the English proficiency of a university graduate whose major is not English and corresponds to CEFR B2. The Advanced Level is

equated to the English proficiency level of a Taiwanese university graduate who majored in English or one who graduated from a university in an English-speaking country, and is aligned with CEFR C1. The Superior Level is approximately equivalent to the English proficiency of a university graduate with “native English-speaking ability” and CEFR C2 (LTTC, 2011d).

The first three levels of the GEPT, Elementary, Intermediate, and High-Intermediate, share a similar format and are given periodically throughout the year at test centers located in cities around Taiwan. The Listening and Reading Modules in each of these levels contain multiple choice items and take between 55 and 85 minutes to complete. The Listening Modules include tasks such as picture description, answering questions, short conversations, and short talks. The voices on the tape speak North American English. The tasks in the Reading Modules include sentence completion, cloze, and reading comprehension items with text types such as notices and ads, letters, newspaper articles and expository essays. The Second Stage Test takes between 50 and 70 minutes to complete. The Speaking Module is tape mediated, and test-takers complete tasks including repeating, reading aloud, answering questions, describing a picture, and discussing a topic. The Writing Module includes Chinese-to-English translation, sentence writing, and guided paragraph writing. The tasks in the Speaking and Writing Modules are marked by pairs of trained examiners according to scoring criteria. The test fee for both stages of the GEPT at these three levels ranges between NT\$1,080 and NT\$2,000.

The GEPT Advanced Level is held once per year, with the First Stage Test given at two test centers and the Second Stage Test conducted only at the LTTC's Taipei campus. The First Stage Test takes approximately two hours. The Listening Module includes short conversations, longer conversations, and long talks, and items consist of both multiple choice and constructed response. The Reading Module includes careful reading as well as skimming and scanning tasks,

also with both multiple choice and constructed response items. The Second Stage Test includes Writing and Speaking Modules. In the Writing Module, test-takers write two essays, one based on verbal input and the other based on non-verbal input, that require them to summarize and give opinions. The Speaking Module of the GEPT Advanced Level differs from the three lower levels in that pairs of test-takers take part in a face-to-face test in which they interact with an interlocutor and each other. The tasks in the Speaking Module include a warm-up interview, a discussion, and a presentation. The test-takers' performance in the both the Writing and Speaking Modules are marked by pairs of examiners according to rating criteria. The test fee for the First Stage is NT\$1,650 and NT\$2,300 for the Second Stage.

The GEPT Superior Level is an integrated test that takes approximately three hours and fifty minutes to complete. This level of the GEPT has been given only once, in 2004, and prospective test-takers must qualify by producing a certificate attesting that they have passed the GEPT Advanced Level or Cambridge CAE or received a TOEFL score equivalent to at least a 280 on the TOEFL CBT or 114 on the TOEFL iBT, or an IELTS score of at least 7. The Integrated Writing Module consists of a 750-word essay based on a video or radio program of 10 to 15 minutes in length and an article of approximately 3,000 words. The Integrated Speaking Module, comprised of a presentation with follow-up questions and answers, lasts approximately fifty minutes. The test fee depends on the type of content required for the test and the number of examinees.

Conclusion

In this chapter, the context and background to this research into language testing policy in Taiwan were discussed. The research context was introduced first, explaining the importance of critical language testing and ethical testing for an investigation of test use. Next, the role of the

English language in Taiwan was reviewed, examining its links to Taiwan's political, economic, social, and educational development. Because this research focuses on a language testing policy that directly affects employees of Taiwan's government, the civil service system was introduced, offering demographic data and details about the recruitment and management of civil servants. Having established the broader context for this thesis, the specific background was reviewed, beginning with the Challenge 2008 development plan that called for improving the English ability of Taiwan's government employees. Next, a timeline for the testing policy was presented, highlighting the key developments in the policy cycle from 2002 to 2011. The significance of the MOE's 2005 adoption of the CEFR Common Reference Levels as English proficiency benchmarks for English learners in Taiwan and the creation of an English examination promotion scoring table for civil servants were discussed. Finally, the chapter ended with an introduction of the content, format, and uses of the various English proficiency tests taken by civil servants in Taiwan in order to obtain promotion scoring points on the basis of the alignment of these exams with the CEFR Common Reference Levels.

Chapter 3

Literature Review

This literature review begins by discussing ancient China's imperial exam system, which remained in existence for over two thousand years. The current testing of civil servants in Taiwan is influenced both by the history of testing in Asia and the development of objective testing methods in the West in the twentieth century. Next, literature related to language policy is discussed, starting with a policy analysis framework that will be employed in this research. Then language testing policies are introduced, beginning broadly by examining different analytical approaches, and then turning to the influence of the CEFR and challenges associated with the adoption of its common reference levels as a proficiency yardstick. Finally, this section ends with a discussion of current English language testing policy issues in Asia and Taiwan. The third section of this chapter addresses the use of high-stakes language tests. The topic of consequential validity of language tests is introduced in order to establish the key concerns for appropriate test use and validation procedures. The theory of assessment use arguments is next outlined, and the potential value of this method for validation is discussed. The issues of test washback and impact are reviewed with the goal of helping to establish the rationale for carrying out this research project. In the final section, the literature review turns its attention to lifelong learning and motivation associated with language learning and testing. The promotion of lifelong learning is one of the rationales behind the effort to encourage civil servants in Taiwan to improve their English language ability. To identify and explain the role of factors that could influence the efforts of Taiwan's government employees to develop their English proficiency, theories of adult learning and education are discussed. Motivation for learning plays an essential role in education and is influenced by testing. Efforts to describe the influence of motivation on

Taiwanese learners of English are reviewed, and the potential implications of such context-specific motivation for civil servants in Taiwan are addressed.

History of Civil Service Exams

The use of examinations as a gate-keeping device for entrance to government employment began with the Chinese imperial examination system. The long history of exam-oriented education in Asia has had a strong influence on test use and educational practice in Taiwan. A review of the history of civil service exams in China, as well as the use of such exams in Europe in the nineteenth century, may offer insight into the use of English proficiency examination as a criterion for the promotion of civil servants in Taiwan. Of particular relevance is discussion of the test preparation strategies employed by the exam-takers, the changes to the exam procedure and contents that were intended to minimize unintended consequences of test use, and the narrowing of curriculum that occurred as a result of the use of high-stakes tests.

Cheng (2010) notes that the imperial examination system, founded in 206 BCE during the Han Dynasty and lasting until 1904, had as its purpose the selection of candidates for service in the imperial government on the basis of their merit rather than their membership in the aristocracy. Cheng notes that the imperial exam had two complementary purposes, the selection of officials and the testing of knowledge and ability. The content and categories of the examination changed over time, with factors such as the separation of the functions of selection and appointment, the adoption of printing, which led to an increase in the number of candidates and increased its difficulty, and the adoption of the eight-legged essay format, which turned the examination into a “contest of regurgitation,” all having an impact. Cheng describes the examination as extremely competitive, noting that 2 million candidates took the exam at the

district level in 1850, and only 300 passed the metropolitan exam, giving candidates a 1 in 6,000 chance of passing all three levels.

Suen and Yu (2006) similarly examine the development of these high-stakes exams, noting the impact of test preparation strategies on exam scores and attempts by exam officials to repeatedly modify the test tasks over time in order to reduce the impact of test preparation. These authors found that the imperial exam system led to a narrowing of the education system and a rise in the use of exam-taking strategies. Scholars writing at that time judged such methods as having the effect of subverting the intended purpose of the exams, thereby weakening the validity of their scores. Suen and Yu conclude that the exam officials were never able to revise the exam content sufficiently to eliminate the influence of exam-taking strategies, and that the continued reliance on high-stakes testing in the modern day is bound to encounter the same difficulty.

The view that these authors share is significant in that it relates a modern phenomenon to a historical trend, but the context of high-stakes testing has clearly been altered in the more than two thousand years since the imperial examination system was first introduced. One could argue that the stakes currently associated with civil service exams are much lower than in past centuries given the development of non-governmental social institutions. Similarly, the test methods now being employed, particularly on proficiency tests that assess communicative competence, are substantially different than those used in early exams. Given the long history of the use of examinations in Asia, it may not be possible to draw too many conclusions about long-term trends based on observation and analysis of the use of test methods in use today or in recent decades.

The adoption of civil service testing in England in 1855 was a direct result of the need to select staff for the British colonial government in India, but it had its root in the introduction of Chinese testing principles to Europe by the Jesuit missionaries who returned from Asia. (Cheng, 2009; Spolsky, 1995) Like the Chinese imperial exams, competitive exams in Europe were a means for replacing patronage in the selection of individuals for government positions, but it was also acknowledged that such testing did produce a narrowing of the education process, subordinating teaching to testing (Cheng, 2009). Spolsky (2009) describes the introduction of competitive exams to the United States in the early twentieth century as being understood to have three purposes: to stimulate education, to exert control over the education system, and to select civil servants or award professional qualifications. It was in the U.S. that the development of intelligence testing and the subsequent founding of statistics occurred, and these were to be key antecedents to the development of objective testing and psychometric measurement, two innovations that increased the reliability of tests while temporarily putting aside questions about the validity of the measurements that were being made.

The introduction of the Chinese imperial examination system to the West was significant in that it influenced examination procedures used for both government employment and education. The development of objective testing and psychometric methodology in the West transformed the traditional testing practices that had been employed in Asia for two thousand years. While the testing methods have changed, the purposes of testing have largely remained the same. Performance on competitive examinations is seen to be an indicator of individual's merit and also of potential contribution to society. The use of tests for the recruitment of civil servants in Taiwan today is seen as democratic, fair, and expedient. It may be that the use of tests as gate-keepers in Taiwan benefits as much from the history of examinations in Asia as

from the development of objective testing in the West; the advantages of the two methods making up for the other's shortcomings. Tests can serve as both an objective method for selection and as an indicator of ethical standards with deep cultural roots.

Language Policies and Testing

This section begins with a discussion of the policy analysis framework that is employed in this research. Next, it considers principles of language planning and policies. Then it turns its attention to the CEFR. Finally, it reviews the literature on language testing policies in Taiwan.

Policy analysis framework. Policy analysis frameworks are useful for describing policy measures in a systematic manner that facilitates analysis of the component stages in a policy cycle and the relationships among the various agencies and institutions. It is believed that the application of a policy analysis framework could clarify the roles of the different agencies responsible for making and implementing the English language proficiency testing policy for civil servants. Further, it may provide a basis for evaluating the effectiveness of the policy and help to identify factors that influence the policy's outcomes.

In 2002, the Geelhoed-Schouwstra (G-S) framework for policy analysis was created for use by the Dutch Ministry of Finance in the evaluation of national and international policies. The framework was based on a conceptualization of the policy-making cycle as being composed of six steps: goals, objectives, methods/instruments, activities, performances, and evaluation. The purpose of the framework was to "identify factors that cause policy outcomes to diverge from the intended results" and may be used both when making and evaluating policy (Geelhoed and Ellman, 2006, p. 1). One drawback of the original framework was that it did not establish the way in which the policy would bring about its intended effects; therefore, the basic framework was expanded to incorporate conceptual and institutional frameworks to account for

key contextual elements that influence policy-makers. The conceptual framework accounts for the effects of ideology, social norms and values, fundamental theories and assumptions, definitions, and attitudes and behaviors and is itself subject to influence by culture, geography, and history, both individual and collective. The institutional framework is composed of elements that include the political, social, and economic setting, the institutional and legal setting, and stakeholders. The authors note that the various elements of the institutional framework will be weighted differently according to the type of policy that is being analyzed, and that education policies (such as the testing policy that is the focus of this study) might warrant much greater attention to elements of social setting than the others. They recommend that the basic framework be applied first in order to identify the key features of a policy, and then the conceptual and institutional frameworks be systematically applied to bring greater transparency to the analysis. A key feature of this extended framework is that it focuses attention on policy stakeholders and the relationships between them that may influence the outcome of the policy. The conflicting interests of the various stakeholders could have a negative influence on a policy, and are thus possible sources of divergence from expected outcomes.

The extended G-S framework is used to trace the development of the testing policy from the goals it was intended to accomplish through the evaluation of the performances that it produced. The conceptual and institutional frameworks provide an opportunity to introduce contextual features into the analysis of the decision making that is at the core of the policy cycle.

Language policies. Policies that call for language testing are part of what Cooper terms *language planning*. This he defines as “deliberate efforts to influence the behavior of others with respect to the acquisition, structure, or functional allocation of their language codes” (Cooper, 1989, p. 45). Language planning plays a role in the political process as decisions are made by

authorities and then implemented for the benefit of society via policies. Cooper outlines a descriptive decision analysis as an examination of “what people actually do in arriving at their decisions, good or bad,” with a focus on the three separate roles of individuals, institutions, and the public arena. He offers this summative statement to express how such an analysis would gather data about language planning decisions: “Who makes what decisions, why, how, under what conditions, and with what effect” (Cooper, p. 88).

Insofar as Cooper addresses acquisition planning as one subset of language planning, he recognizes two fundamental bases for such efforts. These are overt language planning goals, such as acquisition of a second or foreign language, and the methods employed to reach the goal. These methods could include the creation or improvement of opportunity to learn, the incentive to learn, or a combination of both opportunity and incentive. Cooper notes that it is difficult to evaluate the effectiveness of acquisition policies, whether the degree of improvement or the relative contributions of various factors. Generalizing about effective acquisition planning, Cooper makes two points that are relevant to this research into language policies in Taiwan. First, he asserts that unless the language serves a necessary function for the targeted population, the policy is unlikely to be effective. This implies that the goals of the Action Plan adopted by Taiwan's government, which the testing policy is designed to support, are more likely to be achieved if they are based on actual need. If policy makers overestimate civil servants' need for English proficiency, the policy's goals may prove difficult to realize, and resistance to the policy may be encountered. Secondly, implementing a policy may “require repeated efforts by planners to cope with the resistance of those they seek to influence” (Cooper, p. 185). As the testing policy timeline in Chapter 2 shows, the initial targets for the number of civil servants that had demonstrated their English proficiency by taking an English test were not met, and additional

measures to encourage compliance, such as increasing the value of the incentives, were put into effect

Shohamy (2008) traces a development within the field of language testing that sees a shift from focusing mainly on psychometric traits of tests toward a greater interest in test impact, the ethics of testing, fairness, and consequences. Promoting use-oriented design and critical language testing, Shohamy calls for the questioning of tests, not just how tests are designed, or what content they test, but also their societal role, such as how tests-takers are affected by tests, what knowledge is created by tests, what decisions are made based on test results, and what motivations are behind the introduction of tests (Shohamy, p. 363). Testers have recognized the need to examine the role of tests in relation to their powerful impact on society and washback on educational practices. Policy makers also understand the power of tests to influence the behavior of others, and Shohamy (2001) has argued that it is this awareness that has led those in authority to introduce tests as instruments of policy.

According to Shohamy (2008), language testing and language policy influence each other in two directions. One is the policy of introducing the test itself, as an entrance or graduation requirement, or as a promotion criterion, and this perspective, she says, has been the subject of little research, particularly into the intentions, reasons, and arguments for introducing particular language tests (Shohamy, p. 366). The other direction is related to the consequences of introducing a test, specifically the influence on language practices. This research study makes an effort to examine both directions of the English proficiency testing policy. The interviews with representatives of the agencies responsible for formulating and implementing the policy are intended to identify why the goal of encouraging civil servants was adopted and how using tests to certify English proficiency was expected to produce the desired result. The survey of

employees of the three agencies is intended to explore the consequences of introducing the testing policy, looking at its influence on studying, motivation, and on professional duties.

One such consequence of test policies that is of interest to Shohamy is the effect of creating uniformity with regard to ideologies about language learning and use. Writing on the CEFR, Shohamy identifies several areas, such as the absence of language context and advanced language proficiency use, in which the adoption of its rating scales could lead to a prescriptive view about the hierarchy of language acquisition and use that is not based in reality. Rejecting language policies that impose uniform ideologies through the imposition of tests with insufficient regard for their consequences, Shohamy calls for adaptive language policies that recognize that language learners bring different knowledge and experience to the act of language learning. From this perspective, one could expect that a test policy that assumes a uniform need for English among employees in different agencies and different positions might be ill-suited to the needs and interests of those it seeks to influence. An adaptive policy would recognize diversity and avoid imposing standards for language learning and use that are not based in actual conditions.

The way forward, Shohamy proposes, is asking how tests can lead to language policies that do not simply promote a top-down view. From this viewpoint, the challenge for testers is to explore how tests can be “more inclusive, democratic, open, just, fair and equal, and less biased” (Shohamy, 2008, p. 371). Shohamy calls on testers to recognize language tests not just as measurement tools, but as devices with the power to influence education, society, and language policy. She sees that language testing and language policies share a common interest, not to “confirm ‘bad policies’ but rather to pose questions about how language tests can be instrumental in the development of ‘good’ and ‘just’ policies which are more in line with current language

practices” (Shohamy, p. 371). The significance of this view of tests and test policies as powerful instruments is that the power need not be wielded without regard for the interests of those who are affected. The goals for which tests are employed can benefit society and individuals alike, and widening the range of stakeholders who participate in the discussion of how best to utilize tests and formulate testing policies can increase the value that they may produce. Language tests, like other types of tests, have a history and a future. The present task is to develop an understanding of how to maximize benefits of test use in the future based on an analysis of their use in the past.

CEFR. In 2005, the MOE adopted the CEFR Common Reference Levels as an English proficiency yardstick. An English examination scoring table produced by the CPA in that same year equated scores on different English proficiency tests on the basis of their alignment with the CEFR levels. Concern that the Common Reference Levels were not a suitable basis for test comparison led to the withdrawal of the table the next year. The use of the CEFR's Common Reference Levels as proficiency standards in Europe and countries around the world has been controversial. A review of the key issues in regard to the acceptance of the CEFR levels by test developers and test users will help to clarify the impact of the use of this proficiency framework in Taiwan.

The CEFR sets out to establish a common basis for discussion among language professionals, including teachers, curriculum designers, examining bodies, etc. (COE, 2001). Its three main aims are to establish a common language to facilitate discussion of learning objectives and language levels; to encourage reflection on educational practices with regard to learners' needs, learning objectives, and assessment; and to adopt common reference points in accordance with the COE's objectives in language education (North, 2006). In addition to its

role in supporting the development of the CEFR, the COE, through its Committee for Out of School Education and Language Policy Division, was instrumental in promoting the creation of the notional-functional syllabus, the Threshold/Waystage/Vantage proficiency levels, and the communicative approach to language learning (Morrow, 2004; Jones & Saville, 2009).

The origin of the CEFR was a 1991 Swiss initiative to achieve greater transparency and coherence in language education in Europe through encouraging discussion of language teaching and learning in support of international communication and mutual understanding, promote lifelong learning, and to

“develop a Common European Framework of reference for language learning at all levels, in order to: promote and facilitate co-operation among educational institutions in different countries; and provide a sound basis for the mutual recognition of language qualifications; and assist learners, teachers, course designers, examining bodies and educational administrators to situate and co-ordinate their efforts” (Little, 2007, pp. 5-6).

As Little (2007) points out, the CEFR is not language specific and consists of a framework accounting for the different elements of language use and language learning and a definition of six proficiency levels divided into three bands. Furthermore, it adopts an ‘action-oriented’ approach and views language users and learners as ‘social agents’ employing communicative competencies to carry out tasks (strategic actions in which competencies are employed to achieve goals) within a social context (COE, 2001). The framework includes general competences (including knowledge and skills), communicative competences, language use domains, and parameters that shape language use (Little, 2007).

The CEFR provides 54 separate scales for listening, reading, oral production, written production, spoken interaction, written interaction, note-taking, and processing text, as well as a global scale that summarizes proficiency at all six levels: Basic User (A1 and A2), Independent User (B1 and B2), and Proficient User (C1 and C2). Little (2007) notes that CEFR Common

Reference Levels confirmed existing conceptualization of proficiency levels but elaborated them to a degree that had never been achieved previously.

The pedagogic applications of the CEFR are seen in relation to conducting needs analysis of learners, increasing transparency regarding learning objectives, specifying action-orientated objectives for language use, recognizing partial competencies, and conducting learner self-assessment. The CEFR is also intended to be a reference for language assessment and used for specifying the content of examinations, stating criteria for attainment of learning objectives, and describing proficiency levels of exams in order to facilitate comparisons across different qualification systems (COE, 2001, p. 19).

Little (2007) acknowledges that the impact of the CEFR on language teaching and curriculum design has been somewhat limited, with the exception of curricula development projects in Spain and Ireland and teacher training efforts in Poland. The most visible impact of the CEFR has been on language testing. ALTE members swiftly moved to associate their exams with the CEFR six-level scale following its publication. The means by which examination providers could accomplish this goal were not straightforward, however, and this led the COE in 2002 to initiate a project to develop a manual for relating examinations to the CEFR (Little, p. 648). The manual for *Relating Language Examinations to the Common European Framework of References for Languages: Learning, Teaching, Assessment* (COE, 2009) is a guideline for developing, applying, and reporting procedures for situating examinations to the CEFR. This document, referred to as the *Manual*, suggests how to relate examinations to the CEFR, but it does not establish a standard of quality that testing bodies must meet in this process, nor does the COE officially recognize those examinations that have carried this out. These are just two of the problematic consequences of the recognition by Taiwan's MOE of English proficiency tests that

have related themselves to the CEFR. The Common Reference Levels were not intended to serve as authoritative prescriptions of how scores on proficiency tests should be compared to each other.

In addition to the points just mentioned, there has been a fair amount of criticism leveled at the CEFR by researchers. Alderson (2007), who “welcomed the publication of the first draft of the CEFR in 1996,” identifies a number of its limitations. He finds that the CEFR (a) assumes that the same proficiency level is required to perform the same communicative task in different languages without evidence, (b) is written in vague language and is inconsistent in its usage of terms, (c) provides insufficient information, and (d) its scales contain gaps. Alderson also notes problems with the use of the CEFR by exam providers, text publishers, curricula designers, and policy makers. Alignment of an exam with the CEFR is voluntary, and the COE offers no oversight of this process, despite having published the *Manual* as an aid to exam boards who wish to align their tests to the framework. Similarly, publishers and curriculum developers may use claims of CEFR alignment to market their products as they see fit, having no obligation to offer supporting evidence. Policy makers make use of the CEFR levels to define required language proficiency standards for graduation or immigration that may not be achievable or justified (Alderson, p. 662).

Weir (2005) considers the value of the CEFR for use in testing and also finds significant weaknesses in several areas. These include scales that are based on incomplete contextual variables, a failure to address cognitive processing, a lack of description of the quality of performance of language activities, and descriptors that are inconsistent and lack transparency. Weir makes the case that test developers require a framework that comprehensively, coherently, and transparently addresses facets of construct validity including context validity, theory-based

validity, and scoring validity (Weir, pp. 284-287). He cautions that the CEFR is a heuristic device only and should not be used prescriptively, nor should it serve as the basis for making claims about the comparability of different exams (Weir, p. 298).

Fulcher (2004) considers that the use of the CEFR as a tool in support of the COE's goal of harmonization in language policy in Europe is problematic and could lead to an oversimplification of validity issues that promotes an intuitive notional/functional taxonomy of language use to the status of a theory. Fulcher further sees danger in the institutionalization of common perspectives on language embodied in the CEFR for different stakeholders:

“teachers will be led to believe that the CEF scales reflect a hierarchy of second language acquisition; testers will be led to claim that their exams align with the CEFR to gain recognition within Europe and association with prestigious exams; and score users will be led to believe that scores on different exams are comparable or even equivalent based on the fact that the exams were aligned with the CEFR” (Fulcher, p. 260).

The final two of these three points are most significant with relation to the testing policy examined in this research. First of all, testing boards can gain recognition of their tests for measuring the English proficiency of civil servants in Taiwan without researching the suitability of those tests for the intended use. In this era of globalization, open access to markets is considered an essential principle from the perspective of test developers, but it offers little protection to test users who may lack the necessary sophistication to distinguish the relative advantages and disadvantages of different tests. Secondly, the use of the CEFR Common Reference Levels as proficiency benchmarks against which scores on different tests are measured encourages test users to equate scores on different tests to each other. Although the CEFR levels were adopted by the MOE as a practical solution to the problem of recognizing multiple tests for one use, a consequence of this decision is that test users and test-takers are led to believe that the differences in tests are insignificant. If a score on one test can be equated to a score on another,

the implication is that either test is equally suitable for use. Such a conclusion may appear reasonable, but it ultimately leads to inappropriate use of tests.

Fulcher concludes that irrespective of the CEFR's genesis as a means for promoting the political and social agenda of harmonization in Europe, its scaled-can do statements could be of great value for reporting the meaning of test scores to test users (Fulcher, p. 264). This point is significant because it acknowledges that the use of the CEFR levels may have positive value. This value is not, however, in regard to equating scores on different tests. Instead, language learners may use them to set learning goals, teachers may use them to develop curricula, and policy makers may use them to establish standards. The challenge is how to realize the potential benefits that the CEFR offers while avoiding the potential drawbacks.

Language testing policies in Asia and Taiwan. It is helpful to consider the issue of language testing policies in Taiwan from an Asian regional perspective. Many of the same goals behind the adoption of policies to encourage citizens and civil servants in Taiwan are recognized in other Asian countries. In this section of the literature review, problems associated with test-oriented education and their impact on English language testing policies and the effectiveness of measures intended to improve English proficiency in Asian countries are discussed. Research into the use of the CEFR in Taiwan is introduced in order to clarify the key issues affecting the use of SELPT as a promotion scoring criterion for civil servants.

Language testing has played an important role in the social and economic development of East Asia since the middle of the twentieth century. In the national education context, English testing has been integrated into college entrance exams for decades and more recently as a requirement for graduation. Internationally, English tests such as TOEFL and IELTS have been used as a qualification criterion for students seeking admission to universities in English-

speaking countries. Returned foreign students have had a major impact on the social, political, and economic development of East Asian nations, acting as a channel for the transfer of technical knowledge and social values from West to East, but also for their role in the broad discourse between Asia and the rest of the world.

Ross (2008) reviews language testing in the Asian context, and identifies key issues that have an influence on foreign language education and assessment in the region. The increasing importance of English as global language with a strategic competitive value is widely recognized, leading governments to initiate language policies designed to redress deficiencies in English language education. This had led to a conflict between policy makers in the government who seek to introduce centralized language education and testing standards and more conservative individuals within the education system and test-preparation industry that are unable or unwilling to implement reforms. Changes in language testing policies are also seen to produce negative washback on educational practices, and these put a focus on the construct validity of the different language exams available and the appropriateness of their use in various contexts. Finally, the use of foreign language tests as indicators of scholastic or professional merit is also problematic, calling into question the fairness of policies that perpetuate social stratification based on meritocracy.

Wu and Wu (2010) examine changes in English education and assessment policies in Taiwan that have taken place over the last decade, including the adoption of the Challenge 2008 National Development Plan, which prioritized the improvement of the English proficiency of college and university students, teachers, and civil servants, and the recognition of the CEFR common reference levels as proficiency level targets for English learners in Taiwan. The authors situate the adoption of the CEFR common reference levels within the social and political context

of Taiwan, noting the importance of the free-market, personal choice, and the government's reluctance to adopt a single English examination as a standard. The need for a method of equating scores on different examinations to a common standard led to the adoption of the CEFR proficiency levels by the Ministry of Education and the requirement that qualified examinations be aligned with the CEFR common reference levels. There have been problems associated with the use of the CEFR in Taiwan, particularly its role in comparing scores on exams that were designed to be used for different purposes and different populations of test-takers. Another concern is that the publication of a table for the purpose of equating scores on different tests might encourage test-takers to select the test that they feel would be the easiest to pass, regardless of the intended use of that particular test or its appropriateness. The CEFR's influence on the marketing of English examinations in Taiwan is also an issue, considering that exam developers may claim their tests are aligned with the CEFR's common reference levels without providing theoretical or empirical evidence to support that claim. The authors report on a project to relate GEPT reading tests to the CEFR using the *Manual for Relating Language Examinations to the CEFR (Manual)*. The results indicated that the reading texts used in the GEPT increase in complexity, abstractness, and lexical range as GEPT levels ascend. Problems with using the *Manual* were also noted, including a lack of a scale for cognitive features, insufficient precision and clear guidelines, the difficulty of use by teachers and other non-testing professionals, and very few descriptions of performance on tasks at different proficiency levels.

Test Use and Consequence

Language test use. To understand the impact that a testing policy may have, it is necessary to review what language tests are intended to accomplish. Bachman (1990) has identified two major uses of language tests, as (a) sources of information for decision-making

and (b) indicators of abilities for the purpose of evaluating an education program (defined as any situation in which teaching and learning take place, from individual classrooms to nation-wide programs). The first use is obviously the most significant in the context of the policy to recognize passing scores on English proficiency tests as a criterion for the promotion of civil servants in Taiwan. The promotion scoring points that accrue to a civil servant on the basis of English proficiency serve as information for those who award promotions.

Bachman provides a system for classifying tests based on five characteristics, and these are valuable for understanding how tests may both serve the needs of tests users and influence test-takers. These characteristics include (a) the purposes they are used for, (b) content, (c) frame of reference for score interpretation, (d) scoring procedure, and (e) testing methods.

In the education context, language tests are used for making decisions in regard to (a) selection, entrance, and readiness, (b) placement and diagnosis, and (c) progress, achievement, attainment, or mastery. In the context of this study, the first of these decisions is paramount, since civil servants are selected for promotion on the basis of a number of criteria, including English proficiency. As for content, language tests can be classified as based on a theory of proficiency or achievement within a specific domain of content. All of the English tests recognized by the MOE as counting toward promotion scoring are proficiency tests. The frame of reference for language tests is related to how tests scores are interpreted, with regard to the performance of a group of individuals, i.e. norm-referenced, or a specific level of ability, i.e. criterion-referenced. While all of the tests recognized by the MOE claim to have been aligned with the descriptors contained within the CEFR's Common Reference Levels, those level descriptors are not in fact identical to the criteria on which test-takers' performances on tests are assessed; thus, score interpretations based on this alignment must be regarded with some degree

of skepticism in the absence of sufficient evidence. Scoring procedure is significant in that it classifies tests as to whether they are objective, meaning the correctness of a test-taker's response can be determined without requiring scorers to make a judgment, and subjective, in which the scorer must make a judgment about the correctness of a response based on a subjective interpretation of the scoring criteria. Finally, testing methods vary widely, including performance tests in which test-taker performance is expected to be identical to performance in non-testing contexts, and methods such as multiple-choice, completion, dictation, and cloze, which combine different types of instructions, input, and tasks. These five test characteristics suggest that although scores on different tests can be related to the CEFR's proficiency levels, it is problematic to compare scores on these different tests to each other. The validity of such comparisons is addressed in the following section.

Consequential validity. Recognizing that not all tests are equally suited to their uses, test validity has been a subject of interest for as long as tests have been used. Discussions of test validity necessarily incorporate definitions of validity, the process of validation, and the type of evidence that is necessary to make assertions about the degree of validity of interpretations of test scores. The issue of consequential validity is important for this research since the recognition of different tests as measurements of English proficiency in Taiwan is based on the assumption that the use of such tests will result in positive, intended consequences. Whether the use of these English language proficiency tests produces the desired outcome is an open question that can be addressed through investigating the consequential validity of these tests and reviewing the evidence put forward to support such claims. In the following section, literature related to the validation of test use is reviewed.

There has long been an interest in the validity of language proficiency tests. The writing of Messick (1989) on validity has perhaps been the most influential for contemporary scholars, and his framework that identified the facets of validity proposes that validity is a unitary construct with a four-way classification incorporating two facets, the source of justification, and the function of the outcome of testing, has gained wide acceptance. Messick characterizes test validity as “an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores” (Messick, p. 13). Bachman (1990) recognizes a link between validity and the inferences that are based on test scores, and as a result understands validation to be a process of collecting evidence that supports the relationship between test score and interpretation or score use. This notion is fundamental to his view of the evidential basis of validity. On this basis, he proposes that what had formerly been considered different types of validity (content, criterion, and construct), could be more clearly understood as different types of evidence that must be collected in order to demonstrate validity. Bachman discusses the importance of consequential or ethical bases of validity, moving language testing beyond the psychometric properties of tests to consider their function as serving the needs of the education system or the society, agreeing with Cronbach that ability testing is intended to impartially determine “who gets what” (Cronbach, 1984). Recognizing the social implications of language testing, Bachman argues that the rights of test-takers, and the balance between these rights and the values of the society “go to the very heart of the ethical and political values of any society,” and that these have implications for testers and test users that require a consideration of the consequences of testing (Bachman, 1990, p. 280). Bachman proposes that it is the responsibility of testers and test users to provide evidence that the tests in use “are valid indicators of the abilities of interest, and that these

abilities are appropriate to the intended use, and then to insist that this evidence be used in the determination of test use" (Bachman, 1990 p. 285).

These points can be interpreted as suggesting that the use of English proficiency tests within Taiwan's civil service system should undergo a process of validation in order to obtain evidence of the suitability of these tests for this use. This process should be conducted both by testers, the agencies responsible for developing the tests that are being used, and the test users, in this case, Taiwan's central government. Each group has an expertise that the other lacks; testers understand the psychometric features of their tests, while test users understand the political implications of decisions that are made based on test scores.

Weir (2005) accepts the notion of validity as super-ordinate in that it is multi-faceted with complimentary aspects and offers a definition of validation as the collection of quantitative and qualitative evidence to support inferences from test scores. It is the responsibility of testers, he asserts, to present a "clear argument of a test's validity in measuring a particular trait with credible evidence to support the plausibility of an interpretive argument" (Weir, p. 1). Weir explains that consequential validity can be understood as incorporating three factors, (a) those of differential validity, i.e. comparing validity group vs. group, (b) washback, the impact of testing on teaching and learning, and (c) effect on society (Weir, p. 210). He considers that consequential validity is demonstrated through posteriori evidence of washback and effect on society, and proposes that this can be collected through questionnaires and interviews with stakeholders. These methods are employed in the current study to collect evidence of test washback and the effect of tests on society. For this study, society is conceptualized as the context of the workplace in which the civil servants participating in this research labor.

Chalhoub-Deville (2009) addresses the responsibility of test developers and test users to consider the relationship between validity and impact in her paper on the No Child Left Behind policy in the U.S. She identifies the divide between Cronbach and Messick on the relationship between impact and validity, noting that the former preferred to separate impact from validity, while the latter incorporated consequential validity into his unified validity model. For Chalhoub-Deville, the negative unintended consequences of test use are the key issue in terms of consequential validity, and that such impact normally arises from the use of test scores for purposes for which they were not intended (Chalhoub-Deville, 2009, p. 122). She recognizes that test users are often incapable of carrying out validation research, and that although they bear the responsibility for ensuring that the uses of tests are appropriate, they may not have access to the expertise that would enable them to carry out the necessary research that would allow them to do so. Ideally, testers and test users would cooperate to accumulate evidence of appropriate test use, but practicalities may intervene and influence both groups to assume that the other has gathered sufficient evidence. This discussion could be relevant to the use of scores on English proficiency tests as a promotion criterion for civil servants in Taiwan. The government may assume that the test developers have researched this type of use of test scores, and test developers may assume that test users have conducted research into the use of tests for this purpose. The result may be that neither party has addressed the consequences of this specific use. If that were the case, there would be a distinct lack of evidence of the impact of test use, with a potential for negative unintended consequences.

Assessment use arguments. One of the more recent proposals for establishing the validity of the use of language tests is Bachman's assessment use argument. This method outlines a procedure that can be used to construct a convincing logical argument for the purpose

of validating test use. For this purpose, it is necessary to collect information about the context of test use, the identity of test-takers, and the potential consequences of test use.

Recognizing that most tests are used to provide information for making decisions, Bachman (2005) proposes assessment use arguments as a method for structuring test use arguments that link test performance with interpretations, and interpretations with test uses. Bachman (2005, p. 9) adopts the Toulmin argument structure, in which claims are made on the basis of data supported by warrants with backing, and challenged by rebuttals, as a model for his assessment use argument. Bachman's assessment use argument, including both a utilization argument and a validity argument, forms the basis of a validation process that involves building a case that decisions are defensible and providing credible evidence to support that case. Bachman proposes four types of warrants for a utilization argument: (a) relevance (score interpretation is relevant to the decision to be made); (b) utility (interpretation is useful for making the decision); (c) intended consequences (the consequences of using the assessment to make decisions will be beneficial to stakeholders); (d) and sufficiency (the assessment provides enough information to make a decision). Bachman suggests that qualities of tests, including test usefulness, ethics and validity, and fairness, which are seen as critical to test design, development and use, could act as warrants or rebuttals in assessment validity and assessment utilization arguments. He argues that such qualities pertain to specific assessments of individuals or groups of test-takers, and intended uses and potential consequences are local concerns that need to be addressed. The warrants in an assessment use argument help to determine what type and amount of evidence is needed to convince stakeholders that assessments suit their intended purpose. The construction of an assessment use argument would be an appropriate method for validating the use of English proficiency tests as a promotion scoring criterion for civil servants in Taiwan. Such a procedure

would involve gathering evidence to support warrants with regard to relevance, utility, intended consequence, and sufficiency.

Washback and impact. To understand the consequences of the English testing policy that is the subject of this research, it is helpful to review the literature on test washback and impact.

Hawkey (2006) discusses impact and washback from a number of perspectives, including definition and scope, positive or negative implications, intentionality, stakes and stakeholders, and relationships with validity and validation. Hawkey relies on Alderson and Wall's (1993) limitation of *washback* to include only influences on teaching and learning, referring to the wider influences of testing as *impact*. This distinction parallels Bachman and Palmer's (1996) characterization of washback as *micro* and impact as *macro*. Hawkey notes that washback is a subset of the more broadly defined *impact* (Hawkey, 2006, p. 7). Similarly, there is broad agreement that both washback and impact are themselves components of test consequences (Taylor, 2005). Washback is viewed as neutral by virtue of its positive or negative influence on teaching and learning (Alderson & Wall, 1993; Hawkey, 2006). Washback is positive when testing brings about changes that test developers or test users desire, and negative when it produces outcomes that are not desired. Washback that is desired is characterized as *intended*, and that which is not planned as *unintended*. Hawkey (2006) identifies test stakeholders as those with a direct or indirect interest in a testing program, including teachers, students, administrators, government agencies, the public, etc. The widening of stakeholders to include the public and those not directly involved with testing has led to a recognition of social expectations of accountability and professionalism in language testing, a development that is associated with the call for ethical language testing that takes account of the competing interests of different groups

of stakeholders. Hawkey (2006) argues for achieving a balance between the extremes of ignoring consequential validity on the one hand, and investigating so many effects on stakeholders that clear evidence of washback and impact would be impossible to observe, on the other.

Saville (2003) discusses the four qualities around which Cambridge ESOL examinations are developed: validity, reliability, impact, and practicality. It is postulated that an appropriate test development process balances the demands of these four qualities and attempts as much as possible to maximize the value of each without unduly sacrificing the value of the others. Placing excessive concern on reliability could weaken the ability to promote validity; and likewise, an overemphasis on validity could cause practicality to suffer. Saville defines impact as “the influence of the test on general educational processes and on the individuals who are affected by test results (Saville, p. 73), and in this regard would include test washback as well as the wider consequences of test use. Saville notes that Cambridge ESOL considers its examinations to be both fair and appropriate within the educational context in which they are used, and that their impact is intended to be positive. He argues that test developers can promote positive impact by providing the public with greater access to information about testing, as well as providing research and advisory services to test users. Saville (2005, p. 75) outlines questions that test developers can ask to collect information that will allow test impact to be estimated, including (a) Who are the test-takers? (b) Who are the test users, and what are the test results used for? (c) Who is teaching toward the examination, and under what circumstances? (d) What are the effects of the examination on public perceptions generally? (e) How is the examination viewed by those directly involved in education? and (f) How is the examination viewed by members of society outside education?

This present research into the testing policy is designed to collect evidence similar to that described by Saville. The questionnaire elicits data about the test-takers and their views of English tests; the document analysis investigates the use of test results; and the interviews with government officers collect information about how those involved in the use of the test results in the workplace (since this is not an education context) view the examinations and the policy that calls for their use.

Testing and Motivation for Learning

One of the main goals of the testing policy is to encourage civil servants to develop their English proficiency. The ability to communicate in English is valued by the government because it is seen as having an impact on global competitiveness. Typically, language tests are administered in order to collect information that can be used to make decisions. In this case, the decision is whether to award civil servants promotions on the basis of the promotion scoring points that they accumulate. In the context of the testing policy, the test functions as an indirect indicator of a civil servant's merit for promotion. Therefore, while an English test's main purpose would seem to be simply measuring English proficiency, under this testing policy it has an impact on the methods that a civil servant may employ to learn English as well as on motivation to learn. To explore these ideas, the relationship between learning, motivation, and testing will be examined.

Lifelong learning and adult education. One goal of this study is to determine whether the testing policy has achieved its aim of improving the English proficiency of civil servants. Such an improvement in English proficiency would be a result of learning. To understand how adults learn, theories of lifelong learning and adult education are reviewed, noting the definitions of these terms, the contexts in which such learning typically takes place, and suggestions for

optimizing such learning to facilitate not only the development of language skills, but also to promote personal growth and increase access to opportunity.

The term *lifelong learning* came into common usage in the 1960s, and was incorporated as a master concept by UNESCO in 1970 (Medel-Añonuevo, Ohsako, and Mauch, 2001). The policy to encourage civil servants in Taiwan to improve their English ability may be understood in relation to the concept of lifelong learning, which Cropley (1980) defines as learning that lasts the whole life of the individual; leads to acquisition, renewal, upgrading and completion of knowledge, skills, and attitudes in response to the changing conditions of modern life and promoting self-fulfillment; dependent on increasing ability and motivation to participate in self-directed learning activities; and acknowledges the contribution of all available formal, non-formal, and informal educational contexts.

Hughes and Tight (1995), recognizing the importance of learning for the development of individual, organizations, and societies, see lifelong learning taking place within the context of learning organizations and learning societies. Tight (1998) finds that lifelong learning may be understood as a necessity because it speaks of change, is linked to the needs of the economy and the society in today's context of globalization and competitiveness, emphasizes the role of the individual in taking charge of this education, and has the eventual goal of self-fulfillment. Tight asserts that lifelong learning has become a compulsion. In this view, the rhetoric in support of lifelong learning may seem enabling and inclusive, yet the reality is that individuals effectively have no alternative but to participate, and to bear the cost of doing so themselves. He finds that lifelong learning has become an extension of work and "the threat of economic and social exclusion hovers over those who do not take on this responsibility" (Tight, p. 256).

If we consider the policy to encourage civil servants to improve their English from the perspectives outlined above, we may see similarities. The need to improve Taiwan's English proficiency is framed within the context of Taiwan's national development, international competitiveness, and globalization in the goals of the E-Generation Manpower Cultivation Plan, and more specifically in the Government Employee Life-long Learning Promotion Project outlined in Challenge 2008. In that document, lifelong learning is characterized as "a goal we have to strive for," and one for which the government should commit resources in order to "quickly enhance citizen quality and elevate the nation's competitive advantage to international standards" (MOE, 2005). The language of necessity and of a desire for rapid change evident in these passages suggests that national aims will take priority over individual interests in pursuit of a common good, and it may be assumed that this could impact motivation to learn.

Because lifelong learning is focused on encouraging and facilitating the acquisition of knowledge for individuals who have finished their formal education and entered the workforce, it is important to understand how adults learn. The theory of andragogy (Knowles, 1980) is based on assumptions about how adult learners acquire new knowledge throughout their life. In contrast to pedagogy, childhood learning, andragogy assumes that adults are more self-directed in their learning, rely on experiential learning as a basis for new learning, have motivation that is linked to social relevance, and are interested in learning with immediate application. Building on the theory of andragogy, Boggs (1981) supports adult education that contributes to personal growth and the ability to adapt to changes throughout life. He proposes that adult learning should promote skill development and creativity, facilitate progress toward personal goals, and provide access toward greater opportunity.

These ideas are relevant to language testing in that the use of language tests as a tool to encourage adults to improve their English ability would tend to have a greater influence on learning if the content, format, and testing methods were more closely aligned with adults' learning preferences and could motivate them to invest the time and energy necessary. The rationale behind the policy to recognize English proficiency as a promotion criterion for civil servants would tend to support some of the features of andragogy, such as promoting the development of English language skills that may be applied in the workplace, and also providing greater access to opportunity through qualifying for higher status positions that may require English or offer more responsibility. On the other hand, a testing policy would not be seen as promoting creativity, since most tests test-takers tend to adopt test preparation strategies that apply methods such as rote memorization in order to minimize errors. The potential influence of a testing policy on adult learning would depend greatly on factors related to the test-takers' personal goals, their belief about the social relevance of improving their English, and the immediate application of the skills and knowledge that they acquired through their learning prior to taking the test.

Motivation: Learning and testing. Several of the items on the questionnaire used in this study query the participants on the influence of the testing policy on their motivation to learn. Motivation is recognized as essential to learning, and it is hypothesized that the effectiveness of the testing policy could be related to its ability to increase civil servants' motivation to learn English. This review will discuss key theories about the nature of motivation and its influence on learning behaviors and outcomes. Testing is believed to have an impact on motivation to learn, and the results of studies into this relationship could help to explain how the testing policy affects learners.

The discussion of the features of motivation associated with adult learning necessitates a definition of motivation, not least of all because it is a term that is used frequently but about which there are multiple understandings of its exact meaning. In support of his Socio-Educational Model (SEM) of motivation, Gardner (1985, p. 10) defines motivation as “the extent to which the individual works or strives to learn the language because of a desire to do so and the satisfaction experienced in this activity.” This conceptualization is based on the assumption that effort must be expended in order to achieve the goal, and that personal desire and the satisfaction associated with learning the language influence the amount of effort directed at reaching the goal.

Entwistle (1987) reviews ‘mechanistic’ theories of motivation in which motivation was categorized as ‘intrinsic,’ related to the relevance of and interest in learning, and ‘extrinsic,’ associated with the use of external incentives for learning, and discusses their relevance with approaches to learning: (a) deep learning (the intention to reach a deep understanding), (b) surface learning (the intention merely to satisfy task or course requirements that are viewed as external impositions), and (c) strategic (the use of surface and deep learning strategies to maximize attainment). Deep learning is associated with the intrinsic motivation tied to interest in and relevance of learning, while surface learning is linked with anxiety, fear of failure, and extrinsic incentives. Among Taiwan learners, surface learning is seen in the context of test preparation strategies that students employ to deal with tests in subjects in which they have little intrinsic interest but are required to study to fulfill requirements. In Taiwan, students prepare for English proficiency tests that employ test methods based on multiple-choice questions by using memorization and other techniques that result in shallow learning.

Expanding on the SEM, Tremblay and Gardner (1995) differentiate between *motivational behavior* and *motivational antecedents*, the former including characteristics of

individuals that are apparent to observers, and the latter composed of the variables that influence motivational behavior. Through their study on the language attitudes of L2 learners of French in Canada, goal salience, valence, and self-efficacy are identified as motivational antecedents that influence motivational behavior. Tremblay and Gardner note that their conclusions are supported by goal theory, which postulates that goals influence performance by increasing the intensity, persistence, and attention of effort directed toward a goal. Similarly, higher levels of valence, the value that one ascribes to the language and learning the language, and self-efficacy, the belief that one has the ability to reach a future goal, have also been associated with greater motivation in learners (Bandura, 1991, 1999). In the context of the testing policy, valence could have a significant impact on the amount of motivational behavior that English learners among Taiwanese civil servants are likely to express. Higher levels of intrinsic motivation to learn English would likely be associated with greater valence, and thus result in greater effort directed toward improving English proficiency.

Confidence is also seen to be an influence on motivation, related both to expectancy, the expectation that effort will lead to success, and self-efficacy. Clement, Gardner, and Smythe (1980) suggest that the higher the expectation that a behavior will produce a specific outcome, the greater the motivation to do the activity; and the more one believes he or she has the ability necessary to achieve an outcome, the more likely it is to be achieved. These theories suggest that the use of tests could increase motivation if achievement of a particular score (or associated CEFR proficiency level) is adopted as a specific goal. The value that a civil servant attaches to learning English may also predict motivation level for studying, with a higher value being associated with higher motivation to learn English. Self-efficacy, or expectation that one has the capacity to achieve a goal, is also relevant to testing in that test-takers with more experience of

having taken an English proficiency test or more past success in taking such tests, will be more likely to exhibit motivational behavior. Conversely, learners with less confidence in their ability to pass an English test, lower English proficiency, and fewer experiences of success on an English test will be less likely to be motivated to improve their English proficiency to a level that would enable them to pass a test as a criterion for promotion. These hypotheses would support educating civil servants about the methods, content, and format of English proficiency tests in order to improve their confidence in their ability to perform well on the tests that they are encouraged to take in order to earn promotion scoring points.

Motivation to learn is also associated with Deci and Ryan's self-determination theory (1985), in which motivation is seen as linked to three basic psychological needs. These needs are competence, relatedness, and autonomy, and each is related to features of both lifelong learning and andragogy. Competence, or effectiveness in dealing with an environment, suggests that adults are motivated to learn new knowledge and skills that allow them to adapt to a changing environment over a lifetime. Relatedness, or interacting with others, will motivate adults to engage in learning if the knowledge or skills acquired are socially relevant and have practical application. As with lifelong learning and andragogy, self-determination theory addresses adults' need for autonomy, the urge to be an active agent in development and change, proposing that adults are more motivated when learning is self-directed, leads to the achievement of personal goals, and results in greater access to opportunity. In the context of the testing policy, the needs for competence, relatedness, and autonomy influence the motivation of civil servants to improve their English. Testing could motivate learners to continue learning in order to add to their existing competence levels, and this effect would be particularly strong if there is an actual need for English, i.e. if English is related to their job. The significance of autonomy could be

addressed in terms of the civil servants' active participation in self-directed learning that is understood to increase access to opportunity. From this perspective, incorporating English proficiency as a promotion scoring criteria that is voluntary and based on a civil servant's actual need for English in the workplace could produce increased levels of motivation to learn.

While the discussions above have discussed learning motivation generally, there has also been research conducted into the impact of high-stakes testing on students' motivation. Harlen and Crick (2003) note that high-stakes tests are commonly adopted by policy-makers in order to raise academic standards. Kellaghan, Madaus, and Raczek (1996) identify six propositions about the benefits of summative testing for increasing standards of educational attainment: (a) test and examinations indicate standards to learners; (b) high standards can be demanded; (c) tests demonstrate to students what must be learned; (d) rewards and penalties can be applied to test-takers to get results; (e) learners will put more effort into studying to pass tests; and (f) this will be the case for all students. Increases commonly seen in test scores following the introduction of tests are cited as evidence of the positive impact of testing on standards. However, according to Harlen and Crick (2003), these increases are likely to be the result of the greater familiarity of students and teachers with test content and format, and not necessarily associated with an increase in learning.

In a review of research into the effects of testing on learning, Kellaghan et al. (1996) find that motivation is a complex phenomenon, and that too little attention is paid to learners' goal orientation, learner autonomy, and self-efficacy when seeing explanations for learner motivation. Harlen and Crick (2003) conducted a meta-analysis of studies on the impact of testing on students and found that high-stakes testing tended to reduce both learner autonomy and intrinsic motivation and led to shallow learning. Extrinsic rewards were also found to lead to shallow

learning, and motivation was found to decrease when incentives were withdrawn. Sloane and Kelly (2003) note that high-stakes test produce both positive and negative effects in learners, and suggest that test anxiety could be due to insufficient preparation on the part of test-takers rather than some feature of the test itself. They caution that understanding the effect of high-stakes tests on learners requires understanding the interplay of variables related to the social context of learning and testing, and that “teachers and policy makers [should] not blame the thermometer for the fever” (Sloane & Kelly, p. 14).

This discussion of the influence of motivation on learning, and of the relationship between testing and motivation to learn, suggests a complex interrelationship among these different factors. Motivation can be understood as intrinsic or extrinsic, may lead to deep learning or shallow learning, and is related to learners' notions of competency, relatedness, and autonomy. Testing does have an influence on motivation to learn, and this is mediated by factors that are related to learners as individuals and to the context in which learning takes place. Lifelong learning is a concept that is mainly applied to adult education, and if civil servants have a personal interest in the subject, feel that the result of learning can increase access to opportunity, and view the learning as an autonomous process, it is likely that a test policy that promotes lifelong learning as a goal will increase motivational behavior. Relying solely on extrinsic incentives in the absence of sufficient goal salience or valence is unlikely to produce deep learning that will lead to improved English proficiency among civil servants influenced by the testing policy.

Motivation in English learning and testing in Taiwan. While motivation is a psychological construct that helps to explain a learner's drive to develop skills and acquire knowledge, it is likely that motivation is also influenced by the surrounding social and cultural

context. To establish a link between general theories of learning motivation and the conditions faced by English learners in Taiwan, scholars have investigated the influence of extrinsic incentives on motivational behavior. Such research could have significant implications for interpreting the effect of the government's testing policy on the English proficiency of civil servants.

Much of the literature on learner motivation is based on research carried out on second language (L2) learners in western societies, and it has been suggested that results of such studies may fail to accurately address the motivation of English as a foreign language (EFL) learners in Asian societies. In Taiwan, where few English learners have the opportunity to use English in their daily life, integrative motivation (Gardner & Lambert, 1972) is unlikely to have a strong influence on studying behavior. Instrumental motivation, referring to the motivation associated with external incentives such as access to educational or employment opportunity, such as a pay raise or a promotion, has also been shown to have an effect on motivational behavior (Gardner & McIntyre, 1991) of L2 learners, and it is likely that it has a similar effect on foreign language learners. Chen, Warden, and Chang (2005) investigated motivation among English learners in Taiwan, and sought to understand extrinsic motivation associated with instrumental motivation and its influence on learning. The authors adopt the process model of second language motivation (Dornyei and Otto, 1998), which consists of three phases: (a) preactional phase (choice motivation preceding action), (b) actional phase (motivation influencing level of effort applied to learning), and (c) postactional phase (evaluated retrospection after action). The results of their study on 567 adult English learners found that preactional factors, including passing job exams, obtaining higher job security, or getting a raise, were strongly associated with motivational behavior in the study participants. The authors find Norton Pierce's (1995) concept

of 'investment' to be relevant to their results, and hypothesize that the amount of effort that learners invest in study is related to the anticipated return on investment, in other words, the value that they expect to obtain through achieving their educational goals (Chen, Warden & Chang, 2005). The authors postulate the existence of a motivator, the *Chinese imperative*, which is associated with internalized requirements related to the socio-cultural context in which Chinese learners operate, reflecting both test-oriented instruction and family values among other characteristics. Government policies were not considered as a potential motivator in the study, but the potential impact of national development plans and efforts to increase awareness and acceptance of English is acknowledged (Chen, Warden & Chang, p. 624).

The current study does not attempt to prove or disprove the existence of the Chinese imperative motivation, but does, however, find value in an approach that seeks to understand the influence of the specific contextual features that are present in Taiwan. The distinction between instrumental and integrative motivation is useful for discussing the influence of goals on the investment that learners are willing to make in improving their English. The promotion scoring points associated with certifying English proficiency via taking an English test are a tangible reflection of the value associated with a learning goal. It is likely that the value of the promotion scoring points is a powerful incentive that strongly influences the motivation of civil servants to improve their English proficiency.

Conclusion

This literature review has examined a wide range of issues that are relevant to the English testing policy. The history of testing in Asia, and particularly the use of competitive examinations for entry into government, is believed to have a significant impact on the use of tests to encourage civil servants to improve their English proficiency. The subject of language

planning and policy was introduced, reviewing a methodology based on a policy analysis framework and examining issues related to the adoption of the CEFR Common Reference Levels as a language proficiency yardstick by Taiwan's central government. Theories of test validity, particularly consequential validity associated with test impact, were examined, and the potential benefits of assessment use arguments were explored. In the final section, the English testing policy was situated within a context of lifelong learning and theories of adult education in order to hypothesize about the impact of testing on learning motivation within an English as a foreign language (EFL) environment.



Chapter 4

Methodology

The research methods employed in this study were designed to gather a variety of data from a wide range of sources and subject these to analysis in order to answer the research questions. The first research question will henceforth be referred to as RQ1; the second, RQ2; and the third, RQ3.

RQ1: What are the impacts of the policy encouraging civil servants to pass an English language proficiency exam in order to qualify for promotion?

RQ2: Does this policy achieve the aims for which it was created?

RQ3: What factors have a bearing on the achievement of the SELPT policy aims?

There were four main sources of the data gathered for this study: (a) the results of questionnaire distributed to employees of three government agencies; (b) online materials from government agencies in Taiwan; (c) interviews with government officials in a number of different government agencies; and (d) official documents related to the English testing policy.

The questionnaire results were primarily used to assess the impact of the testing policy on civil servants, the focus of RQ1. The questionnaire collected information on the participants' demographics, English proficiency level and need for English at work, English testing and studying experience, impressions of English proficiency tests, and evaluation of the effectiveness of the testing policy. The majority of questionnaire items collected quantitative data via Likert scales, but several items asked participants to provide open-ended comments.

Publicly-available materials from the government were accessed through the Internet. These provided information about the structure of Taiwan's government and the civil service system, the roles and responsibilities of various agencies, the national development plans, and

measures that were specifically related to the English testing policy. This information was essential for understanding the background and context of the testing policy, the policy's initial goals, the government's public statements with regard to the achievement of the policy's objectives, and the subsequent development of further aims. Such data contributed to answering RQ2 and RQ3.

Speaking directly with officials involved at different stages within the process of developing and implementing the testing policy gave key insight into the nature of the policy cycle. The primary objective of the interviews was to gather data that would contribute to answering RQ2 and RQ3. Interview questions were drafted prior to the interviews and provided to the subjects so that they could prepare their answers and supplemental data. The results of the interviews differed according to both the willingness of the participants to address the questions and the nature of further information that the participants voluntarily provided.

The interviews with officials enabled the researcher to request access to documents that were relevant to the testing policy. The various documents that were made available provided details about the policy's specific objectives, the measures that were used to achieve them, and the results that were reported. The data contained within these documents contributed to answering RQ2 and RQ3.

This thesis research may be characterized as both a study of test impact and an analysis of language testing policy. Three research methods were employed in this study. The first was using a questionnaire to survey employees of three government agencies with regard to the impact of the English proficiency tests. The second was conducting interviews with representatives of seven government agencies involved in making and implementing the English

testing policy. The third was utilizing the G-S extended policy analysis framework to integrate the information gathered from the interviews and both the online and government documents.

Questionnaire

The use of a questionnaire in this research necessitated the development of a survey instrument, the handling of the completed questionnaire, and analysis of the results.

Stages of questionnaire research. The first stage of the use of the questionnaire was the development of the survey instrument, and this included drafting, translation, pilot testing and revision. The second stage involved the distribution and collection of the physical copies of the questionnaire, as well as the coding of the data into an electronic spreadsheet document. In the final stage, analysis of the resulting data was carried out.

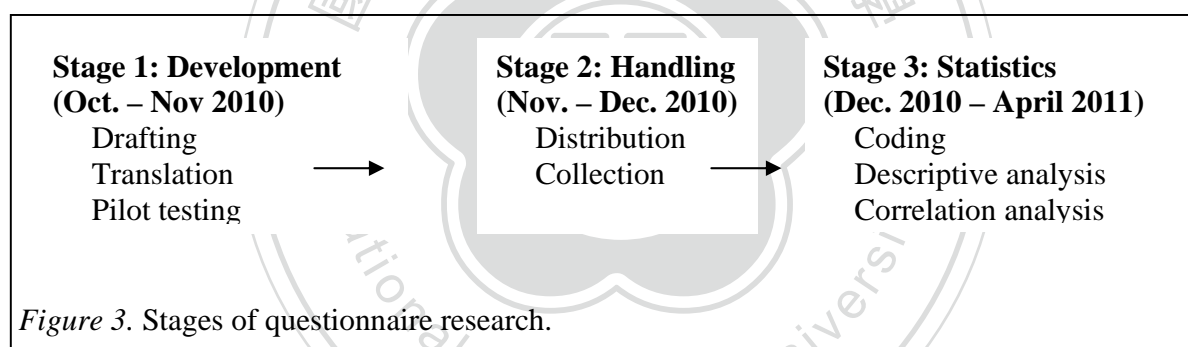


Figure 3. Stages of questionnaire research.

Development. Test impact studies have conventionally made use of stakeholder questionnaires to assess the influence of tests. As this research aimed to understand the consequences of a testing policy for civil servants who were directly affected by its implementation, a questionnaire was adopted as a suitable approach to gathering standardized data. The content of the questionnaire was designed to gather demographic information about the participants, particularly with regard to their employment, as well as information about their experience taking English proficiency tests, their study of English, impressions of English tests,

and their evaluation of the effectiveness of the testing policy. While earlier test impact studies were used as references, an entirely original instrument was designed in order to most ideally suit the focus of the research questions as well as the unique context in Taiwan.

The drafting of the questionnaire went through numerous stages in order to increase the likelihood that it would yield the desired data. The earliest drafts of the questionnaire were completed in English, and these were reviewed by both Taiwanese scholars and policy stakeholders, all of whom offered valuable comments with regard to revisions of the content and structure. The questionnaire was then translated into traditional Chinese (the writing system used in Taiwan) by a native speaker of that language. Several cycles of review and revision with the aim of clarifying the language used in the items and refining the format were carried out prior to the thesis proposal presentation to the thesis advising committee on October 27, 2010.

A draft of the questionnaire was used in a pilot survey on November 3, 2010. The participants in the pilot survey were fifteen employees of the CBC that attend a weekly English course taught by the researcher at the bank's headquarters. The researcher first introduced the topic of the study and then explained that the results would be used in a master's thesis. After the participants verbally gave their consent to take part, they completed the questionnaire during their class meeting. It took the participants between ten and fifteen minutes to complete the survey, which was the length of time targeted by the researcher. Thus, it was found that the length of the questionnaire was appropriate. One questionnaire was returned without any responses. The participants in the pilot provided oral feedback on the questionnaire and commented on the wording of several of the items, specifically suggesting that age should be reported as a range rather than a specific year. Following the piloting procedure, final revisions to the questionnaire were made.

Handling. Copies of the questionnaire were distributed to employees of the Central Bank of China (Taiwan), referred to as the CBC, the Ministry of Finance (MOF), and the Ministry of Economic Affairs (MOEA). These agencies were chosen for several reasons: 1) all are central government agencies; 2) all are involved in financial and economic matters in which foreign language proficiency is assumed to be of high value; 3) all employ a relatively large number of employees; and 4) the researcher had access to employees of these agencies through professional networks. A narrower target population was chosen rather than a wider one in the hope that deeper insight could be revealed into the impact of the testing policy on employees of agencies with similar areas of responsibility. Due to the size of these agencies, it was expected that the employees of these three agencies and their subordinate organizations would be engaged in a wide variety of jobs, and that the impact of the policy would affect them differently based on their unique employment conditions, education background, age, and gender.

The questionnaire was distributed in three stages: first to employees of the CBC, then to those working at the MOF, and finally to staff of the MOEA. The researcher had a goal of gathering a total of 300 completed questionnaires, with approximately 100 coming from each of the three agencies. Representatives of the different agencies were contacted by the research, and they consented to distribute the questionnaire among their colleagues and take responsibility for returning them to the researcher. This stage of the research took place in November and December 2010. Due to the manner of distribution of the questionnaire, it was not possible to monitor what percentage of participants chose not to complete the questionnaire. The results of all copies of the questionnaire that were completely or partially completed were included in this study. Items that lacked responses were treated as missing items and excluded from statistical analysis.

Statistics. The data collected from the completed questionnaires were transferred to a Microsoft Office Excel 2003 spreadsheet. The 17.0 version of the Statistical Package for Social Science (SPSS) software was used to prepare descriptive statistics and conduct statistical analysis. Descriptive statistics are used to demonstrate the trends in the results. Correlation analysis was conducted to identify correlation between individual independent and dependent variables. Graphs and tables are used to present key findings where appropriate.

Content. The questionnaire includes 44 items that are divided into four sections: demographics, English, test impression, and policy effectiveness. The Chinese and English versions of the questionnaire can be found in Appendix C and D, respectively.

Demographics. In this section, participants provided information about their employment, education, age, and gender. With regard to employment, the questionnaire asked for the name of the agency the participants works for (Item A), their civil service grade (Item B), department (Item C), job title (Item D), and number of years of service (Item E).

Civil servants in Taiwan are divided into three ranks and fourteen grades. Unlike the MOF and MOEA, the CBC is an independent agency and maintains a separate but similar grade system for its employees. It is possible to equate the levels in the different system using an official table created for that purpose. To facilitate analysis of the participants' responses to the questionnaire, the researcher divided the civil service grades into five levels: Elementary A (Levels 1 through 3), Elementary B (Levels 4 and 5), Junior A (Levels 6 and 7), Junior B (Levels 8 and 9), and Senior (Levels 10 through 14). The use of five levels was chosen by the researcher to more clearly differentiate the participants according to their seniority within the civil service system.

The number of departments (or subordinate agencies) that the participants work within was extensive, numbering over 22 for the Ministry of Economic Affairs alone, so a system was devised to divide participants within the same agency into groups of approximately the same size based on the nature of their work. Departments with study participants numbering approximately 20 or above were not altered, while departments with fewer participants were combined to produce groups with at least twenty members. Ten such groups were ultimately created, and these will be discussed in Chapter 5.

The *Job title* item yielded data that was difficult to manage due and was ultimately not used in the analysis. A variety of job titles are used in the different agencies, and with a few exceptions, it is not possible to link titles in one agency with those in another. Further research on this topic may wish to adopt an approach that divides jobs into supervisory and non-supervisory. It was discovered during the analysis of the questionnaire that a similar approach is used by the CPA when it reports statistics regarding classes of civil servants.

Participants indicated the length of their employment with the government by writing the number of years they had served as civil servants in a blank provided. The information gathered about the participants' civil service grade and their length of service made it possible to group the participants in terms of their seniority and levels of responsibility. It was hypothesized that these two factors might play a role influencing the participants' evaluation of the policy's impact.

The questionnaire asked participants to indicate the highest level of education (Item F) that they had achieved by choosing from *college graduate*, *university undergraduate*, *university graduate level*, and *university post-graduate level*, as well as their major field of study (Item G). Five categories of major subjects were provided, *arts and literature*, *law/administration/sociology*, *finance/economics*, *physical science*, and *other*. The questionnaire

instructed those selecting other to indicate their specific major by writing that information on a line provided for that purpose.

Finally, data about the participants' gender (Item H) and age (Item I) were collected. The results of the pilot test suggested that participants would be less willing to indicate their actual age, so seven 5-year categories from 25 to 60 (plus one group for those over 60 years of age) were provided for the participants to choose from.

English language. In this section, data was gathered from the participants about their English ability, the importance of English in their workplace, their experiences of taking an English language proficiency test, and their past and current study of English.

Self-assessment of English ability. The participants were asked to provide a self-assessment of their English ability by completing one 5-point Likert scale (Items 1-4) for each of the four language skills (listening, speaking, reading, and writing). The five choices were labeled *very poor*, *poor*, *not bad*, *good*, and *very good*. Each choice was assigned a value from 1 (*very poor*) to 5 (*very good*). These results are used in two different analyses. First, the results from the 5-point scale are transferred to a three-point scale (*above average*, *average*, and *below average*) to facilitate discussion and comparison with similarly constructed Importance of English data. *Above average* includes *good* and *very good* responses; *average* will include only *not bad* responses; *below average* will include *very poor* and *poor* responses. Then the sum of each participant's scores for the four language skills is divided by four to produce a mean value that is reported as the participant's overall *self-assessment average*.

Importance of English for work. The questionnaire also asked participants to indicate how important English proficiency is for their job by completing one 5-point Likert scale (Items 5-8) for each of the four language skills. The five choices were labeled *no importance*, *not very*

important, no opinion, important, and very important. Each choice is assigned a value from 1 (*no importance*) to 5 (*very important*). Like the self-assessment of English data, these results are used in two different analyses. First, the 5-scale results are transferred to a 3-point (*above average, average, below average*) scale for discussion and comparison with the self-assessment data. Then the sum of each participant's scores for the four language skills is divided by four to produce a mean value that is then reported as the participant's overall *importance of English average*.

The participants were further asked to provide qualitative comments on the overall importance of English for their job. These comments were collected in their raw form for qualitative analysis and converted into quantitative data by assigning nominal values to seven typical terms expressed in the responses: *reading, translating, writing reports, writing letters/email, attending international conferences, escorting foreign visitors, and using the telephone*. A maximum of three responses per participant were collected.

Testing experience. The questionnaire asked participants to provide information about the English proficiency tests (Item 9) that they have taken, including the tests' names, the reason they were taken, and the scores that were awarded. Participants could select from eleven separate tests, TOEFL iBT, TOEF CBT, TOEFL PBT, IELTS, TOEIC, BULATS, FLPT, GEPT Elementary, GEPT Intermediate, GEPT High-Intermediate, GEPT Advanced, or choose *other* and provide the name of another test that they had taken.

Participants were also asked to select from one of four choices to indicate their reasons for taking each test: *study, work, interest, and other*. Three additional responses were collected in order to account for participants who indicated multiple reasons, i.e. *study + work, work + interest, and study + interest*.

For the TOEFL tests, IELTS, TOEIC, and BULATS, and FLPT, blanks were provided for participants to write down the numerical score that they received. For each of the GEPT tests, participants could indicate whether they had passed or failed. The scores reported by the participants were equated to the Civil Service English Examination Scoring Table, and a CEFR level was assigned to each score. For participants who provided scores for more than one English proficiency test, the highest of those was reported as that participant's CEFR level.

Using the information provided on the questionnaire, data about the number of English tests taken by each participant was also collected for analysis. A further question asked participants to indicate whether they had never taken an SELPT, and if not, to provide a reason (Item 10). The qualitative reasons reported by the participants were grouped based on similarity of ideas expressed, and each group was assigned a nominal value to allow for quantitative analysis.

English study. Test impact studies typically attempt to measure the influence of tests on studying and teaching methods, known as washback. This questionnaire asked participants to respond to four questions about their study or practice of English with relation to the test. Item 11 asked participants whether they currently study English, and Item 12 whether they studied or practiced English before preparing to take an English proficiency test. Item 13 asked whether the participants attended a test preparation course before taking an English proficiency test. Finally, Item 14 assessed whether the participants continued to study or practice English regularly after taking an English proficiency test.

Test impression. One of the key purposes of this questionnaire is to identify how the participants view English proficiency tests in general. It is expected that their perceptions of these tests may have an influence on the tests' impact as well as the effectiveness of the testing

policy to achieve its aims. In this section, participants completed seven items by responding to statements about their impression of English proficiency tests by selecting from 5-option Likert scales. Numerical values were assigned to each of the Likert options, from 1 (*strongly disagree*), to 5 (*strongly agree*). The seven items gathered information about confidence (Item 15), knowledge of test format and content (Item 16), beliefs about test accuracy (Item 18), test selection strategies (Items 17 and 19), the impact of anxiety on test scores (Item 20), and the impact of the testing policy on the study and practice of English (Item 21). One final item in this section item invited participants to respond to an open-ended question on the influence of the English proficiency testing policy on the participants' use or study of English. This item differs from the others in this section not only because it is open-ended, but more importantly, because it asks about the impact of the testing policy.

Policy effectiveness. In contrast to the items in the preceding section that asked about the participants' impressions of English proficiency tests in general, those in this section focus on the participants' evaluation of the effectiveness of the government's policy of recognizing scores on English proficiency tests as a promotion scoring criterion. The questionnaire collected the participants' responses to nine statements through the use of 5-option Likert scales, and numerical values were assigned to the responses, from 1 (*strongly disagree*) to 5 (*strongly agree*). Three of the items (Items 22, 23, and 26) addressed the testing policy's impact on motivation to study English. It is hypothesized that the testing policy will have a significant effect on motivation because of the relatively high value of the promotion scoring points associated with the proficiency levels linked to scores on the English proficiency tests. Other items queried participants' about their preference for the test policy (Item 24), the fairness of the test policy (Item 27), and the impact that preparing for a test would have on their English (Item 25). Two

items were designed to assess the impact of the testing policy on the participants' confidence to use English at work (Item 28) and on their ability to use the language to carry out their duties (Item 29). The final item (Item 30) asked participants to evaluate the impact of the policy on the professional quality of the civil service as a whole. Items 23, 24, and 27 asked for agreement or disagreement with a statement that included *not* in order to discourage participants from simply marking all items with the same response. The numerical values assigned to the responses for these items were reversed from those used with the other six items; a value of 1 was assigned to *strongly agree* and 5 to *strongly disagree*. All of the questions in this section were designed to reveal whether the participants had a positive or negative evaluation of the policy's effectiveness. Responses that strongly agree or agree with Items 22, 25, 26, 28, 29, and 30, or *strongly disagree* or *disagree* with Items 23, 24 and 27, indicate a positive evaluation of the policy. As this is the case, the sum of the responses to the items in this section can be calculated and serve as an indicator of each participant's overall evaluation of the testing policy.

Interviews

The researcher conducted interviews with representatives of seven separate agencies of Taiwan's central government between September 2010 and March 2011. The first interview with a Director-General of a Ministry of Finance agency was conducted informally, but the latter six were semi-structured and based on sets of questions that were supplied to the interview subjects prior to the interview. The focus of each interview varied according to the role of the agency in the SELPT promotion criteria policy cycle and the subject's position and responsibilities.

The questions used in the six semi-structured interviews were drawn up by the researcher in English and revised after discussion with the thesis advisor. After revisions to the original English versions of the questions were made, each set of questions was translated into Chinese

(using the traditional Chinese characters used in Taiwan) by a native speaker of Chinese who has a professional background in English testing and education. The Chinese versions of the questions were revised in consultation with the thesis advisor prior to the presentation of the thesis proposal to the thesis committee on October 27, 2010. The interview questions in both English and Chinese can be found in Appendix E.

The interview subjects, except for Subject 1 from the Ministry of Finance, were contacted with registered letters sent to the Secretariat Departments of the various agencies in November 2010.² These letters identified the researcher and explained the topic and purpose of the thesis research. Representatives of the Examination Yuan, the CBC, and the RDEC contacted the researcher within days of receiving the registered letter and requested copies of the interview questions for their review.

Following this review, interviews with representatives of these agencies were then scheduled to take place in November 2010. A representative of one agency, the Ministry of Finance, contacted the researcher by telephone and requested a copy of the interview questions. The questions were provided, but no further communication was initiated by personnel of the ministry. The researcher decided not to pursue further contact with the ministry through this channel, assuming that the agency had chosen not to participate in this aspect of the research project.

The researcher attempted to contact the remaining three agencies via email, and was successful in establishing communication with the Ministry of Economic Affairs in December 2010. After receiving a copy of the interview questions, a representative of the ministry's Personnel Department scheduled an interview for early January 2011.

² The subject of the first interview is personally known to the researcher and was contacted directly.

After having made several unsuccessful attempts to contact the MOE and the CPA, the researcher wrote to the Office of the President in January 2010 via a website provided to residents of Taiwan who are seeking help with matters related to the government. A formal request for assistance was submitted. In early February, an email from that office was received, informing the researcher that the request would be forwarded to the relevant agencies. In late February and early March 2011, communication was established with these two agencies, and interviews with representatives were scheduled after copies of the interview questions were provided for review.

Table 2

Interviews: Subjects and Schedule

#	Agency	Subject's position	Language	Date
1	MOF	Director-General	E	09/21/2010
2	Examination Yuan	Administrative Vice-Minister	E / C	11/19/2010
3	CBC	Personnel Department Staff	E / C	11/23/2010
4	RDEC	Executive Officer	E	11/23/2010
5	MOEA	Personnel Department Chief	E / C	01/04/2011
6	CPA	Human Resources Section Chief	E / C	03/09/2011
7	MOE	Social Education Section Chief	E	03/29/2011

Note. E = English; C = Chinese.

All seven of the interviews took place in the offices of the various agencies. All of the interviews were digitally recorded, and detailed notes of their contents were produced. Interviews #1, #4, and #7 were conducted in English; Interviews #2, #3, #5, and #6 were conducted in a mixture of both Mandarin Chinese and English. A Taiwanese bilingual Chinese-English research assistant accompanied the researcher to Interviews #2, #3, #6, and #7 to aid in interpretation. An English teacher employed by the MOEA was present at Interview #5 to aid in

interpretation and to provide information about the ministry's English training program for its employees.

Each interview lasted between 30 and 90 minutes, depending on the subjects' willingness to specifically address individual questions and to elaborate on their answers in general. Several of the subjects provided documents related to the testing policy that had been produced by their own agency or by other agencies involved in the policy cycle. The contents of the interviews and the documents that were provided by representatives of the agencies will be discussed in Chapter 5.

Online government publications and internal documents

Government agencies in Taiwan maintain an online presence, predominantly in Chinese, but also in English and other foreign languages. Information about the organization, mission, policies, regulations, and achievements of these agencies, as they relate to the enhancement and testing of English proficiency, were collected by the researcher over the course of this research project. The websites from which this information was gathered are listed in the References section of this study. Such information was used as reference when drafting the questions for the interviews with government officials and when filling in the policy analysis framework.

When conducting interviews with government officials, requests for documents relating to the testing policy were made. Numerous documents containing details about specific measures employed within policy-implementing agencies, records of achievement of targets, and revisions to the promotion criteria standards were provided. As these documents were intended for use within the government, they were all written in Chinese. Translators were employed to create summaries of key portions of these documents, and data from these are discussed in Chapter 5.

Extended Geelhoed-Schouwstra (G-S) policy analysis framework

In order to provide a structure for the discussion of the results of the interview, and also to relate them to the questionnaire results, a policy analysis framework is employed. The extended G-S policy analysis framework was adopted both as a model for conceptualizing the policy-making process and also as a framework for analyzing the role of the various participants and stakeholders in the process. The original 2002 G-S framework includes six steps within the policy cycle, and the 2006 extended framework incorporates both a conceptual framework and an institutional framework that provide a specific context, consisting of geographical, cultural, and historical factors, for the policy cycle. As recommended by the authors of the G-S framework, the framework's six stages were outlined to identify the key features of the policy, and then the conceptual and institutional frameworks were integrated into the analysis. This procedure was carried out in order to identify the separate and potentially conflicting interests of the policy stakeholders, both the formal institutions (agencies) as well as the informal (individuals), that may be linked to the policy outcomes that diverged from policy-makers' expectations. The extended G-S policy analysis framework is presented in Figure 4.

The first of the six steps in the G-S framework is identified as *Goals*, which includes both strategic and tactical goals. Strategic goals are broader and more long-range than tactical goals, which may serve as benchmarks and facilitate progress toward strategic goals. The research gathered data from documents that spell out the government's goals, including the Challenge 2008 national development plan and the 2009 Plan for Enhancing National English Proficiency. This data established a basis for comparing changes to the goals over time and also provided insight into the establishment of specific objectives.

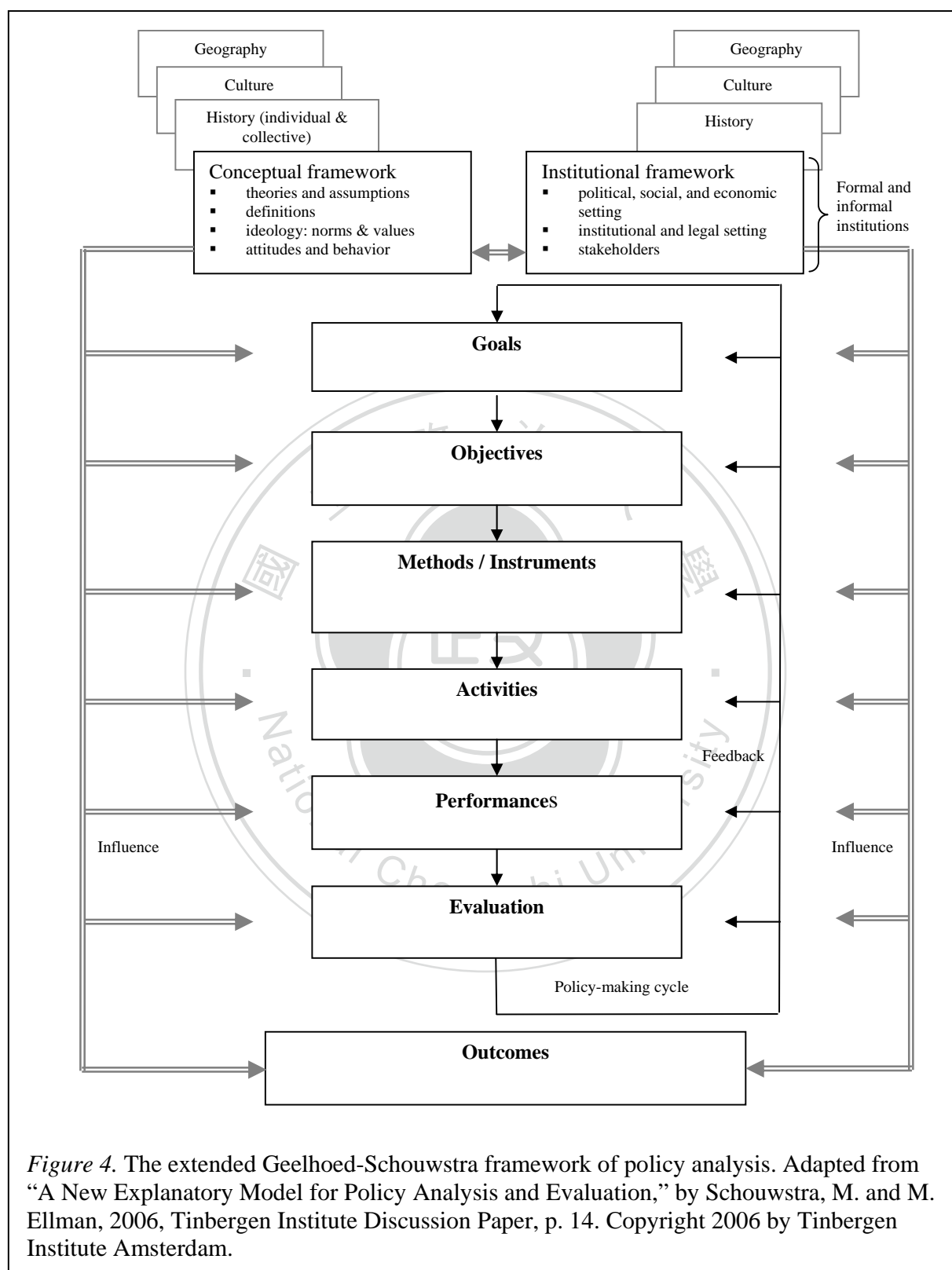


Figure 4. The extended Geelhoed-Schouwstra framework of policy analysis. Adapted from “A New Explanatory Model for Policy Analysis and Evaluation,” by Schouwstra, M. and M. Ellman, 2006, Tinbergen Institute Discussion Paper, p. 14. Copyright 2006 by Tinbergen Institute Amsterdam.

The RDEC has the responsibility for defining the operationalized goals that make up the *Objectives*, the second stage of the policy cycle as conceptualized in the G-S framework.

Whereas goals may be less likely to change in the short-term and are more closely related to wider social goals, objectives may be more sensitive to feedback from within the policy cycle and thus subject to more frequent revision. Difficulties associated with methods or activities may influence the policy-makers to alter objectives in order to more effectively work toward goals. The results of the interview with the representative of the RDEC clarified both the broader goals that guide the policy cycle and the objectives that link the upstream goals and downstream methods.

For the third stage of the policy analysis framework, *Methods/Instruments*, the roles of the MOE and the Examination Yuan were investigated. It is at this stage of the policy cycle that experts are called upon to discuss the relative merit of various alternatives. These would include proposals for adoption of a particular English proficiency test or multiple tests as the standard by which English proficiency can be measured, and the consideration of different language proficiency frameworks that could be employed. The outcome of this stage of the cycle is the selection of one or more methods that are intended to achieve the goals and objectives set forth in the previous stages of the cycle.

The fourth stage of the cycle, *Activities*, includes the development and implementation of specific measures that facilitate the use of the chosen methods. In the context of this study, this stage of the policy cycle included actions taken by the CPA, the three agencies (CBC, MOF, and MOEA), and the civil servants themselves. The interviews with the representatives of those four agencies produced evidence of the specific measures that were utilized to achieve the policy's

goals and the rationale behind their selection. The questionnaire results provided information about how the civil servants responded to the methods and activities adopted.

The fifth stage of the G-S policy analysis framework, *Performances*, examined the immediate results of the testing policy in terms of the quantitative measures used by policy-makers in order to assess how effectively the policy's goals and objectives have been met. These included the overall percentage of civil servants that have demonstrated English proficiency by passing a recognized English proficiency test, as well as the percentage of employees of specific agencies that have reached that goal. The primary sources of information for this stage of the analysis were government documents; interviews with representatives of the CPA, CBC, MOF, and MOEA; and the results of the questionnaires completed by the participants. While the interviews with the representative of these agencies' personnel departments facilitated analysis from a top-down perspective, the questionnaire data contributed to a greater understanding of the impact of the English testing policy from a bottom-up view that complements the preceding analysis.

The sixth and final stage of the G-S policy analysis framework is *Evaluation*, and it is at this phase of the investigation that attention is turned to comparisons of the actual results versus the desired outcome with regard to the policy's objectives and the activities that were designed to achieve them. When methods or activities within the policy cycle were altered as a result of analysis of performance, it was an indication that evaluation has taken place. The interviews with representatives of the RDEC, MOEX, CPA, and MOE produced evidence of the evaluative measures that were carried out and their influence on the policy cycle.

Outcomes, which appears at the bottom of the framework, is a product of the six stages of the policy cycle under the influence of the conceptual and institutional contexts. In one sense,

policy outcomes are external to the policy itself since they may be the result of factors included within the policy cycle and also contextual elements that are exogenous to the process of making and implementing policy. Policy consequences that are not measured as performance indicators and used for evaluation may be considered as outcomes. These outcomes may include impact on formal institutions that are directly involved in the policy, such as policy-making agencies, agencies that implement policies, or organizations that develop and administer English proficiency tests; individuals within these institutions that are directly influenced by the policy; institutions that are indirectly influenced by the policy, such as private companies that adopt similar strategies based on a policy's effectiveness; and individuals in the society at large that come under the indirect influence of the policy's consequences. Unlike performances, which are relatively easy to quantify, outcomes are more complex and thus more difficult to assess.

The conceptual and institutional frameworks that appear at the top of the extended G-S policy analysis framework are intended to allow for consideration conceptual and institutional factors that may have an influence on the policy's consequences. The conceptual framework included theories and assumptions unique to Taiwan, not only those that are related to English proficiency and testing, but also values and beliefs about employment, motivation, and the civil service. The institutional framework similarly influences the policy cycle through the interaction of policy stakeholders in a specific context, including political, social, economic, and legal factors. As the framework indicates, these factors influence all stages of the policy cycle, and therefore have an impact on the choice of goals, objectives, methods, and activities, the quality of performance, as well as the methods of evaluation and the actions taken based on the results of evaluation. The study of the impact of English language tests typically focus on how the use of such tests affects stakeholders, including test-takers, test users, test developers, and the wider

society as well. It is the integration of conceptual and institutional factors into the policy analysis that provides a rationale for conducting a study that incorporates features of both policy analysis and test impact methodology. The survey of test-takers seeks to explore how *micro* factors relate to the perceptions and actions of individuals influence the *macro* outcomes of an English language proficiency testing policy in Taiwan.

Conclusion

The research methodology employed in this study incorporates a variety of approaches in order to gather quantitative as well as qualitative data. The review of key documents, including those available publicly via the Internet or those circulated within the relevant government agencies to communicate the policy aims, elements, and outcomes to the various stakeholders, facilitated an analysis of the English testing policy for civil servants in Taiwan. The interviews with government officials involved in making and implementing the policy provided data that were used to complete the G-S policy analysis framework and increase the transparency of the policy cycle. It was anticipated that reviewing the roles of the various agencies within the process could shed light on factors that have a significant influence on the outcomes of the policy. Recognizing that research that solely examines policy from a top-down perspective could result in an overly narrow view of the process, it was hoped that the survey of the employees of the CBC, MOEA, and MOF could give equal weight to a bottom-up view of the impact of the policy. Collecting the experiences and beliefs of the civil servants who are affected by the English testing policy is critical to gaining an overall view of this issue. Therefore, the qualitative comments about the impact of the policy that were expressed by the participants via the questionnaire reveal insight into the features of the policy that affect civil servants as individuals

and as members of the agencies that play a key role in Taiwan's national development and influence its global competitiveness.



Chapter 5

Results and Discussion

The results of the data analysis procedure are presented in this chapter, and they reflect the multi-method approach utilized in this research. First, the results of the questionnaire completed by employees of the CBC, MOEA, and the MOF are presented, focusing first on the descriptive statistics summarizing the responses of the participants, and then discussing the significance of these results. The results of statistical analysis of the questionnaire data are offered next. Then the results of the interviews with representatives of seven central government agencies are summarized. The key issues discussed in the different interviews and gathered from official documents are presented with reference to the G-S policy analysis framework, first focusing on the government goals, and then the successive stages of the policy cycle, namely objectives, methods, activities, performance, and evaluation.

Questionnaire

The questionnaire results are presented and discussed in the following sections. See Appendix F for the complete results.

Demographics. The results from this section are found in Table 3. The participants in the survey are employees of the CBC, the MOF, and the MOEA. According to CPA data from 2009 Q2, the total number of CBC employees at the elementary, junior, or senior rank totaled 1,060. This number is far less than the number of employees of similar status at the other two agencies. If the employees of state-run enterprises are included in the totals, the workforce of the MOEA numbers 24,078, while the MOF comprises 24,854. The sampled population of the CBC workforce accounts for 10% of the total number of employees, while the MOEA and MOF samples account for just 0.3% and 0.4%, respectively. If the total number of employees of the

MOEA is calculated as including only those working in staff units or administrative agencies, the total workforce totals 4,506. Based on this number, the sampled population from MOEA makes up 1.7% of the total number of civil servants employed by that agency.

Comparing the data for the study's sample population to CPA data reveals significant differences. Women make up a greater percentage in the sample than in the CPA data, 56.7% compared to 43.5%, respectively. Civil servants at the elementary rank comprise a smaller percentage in the sample population than in the civil service as whole, 11.3% compared to 40.5%, respectively. In terms of age, younger people made up a greater portion of the total in the study sample than in the CPA data. Civil servants between the ages of 25 and 39 made up 51.2% of the sample population compared to just 27.6% of the civil service as a whole. The education level of the study sample was considerably higher than the population of civil servants as a whole. Participants with master's or doctorate degrees made up 43.7% of total, while in the civil service as a whole, only 19.4% of employees possess such degrees. It was anticipated that the demographics of the study sample would differ significantly from that of the civil service as a whole. It is assumed that the three agencies, all concerned with finance and economics, require personnel with a relatively high level of both education and English ability due to the highly professional nature of their work.

Table 3

Participant Demographics

Gender		(N=277)					
Female	Male						
56.7%	43.3%						
Agency		(N=282)					
CBC	MOF	MOEA					
37.9%	34.8%	27.3%					
Department ^a		(N=252)					
CBC Foreign Exchange		12.3%					
CBC Various		7.9%					
CBC Financial Inspection		8.7%					
CBC Economic Research		11.5%					
MOF Local Tax Agencies		9.9%					
MOF National Level Agencies		12.1%					
MOF Customs		9.1%					
MOEA Taiwan Power Co.		10.7%					
MOEA Staff Units		7.9%					
MOEA Administrative Agencies		8.3%					
Civil Service Grade ^b		(N=282)					
Elementary A	Elementary B	Junior A	Junior B	Senior			
1.8%	10.5%	36.3%	47.4%	5.8%			
Years of Service ^c		(N=258)					
1-5	6-10	11-15	16-20	21-25	26-30	31-37	
40.7%	11.6%	11.2%	13.2%	10.1%	11.6%	5.2%	
Age		(N=281)					
25-29	30-34	35-39	40-44	45-49	50-54	55-59	60+
15.3%	23.8%	12.1%	15.7%	13.2%	11.4%	7.1%	1.4%
Education Level		(N=276)					
5-year College	Univ. Undergrad	Univ. Graduate	University Post –grad				
7.2%	49.3%	41.7%	1.8%				
Academic major		(N=271)					
Finance / Economics	Other	Law / Admin. / Sociology	Physical Science	Arts and Literature			
55.7%	17%	12.2%	11.8%	3.3%			

^aThis variable was transformed by combining members of different departments to produce groups of comparable size in which the members have similar areas of responsibility.

^bThe civil service grades reported by the participants were transformed to reduce the number of grades (from 14 to 5) and to allow for comparison among agencies with different grade systems.

^cThe participants reported this variable as a specific number of years. The groups were formed after collection to facilitate presentation.

English language. This section of the questionnaire includes information about the participants' self-assessed English ability, the importance of English for their work, the experience taking English language proficiency tests, and their study of English. The general aim of this section is to: (a) present data that demonstrates general trends and average figure in order to accurately characterize the sample population, (b) suggest explanations for the observed data, and (c) provide evidence to support conclusions about the effectiveness of testing policy measures.

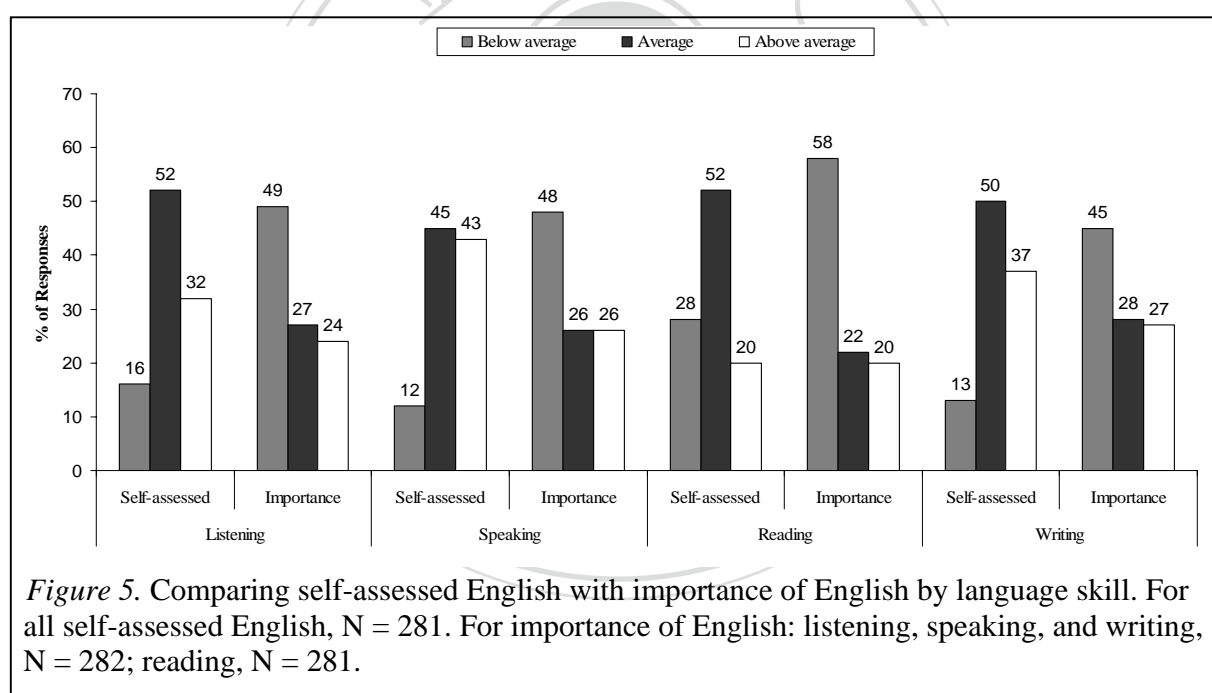
The self-assessed English data demonstrate the participants' abilities for each of the language skills and indicate their relative strengths and weaknesses. This data represents the supply of English language resources that the government may draw on. The data about the importance of English indicates the relative importance of each of the English language skills for the participants' work and represents the government's demand for English language resources. The participants' comments about the work activities for which they most commonly use English are also reported.

The self-assessed English and importance of English results are compared to each other in order to demonstrate whether there is a surplus or shortage of English language resources for each of the language skills. Similarly, overall averages for both self-assessed English and importance of English are compared to each other to indicate whether there is a relative shortage or surplus of English language resources at lower, middle, and higher proficiency levels.

The English proficiency testing experience data includes the participants' reasons for taking or not taking an English test, the number of tests they have taken, and the participants' CEFR levels based on their self-reported test scores. Finally, the English study data helps to characterize how the participants' study of English is related to English proficiency testing. Of

particular interest is whether taking an English test leads participants to attend test preparation courses or continue studying English after taking a test.

Self-assessed English ability. The participants assessed their English ability on a 5-point Likert scale for each of the four language skills (listening, speaking, reading, and writing). The choices were labeled from worst to best: *very poor*, *poor*, *not bad*, *good*, *very good*. To facilitate comparison, these responses have been combined to form three groups: *below average* includes responses of *very poor* and *poor*; *average* includes only *not bad* responses; *above average* is comprised of *good* and *very good*. Refer to Figure 5 for the results.



Approximately one-half of the participants rated their English ability as average in all four language skills, ranging from a high of 52% for listening and reading to a low of 45% for speaking. Speaking and writing had the highest percentage of participants who rated their ability to be *above average*, at 43% and 37%, respectively. Reading had the highest percentage of

participants reporting that their English ability was below average, at 28%. These results suggest that participants are most confident in their speaking and writing and least confident in their English reading ability.

Importance of English. The participants indicated the importance of the four English language skills for their work by selecting from one of five choices on a Likert scale. The choices were labeled from least important to most important: *no importance*, *not very important*, *no opinion*, *important*, and *very important*. These responses have been combined to form three groups: *below average* includes responses of no importance and not very important; *average* includes only no opinion responses; *above average* is comprised of important and very important. The most frequent response for each of the language skills was *below average*, ranging from a high of 58% for reading to a low of 45% for writing. These results suggest that approximately one-half of the participants estimate English ability to be of little importance to their work, while approximately one-quarter find English to be of average importance and one-quarter believe it to be of above-average importance.

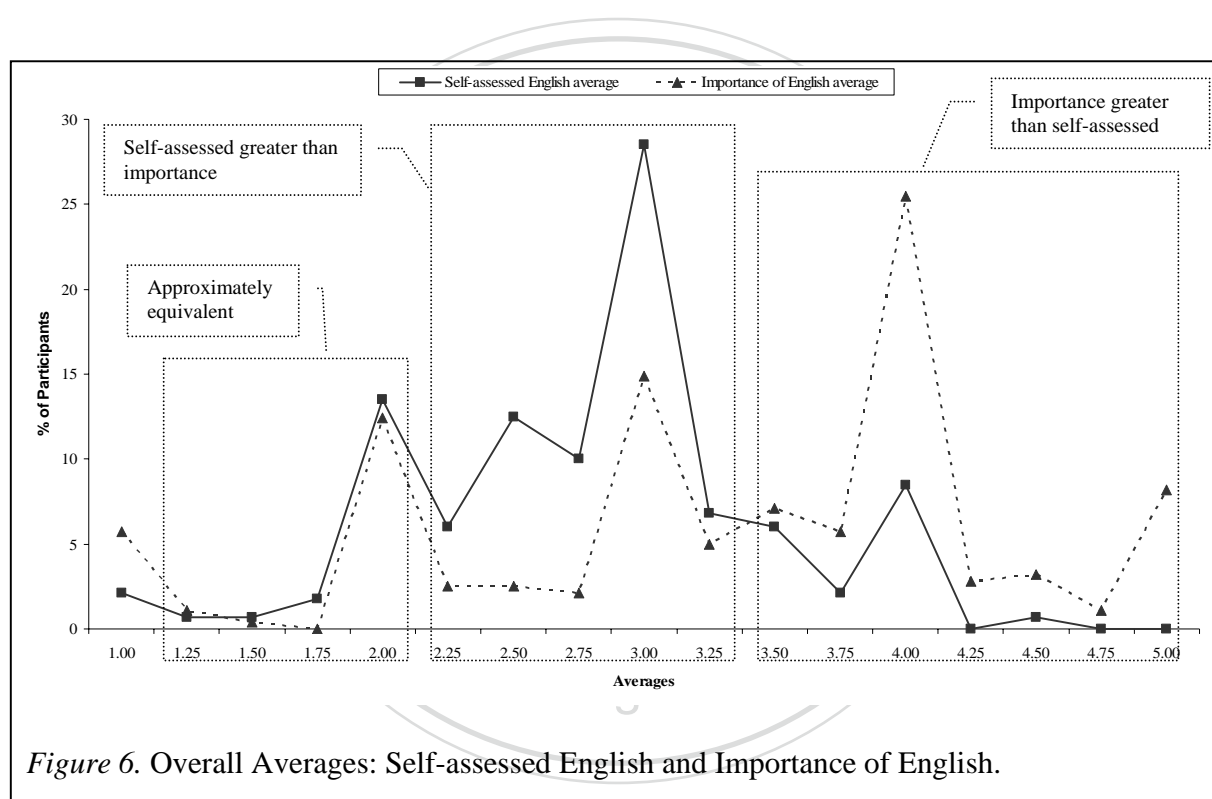
Comparing importance of English to self-assessed English by language skill, a clear trend emerges. For each of the language skills, the participants indicate that their supply of English language resources is generally more than adequate for the demand for those skills in their jobs. Only for reading is the figure for above-average self-assessed English, 20%, less than the accompanying figure, 21%, for importance of English. However, 52% of the participants report their English reading ability is average, while only 22% claim that English is of average importance for their work. For the other three language skills, the above-average self-assessed figure is greater than the above-average importance figure. The greatest difference is noted for

speaking, with 43% of participants rating their speaking ability to be above average versus only 26% who judge English speaking to be of above-average importance for their work.

In addition to rating the importance of English for their jobs, the participants briefly commented about how they used English for their work. In all, 125 participants, or 44.3% of the total, provided 166 responses to this item. Their responses were translated into English and grouped into 7 categories. Of those who responded to this item, 67% reported just one activity requiring English language proficiency, while 29% reported two activities, and just 4% reported three. Reading was the activity reported most frequently, making up approximately 37% of all the responses. Attending international conferences and talking with foreign visitors each comprised 21% of the total. Translating made up 8% of the total, while writing reports contributed 7%. Writing letters or email and using the telephone each made up approximately 3% of the total.

The responses to this item give a significantly different view of the importance of English language skills for the work done by the participants than do the responses to the quantitative items reported above. One reason for the difference could be that the response rate to the quantitative items is much higher than for the qualitative item, but the divergence may also be explained by the fact that the qualitative item required the participants to think more deeply about their actual duties than they did for the quantitative item. It would be valuable to carry out a more thorough English language needs analysis for civil servants. Such a study could help to identify appropriate English proficiency levels for different jobs and civil service grades, to aid in the design or selection of English proficiency tests for civil servants, and to guide the development of curricula for English language training programs.

Overall averages: Self-assessed English and importance of English. The choices for each language skill for self-assessed English and importance of English were assigned a numeric value, with *very poor/no importance* equal to 1, and *very good/very important* equal to 5. The sum of each participant's responses for the four skills for both self-assessed English and importance of English were divided by four to produce an average figure. The frequencies of these averages are presented in Figure 6.



Comparing the average figures for *self-assessed English* and *importance of English* reveals three distinct trends. The percentage of participants in jobs for which English is of below-average importance, from 1.25 to 2.0, is approximately equivalent to the percentage of participants with English proficiency that is below average. The percentage of participants who

report that English is of average importance, from 2.25 to 3.25, is lower than the percentage who report that their English proficiency is average. Finally, the percentage of those who report that English is of above average importance for their work, from 3.5 to 5.0, is greater than that of those who claim their English proficiency is above average.

These results suggest that in these three agencies, there is a surplus of civil servants with an average level of English proficiency but a shortage of those with English proficiency that is above average. In terms of promotion scoring, the scoring values associated with basic English proficiency levels may be adequate to motivate employees to improve their English. This conclusion is supported by the fact that there is surplus of participants with English proficiency of an average level. On the other hand, the apparent shortage of participants with above-average English proficiency suggest that the promotion scoring values associated with higher levels of proficiency could be insufficiently high to motivate further improvement. There is a potential danger, however, of assigning promotion scoring values that are too high, since such an action could result in civil servants ignoring their duties and devoting an inordinate amount of time to improving their English. Therefore, it may be prudent to increase opportunities to improve English proficiency for civil servants in positions in which English is of average or above-average importance without appreciably altering the weighting of promotion scoring values.

Testing experience. A total of 270 participants provided information about whether they had taken an English proficiency test recognized by the MOE. Anticipating potential difficulties in calculating the number of participants who had taken an English test, several methods were employed to arrive at this figure. First of all, in answer to Item 9a, 182 participants, or 64.5% of the total study sample, reported the name of at least one English test they had taken. Another item (10a), found that 31.9% of the participants reported they had never taken an English

proficiency test. The sum of these two numbers is 96.4%, and while it is not equal to 100%, the fact that it is relatively close may suggest that the 64.5% figure may have some basis in reality. It should be emphasized that this number represents the percentage of the participants who have reported having taken, not having passed, an English proficiency test. Of the 282 participants in the study, 124 (or 43.9%) provided a score that could be equated with a CEFR proficiency level. It is not possible to say what percentage of those who did not report a score on a recognized English proficiency test achieved an A2 on the CEFR scale.

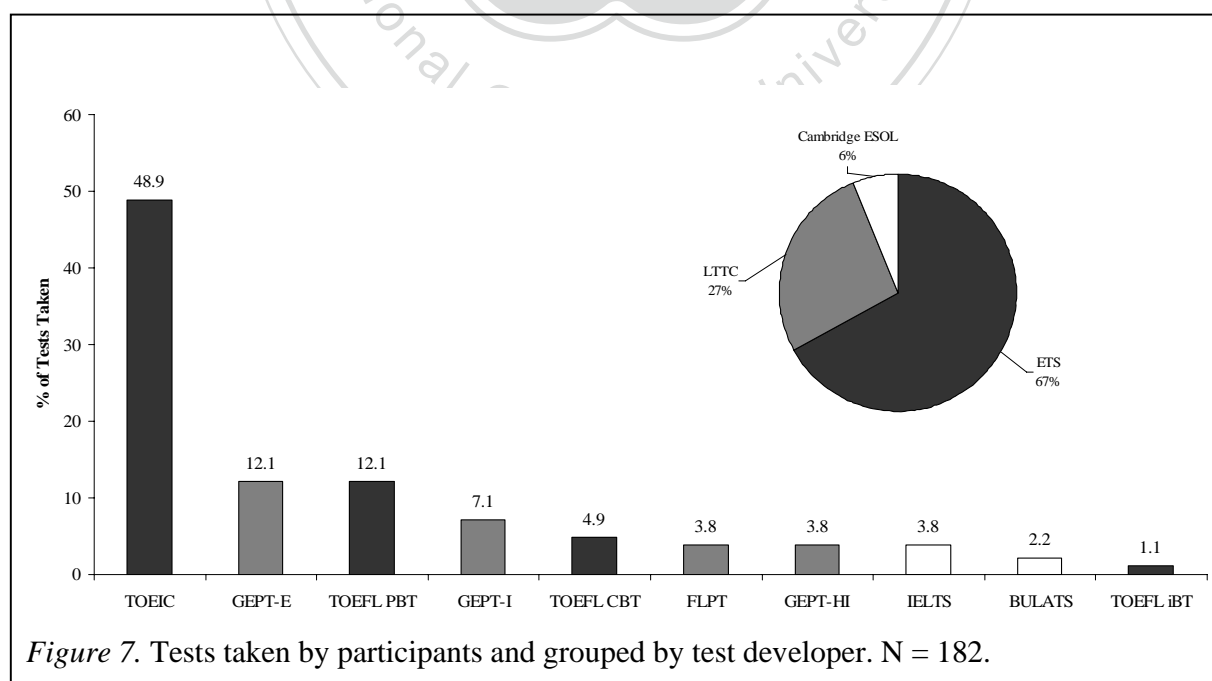
The figure of 43.9% is significantly higher than the 25% figure reported by the CPA. This discrepancy could have several explanations. Firstly, the number of participants who claim to have reached the A2 level may be either higher or lower than the actual figure by an undetermined degree. Secondly, the sample may not reflect the population of civil servants as a whole based on the fact that the study focused on three agencies whose need for English-speaking employees was estimated to be relatively higher than average. Overall, the margin of error could be substantial, and it is difficult to conclude that the figures arrived at by this study have a high degree of confidence.

Of those participants who reported not having taken an English proficiency test, 43.1% claimed that taking such a test was unnecessary. There was no information provided about why taking an SELPT was not needed, but it may be assumed that English proficiency is not important for these participants' jobs. If that were the case, then English examination promotion scoring points might have little impact on the likelihood of receiving a promotion. Alternatively, those participants may believe that English proficiency would have little impact on their opportunity to receive a promotion. In that case, the promotion scoring associated with achieving the CEFR A2 level may have little ability to motivate them to improve their English.

A further 29.3% reported that they were not yet ready to take an English test. It seems likely that these participants may plan to take an English examination in the future. A further 12.1% of the participants reported that they had no time to prepare to take an English proficiency test, while 10.3% were not interested. Of these two groups, the former is more likely than the latter to decide at some point in the future to take an English test.

A total of 52.2% of those participants who provided information about having taken an English test reported taking just one, while 11.1% reported having taken two. Fewer than 4% took more than two English tests, with one participant reporting having taken five tests, the highest number of tests taken by any participant.

In all, the participants reported taking 182 separate tests. Of these, 48.9% were TOEIC, with GEPT-E the next most commonly cited, making up 12.1%. Overall, 67% of those participants who reported having taken an English proficiency test took a test developed by ETS, while 27% took an LTTC-developed test, and just 6% took one of Cambridge ESOL's tests. Refer to Figure 7 for a detailed breakdown of the tests taken.



The participants were not asked to give a reason for choosing to take a particular English proficiency test. It would be useful for future studies on this topic to enquire into the factors that influenced this decision. Given this limitation, the observed preference for taking TOEIC by the study's participants is likely attributable to several factors. First, TOEIC maintains a high visibility in Taiwan due to advertising and the marketing efforts of ETS's local agent.

Furthermore, TOEIC's brand image benefits from its association with TOEFL, which has a long history in Taiwan. Thirdly, it is claimed that TOEIC is particularly suited to testing the English skills that are used in the workplace. It is likely that a narrower focus for the intended uses of an English proficiency test would support a claim that a test is better suited to a particular use. A testing system such as the GEPT, with a more broadly defined *general* intended use, encompassing both academic and workplace contexts, could be perceived as less suited to a specific use than a test with a narrower context of use. Finally, the fact that TOEIC did not include a speaking test until 2006 may have given test-takers the impression that it would be easier to earn a higher score on TOEIC than tests that assessed all four language skills. The importance of this factor is undoubtedly decreasing now that TOEIC assesses all four skills.

Approximately one-half as many participants reported taking one of the GEPT levels as those taking TOEIC. Over one-half of those who took the GEPT reported having taken the Elementary level. It is likely that participants may elect to take the GEPT Elementary test prior to taking the higher levels. Passing the first level of the test would probably give test-takers increased confidence in their English ability and increase their familiarity with test format and content. The GEPT has relatively high visibility in Taiwan, but the fact that it is not administered by a private, for-profit company suggests that the marketing efforts employed to promote it are considerably more limited than those for TOEIC. This undoubtedly works to diminish its

popularity among English learners and test users. However, its association with the government by virtue of the MOE's support for its development contributes to its acceptance. It is likely that its local origin may appeal to some learners who may feel that it is more appropriate for English learners in Taiwan, but its local roots may also diminish its acceptance among learners who believe that a test developed in an English-speaking country may be more authentic. The GEPT's general lack of international recognition may also tend to diminish its popularity relative to tests developed in English-speaking countries. Finally, the fact that the GEPT is a four-skills test, and therefore potentially more difficult on which to receive high marks, might also influence English learners to opt for a test with no speaking portion.

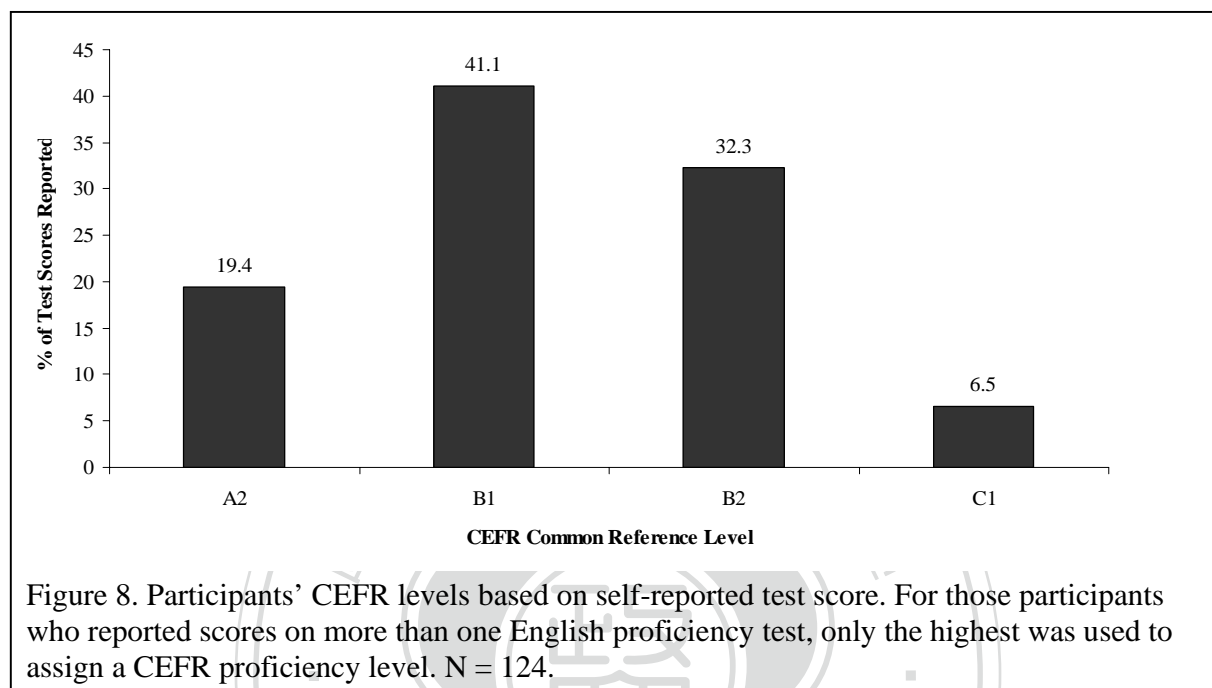
The third most frequently taken English test was the TOEFL PBT, which has not been administered in Taiwan since 1998. TOEFL CBT, administered from 1998 to 2006, made up just 4.9% of the total; and the TOEFL iBT, administered in Taiwan since 2006, comprised just 1.1% of the total. Given that TOEFL is primarily targeted at those desiring to pursue advanced studies in an English-speaking country, it is not surprising that the percentage of participants who reported having taken one of the forms of this test is so low. It is likely that those who reported having taken one of the TOEFL tests may have studied in the U.S.A. or at a university in another English-speaking nation where the TOEFL is recognized for admission.

The FLPT, which makes up 3.8% of all tests reported as being taken, has a long history of use in Taiwan, particularly by government agencies. Compared to the GEPT, its visibility is lower, but the fact that it is basically norm-referenced rather than criterion-referenced, could be one rationale for its continued popularity. Its more familiar format and content may be attractive to some English learners. In fact, it was reported by the MOEA that its English training program includes a course to prepare students for taking the FLPT.

While tests developed by Cambridge ESOL have a generally high reputation internationally, only 6% of the tests taken by the participants were developed by this organization. The Academic Module of the IELTS test is commonly taken by students who wish to pursue advanced education at universities in English-speaking countries such as the U.K. and Australia. IELTS made up only 3.8% of tests reported as being taken. No information was collected about which IELTS modules the participants took. BULATS, which is intended specifically to assess the English proficiency of learners who wish to use English in the workplace, made up just 2.2% of the tests that were reported as being taken by the participants.

Of the 182 English tests reportedly taken, reasons were provided in 179 instances. *Work* was the reason cited most often, accounting for 51.4% of the responses. *Study* was cited in 20.1% of the instances, and *Interest* accounted for 16.2% of the total. These results suggest that most of the participants are motivated to take an English test because of the promotion scoring incentive. This conclusion is supported by the fact that TOEIC and the various GEPT tests enjoy the greatest popularity among the participants, whereas TOEFL and IELTS, which are mainly used by learners who wish to study overseas, are taken less frequently by the participants. While it may be useful to recognize multiple English proficiency tests for certifying civil servants' English proficiency, it is uncertain whether this situation serves the interest of the government's effort to promote the English proficiency of its employees. It may in fact be more appropriate to determine whether one or more than one test would be most suitable; however, practical considerations regarding the testing market in Taiwan may prevent such a policy from being carried out. If such an action were taken, evidence supporting the validity of such a decision would be essential.

The participants' self-reported test scores were used to assign CEFR proficiency levels based on the 2005 CPA English examination scoring chart. Refer to Figure 8 for the results.



Of those participants who reported English test scores, just 19.4% were at CEFR A2, the minimum level recognized as qualifying for promotion scoring points. Learners at this level, according to the CEFR global scale, can understand and exchange “information on familiar and routine matters.” More than 80% of the participants who reported test scores had achieved a CEFR level above A2, suggesting that the average English proficiency of civil servants is B1 or B2. The global scale of the CEFR Common Reference Levels describes learners at the B1 level as able to understand the main points “on familiar matters regularly encountered in work,” while at the B2 level, learners can “understand the main ideas of a complex text” and “can interact with a degree of fluency and spontaneity....” Based on these descriptors, civil servants with English proficiency at the B2 level would have little difficulty using English for a fairly wide range of

purposes at work, while those at B1 level would be capable of giving brief reasons and explanations and producing simple texts on familiar topics.

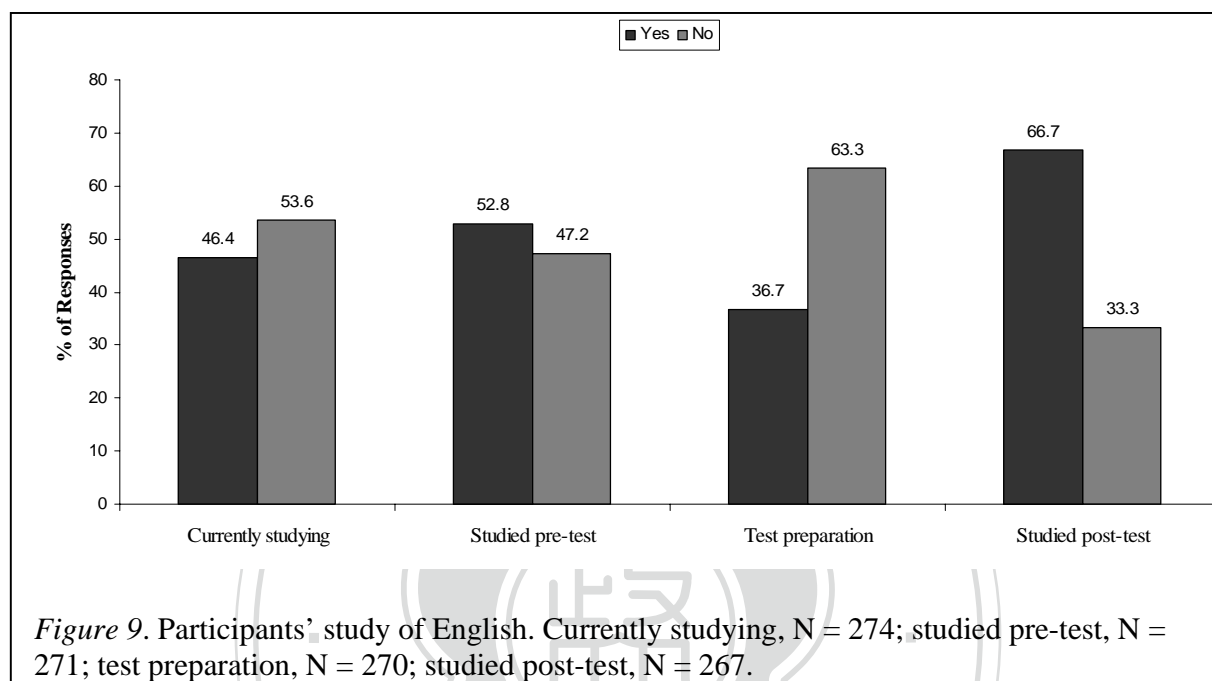
The relatively low percentage of participants reporting scores at the C1 level indicates that few civil servants have achieved English proficiency at that level. According to the CEFR global scale, learners at this level can understand more demanding texts with implied meaning, and communicate fluently, flexibly, and effectively for a range of purposes. Achieving English proficiency at this level would certainly require a considerable investment of time and energy, and it is unlikely that many civil servants would find it necessary to use English in this manner for their work.

It would be advisable for agencies encouraging their employees to improve their English for use at work to conduct needs analyses in order to understand what specific language skills would be most useful for different positions. The CEFR proficiency levels were not specifically designed for use by English learners in Taiwan, nor for civil servants, so investigation of actual needs would allow the government to set appropriate targets and implement effective measures to reach them.

English study. Fewer than one-half of the participants reported that they were currently studying English, while just over half studied before taking an English proficiency test. This indicates that participants were somewhat more likely to study English before taking a test than after, but the difference between these two measures is not great. Refer to Figure 9 for results.

Approximately two-thirds of participants reported that they studied English after taking an English test, but one-third of them stopped studying at some point after taking an English test and before completing the questionnaire. This could be interpreted to suggest that taking a test

motivated the participants to continue studying English, but that approximately 20% of those people ultimately lost interest in actively studying.



Only 36.7% of participants reported attending a test-preparation course prior to taking an English test. This suggests that most participants were somewhat confident in their English ability prior to taking the course, or that they doubted that attending a test-preparation course would increase their score.

Test impression. The participants' responses to the seven items in this section of the questionnaire indicate their perceptions about English proficiency tests in general. The results are summarized in Table 4.

Table 4

Test Impression Results

Test impression variables	N	M	SD
NRV = Anxiety affects test score	279	3.72	0.77
ESY = Choose easiest test	278	3.51	0.91
INF = Test preparation influences English use	279	3.39	0.86
KNO = Sufficient knowledge of test format and content before test	278	3.30	0.90
ACC = Language tests are accurate	278	3.15	0.86
PRC = Confident in English ability prior to taking test	277	3.03	0.88
LOC = Local tests are more appropriate	279	2.99	0.68

The participants expressed the strongest agreement with NRV ($M = 3.72$), indicating that test anxiety has a negative effect on language test scores. Given that the participants considered test anxiety to be a serious concern, the strong support for ESY ($M = 3.51$) is not surprising. The participants' preference for taking what they believe to be the easiest test could reduce the effectiveness of the testing policy. If civil servants truly believe that one test is easier than the others, they may decide to take that test regardless of its appropriateness for learners in Taiwan or its relation to the English skills that civil servants may need. It is possible that one reason for the popularity of the TOEIC, at least prior to the introduction of the speaking and writing tests in 2006, is the perception that it may be the easiest of the tests recognized by the MOE.

As for the influence of the testing policy on the participants' use or study of English, there was fairly strong support for such an impact. The responses for INF ($M = 3.39$) indicate that 49.1% of the participants believe that the policy has such an effect, while the remainder have no opinion (35.1%) or disagree/strongly disagree (15.8%). These data suggest that nearly one-half of the participants perceive the test content and tasks as relatively unrelated to their actual use of English. These participants may see little connection between developing their English proficiency to a degree that would allow them to qualify for English examination promotion

scoring points and their actual use of and need for English. If test-takers believed that an English proficiency test assessed their ability to use English in authentic situations, it is more likely that the participants would report that preparing for an English test would have a greater influence on their use of English. This implies that the government's policy of recognizing English examinations as a criterion for promotion scoring would be more effective at encouraging civil servants to improve their English if the recognized examinations were more closely related to the English skills that are useful in the workplace.

Just over one-half, or 50.7%, of the participants reported that they agreed/strongly agreed with KNO ($M = 3.30$) that they had sufficient knowledge of a particular English test's format and content prior to taking it. In contrast, just 34.1% of participants agreed/strongly agreed with PRC ($M = 3.03$) that they were confident of performing well on an English proficiency test before taking one. Greater familiarity with an English exam that one intends to take would probably give a prospective test-taker greater confidence in English ability and reduce anxiety. If a test-taker's English proficiency is considerably lower than the level required to pass the test, the influence of such knowledge may have limited benefit. However, for test-takers whose English ability is close to the target level, knowledge of a test's format and content could have a significant impact on the outcome of the test. The fact that approximately one-half (49.3%) of the participants disagreed/strongly disagreed or had no opinion about sufficient knowledge of English proficiency tests indicates that it may be advisable for test users to provide access to information about such tests. Those potential test-takers with little knowledge of English proficiency tests or experience taking them may elect to attend test preparation courses to increase the likelihood of earning a higher score, potentially resulting in shallow learning with little practical advantage for developing communicative ability.

Of the participants, 38.8% agreed/strongly agreed with ACC ($M = 3.15$) that English proficiency tests accurately assess the English ability of learners. A slightly larger portion, 39.9%, had no opinion, while just 21.3% disagreed/strongly disagreed. Overall, this indicates that the majority of participants either accept that English proficiency tests are accurate or do not consider test accuracy to be a critical issue.

Finally, the participants expressed the least agreement with LOC ($M = 2.99$) that local English proficiency tests are more suitable for civil servants in Taiwan than tests developed overseas. In all, 64.9% of participants reported having no opinion on this issue, with the remainder split nearly evenly between agree/strongly agree and disagree/strongly disagree. On this basis, the distribution of English tests taken by the participants does not seem to be strongly influenced by the origin of the individual tests.

Qualitative results. Approximately 23% of the participants provided comments in response to an open-ended item that asked them about the testing policy's influence on their study of English. Of the 64 comments that were provided, 63 of these were written in Chinese, and one was in English. Two of the comments were unrelated to the policy or its impact. The 62 valid comments were divided into six groups based on the topics that they addressed. These included (a) need for English, (b) motivation, (c) promotion, (d) interest, (e) languages other than English, and (f) various. The comments are summarized in Table 5 and discussed in the following paragraphs.

Table 5

Summary of Participants' Comments on English Testing Policy

Group	Comments	n
Need for English n = 13	No need for English	5
	English is needed	3
	If English needed, policy is valuable	2
	If English needed, policy is unnecessary	2
	Policy focuses on test, not need for English	1
Motivation n = 10	Policy provides positive motivation	6
	Policy provides negative motivation	2
	Failing a test lowers motivation	1
	Motivation is necessary	1
Promotion n = 7	English has no impact on promotion	4
	Professional skills should be promotion criteria	2
	Study English only for promotion, not for work	1
Interest n = 6	Study should be based on interest	2
	Studies English because of interest	1
	Takes test because of interest	1
	Without interest in English, impact is negative	1
	Studies because of policy, not interest	1
Other languages n = 5	Other languages should be recognized	5
Various n = 21	Studying takes time away from work	4
	Tests should not be required	3
	Studying is necessary	3
	Against test policy	2
	Policy has no influence on study of English	2
	Tests are expensive	2
	Tests are inaccurate	1
	Policy is unfair	1
	Waste of administrative effort	1
	Doesn't understand policy	1
	No experience with speaking/writing tests	1

Note. All but one of the comments were provided in Chinese and translated into English. These were summarized for ease of presentation. N = 62.

Of all the topics discussed in the comments, the need for English was mentioned the most frequently. Most of these comments related to the participants' own experience, but several

discussed the value of the policy if English were important for a particular position. Based on these comments, we may conclude that the need for English is seen as an important factor with regard to the value and impact of the policy. Of those comments about motivation, the majority expressed the opinion that the policy can provide motivation, although there were concerns expressed about negative motivation to study English associated with the testing policy, particularly if civil servants had no interest in English, or if they failed a test.

Seven comments specifically addressed the relation between the testing policy and promotions. More than half reported that English proficiency would have no impact on promotion opportunities, and it may be assumed that these individuals hold positions in which English proficiency is of little value. While the promotion scoring value of English proficiency may act as a powerful incentive, this effect may be problematic unless there is a need for English in a particular department or position.

Six comments addressed the importance of interest in English, and there was general agreement that the study of English should be based on an individual's interest in the language and not on policy incentives. Several comments indicated that the participants have an interest in languages other than English. This is significant given the fact that proficiency in other languages was recognized as a promotion scoring criterion in the 2011 amendment to the promotion scoring criteria. This indicates that the revised policy may be more acceptable to civil servants and have a more positive impact on them.

Other comments about the policy addressed a range of different issues, with the common element being that most expressed a negative view of the policy's impact. Several referred to specific problems associated with the policy, such as the time that civil servants would need to

devote to studying and the expense of taking tests, while others expressed more general dissatisfaction with the policy.

Overall, the comments collected in response to this item reveal the issues that are important to the participants in the study. While there is no absolute agreement with regard to any of the issues raised, trends may be observed, particularly with regard to the importance of English for particular jobs, the powerful incentive that the policy exerts, and the recognition that interest is important to an individual's study of English or other languages.

Policy effectiveness. The nine items in this section of the questionnaire were designed to assess the participants' evaluation of the testing policy's effectiveness. Refer to Table 6 for a summary of the results. The variables in Group 1, with a mean value >3.3 , indicate a strongly positive evaluation of the English language testing policy. Those in Group 2, with a mean value <3.3 and >3.0 , demonstrate a moderately positive evaluation. Variables in Group 3, with a mean value <3.0 , indicate a negative evaluation of the testing policy.

The responses for the variables in Group 1 indicate that the participants generally believe that the testing policy has a positive impact. The data for NOT ($M = 2.02$) suggest the participants have an intrinsic interest in studying English, and do not seek to improve their English solely to obtain promotion scoring points. The fact that the participants show fairly strong agreement with the idea that studying for an English test improves one's overall English ability, demonstrated by ABL ($M = 3.54$), is also an indication that the testing policy could encourage civil servants to prepare for taking English proficiency exams. If there were low support for this variable, it is unlikely that participants would be encouraged to prepare for and take exams since there would be little expectation that such preparation would have a positive impact on their English ability. The response trend for MOT ($M = 3.34$), with more than 50% of

participants agreeing that the policy provides them with an effective incentive to improve their English, provides further evidence of their positive evaluation of the testing policy. It may be that the effectiveness of extrinsic incentives is enhanced when such rewards encourage behavior that individuals already support.

Table 6

Policy Effectiveness Results

Group	Variables	N	M	SD	Policy evaluation
1	NOT = Study English (not) just for promotion	272	2.02	0.76	Strongly positive
	ABL = Preparing for test improves English	272	3.54	0.78	
	MOT = Policy gives appropriate incentive	272	3.34	0.94	
2	IMP = Policy has a positive impact on civil service	272	3.23	0.94	Moderately positive
	CON = Passing test increases confidence in English	271	3.18	0.91	
3	INT = Policy increases interest in English	272	2.78	0.96	Negative
	JBK = Better English does (not) affect job skills	272	3.30	0.90	
	FAR = Policy is fair to all civil servants	272	2.64	0.97	
	PRF = Prefer tests were (not) a promotion criterion	269	3.43	0.88	

Note. Three variables (NOT, JBK, and PRF) are based on items written as negative statements. A response of disagreement indicates a positive evaluation of the policy.

The results of the variables in Group 2 also indicate that participants are generally in favor of the policy, but a more moderate evaluation of the impact of the policy is suggested. The responses to CON ($M = 3.23$) suggests that many of the participants believe that passing an English proficiency test would increase their confidence in using English at work, while a significant portion are unsure about this impact. Similarly, IMP ($M = 3.18$) indicates that the participants are divided over the policy's impact on the civil service, with approximately

equivalent proportions agreeing that the impact is positive or expressing uncertainty about its overall value. These results could be influenced by several conditions. If the participants have little opportunity to use English at work, they may believe that improved English proficiency would offer them little additional promotion opportunity. They would therefore likely assign a relatively low value to the testing policy. If English were important for these participants' duties, the policy could be seen as redundant and offering little actual value. Such civil servants would already have developed substantial English proficiency in order to carry out their duties, and would realize no additional benefit from the policy. Overall, however, these results can be seen as relatively positive, since approximately 75% of the participants do not indicate that the testing policy has a negative impact.

The variables in Group 3 identify specific impacts of the testing policy that the participants may believe have a limiting influence on the policy's effectiveness. INT ($M = 2.78$) demonstrates that a significant portion of the participants believe the policy did not increase their personal interest in improving their English. This is not unexpected, since the testing policy operates as an extrinsic motivation and does not directly influence civil servants' personal interest in the English language. The results for JBK ($M = 2.7$) show that the participants generally do not believe that improving one's English would enhance their ability to carry out their work. This result could be explained by the fact that English proficiency may be valuable but not essential for most jobs in the CBC, MOEA, and MOF; thus, improving one's English would not necessarily help one to work more productively. Furthermore, FAR ($M = 2.64$) indicates that most participants were of the opinion that the policy was not fair to all civil servants. This suggests that the participants believe that those civil servants with higher English proficiency might enjoy a significant advantage in terms of opportunity for promotion relative to

those with lower English proficiency. Introducing a new promotion scoring criterion is likely to be controversial given that those civil servants who have labored for years without such an incentive would be at a disadvantage compared to those who more recently joined the civil service. At most government agencies, the number of staff is capped, and the climb up the career ladder from the lower to the higher civil service grades may be quite slow. In that case, introducing English proficiency as a promotion criterion could have the result of creating a fast track for younger civil servants. Finally, the responses for PRF ($M = 2.57$) imply that a majority of the participants are personally opposed to the use of English proficiency tests as a promotion scoring criterion. Most of the participants in the study had been civil servants before English proficiency was adopted as a promotion scoring criterion, and they may thus feel that the policy interferes with their expectations for future promotions.

Analysis of correlation among test impression and policy effectiveness variables. To gauge the relationships among the variables in the test impression and policy effectiveness sections of the questionnaire, correlation analysis was carried out using SPSS 17.0. The full results of this correlation analysis are reported Appendix G. The pairs of variables with Pearson correlation greater than .300 are included in Table 7.

A moderate to good correlation ($r > .500$) was noted among six variable pairs formed by four separate policy effectiveness variables: FAR, INT, MOT, and IMP. The fact that three of the six variables pairs with correlation stronger than .500**, FAR/IMP, MOT/IMP, and INT/IMP, include IMP, which was designed to represent a positive evaluation of the testing policy, suggests that IMP functioned as it was intended to. These results imply that fairness, the policy's influence on motivation, and the participants' personal interest in improving their English proficiency are all of critical importance to the participants' evaluation of the testing policy's

effectiveness. Motivation to study English and interest in English were also found to be important in the participants' qualitative comments.

Table 7

Pairs of Variables with Significant Pearson (r) Correlation >.300

Variable Pairs	r	Variable Pairs	r	Variable Pairs	r
FAR/IMP	.593**	FAR/CON	.467**	ACC/MOT	.356**
MOT/IMP	.585**	NOT/IMP	.462**	ABL/FAR	.356**
INT/FAR	.575**	NOT/MOT	.440**	ACC/CON	.356**
INT/IMP	.551**	PRF/MOT	-.440**	PRF/GBK	-.347**
INT/MOT	.548**	ABL/MOT	.435**	PRF/IMP	-.347**
MOT/FAR	.542**	INT/PRF	.409**	GBK/IMP	.322**
CON/IMP	.490**	INT/CON	.405**	ACC/IMP	.317**
ABL/CON	.488**	PRC/KNO	.374**	ACC/ABL	.312**
NOT/FAR	.480**	INT/ABL	.373**	ACC/INT	.311**
PRF/FAR	-.480**	ABL/IMP	.369**	LOC/CON	.301**
MOT/CON	.473**	LOC/ACC	.357*		

Note.

* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

A fair degree of relationship was found among 26 variable pairs. These results may suggest that the variables chosen for the questionnaire represent important factors that link the participants' impression of English proficiency tests and the effectiveness of the testing policy. Of those pairs of variables with fair to moderate significant correlation, only those pairs formed with PRF, which asked participants to agree that they would prefer that English proficiency were *not* a promotion criterion, demonstrate a negative value. In most cases, this was anticipated. However, it was expected that the PRF/GBK pair would result in a positive value since both of these variables include *not* in their wording.

Three test impression variables, ESY, NRV, and INF were found to exhibit only small or no significant correlation with the other variables. It is interesting that the responses to these three items demonstrate the highest levels of agreement among the participants, as shown in Figure 10, yet there is little significant correlation among the participants' responses to these

items. It may be that the ESY and NRV variables reflect factors about which the participants share similar impressions, but that they are not closely related to features of English proficiency tests. The absence of significant correlation among INF and other variables could be related to the fact that the participants' qualitative comments revealed that the testing policy had little influence on their use of English in the workplace.

Overall, the results of this analysis suggest that there is moderate correlation among 13 of the 16 variables in the test impression and policy effectiveness sections of the questionnaire. Several inferences can be made based on these results. First, the testing policy does have a powerful influence on the motivation of the participants to improve their English proficiency. This suggests that including English proficiency as a criterion for promotion scoring is likely to produce the desired outcome, increasing the percentage of civil servants who achieve English proficiency at the level of CEFR A2 or above. This outcome is likely to be mediated, however, by the civil servants' personal interest in improving their English ability. Those with a higher personal interest in improving their English perceive the value of the testing policy to be greater, whereas those with a lower personal interest in improving their English judge the testing policy to be of less value. This highlights the power of the language testing policy to affect the behavior of civil servants. Those individuals with less personal interest in developing their English proficiency may feel that the policy exerts an uncomfortable amount of influence, but in order to increase their chances of future promotion, they may feel compelled to invest time and energy in improving their English ability.

Secondly, the issue of the fairness of the policy's impact is also critical. If the weight of the promotion scoring points associated with English proficiency relative to other promotion scoring criterion is too great, the result could be to cause civil servants to form a negative

impression of the testing policy. This could potentially be a source of friction among civil servants, creating dissatisfaction among those with lower English proficiency or those in positions in which English is of less importance for their work. This implies that the weighting of English proficiency as a promotion scoring needs to be moderated to account for the differences in the civil servants' intrinsic interest in improving their English proficiency as well as the importance of English for different positions within the civil service. These results could, therefore, be used as evidence to support maintaining English proficiency as a promotion scoring criterion, but with the caveat that the weighting of this criterion be determined by individual agencies or units within agencies based on the importance of English proficiency.

Policy Analysis

In this section, the results of the analysis of the English testing policy are reported and discussed. First, the interviews with representatives of the various government agencies will be summarized. Then the stages of the policy cycle will be analyzed with reference to the extended G-S policy analysis framework. Finally, the conceptual and institutional frameworks will be discussed.

Interview results. In this section, the results of the interviews with officials of the government agencies that played a role in testing policy are reported. Although scripted questions were provided, the subject were given the freedom to address specific items from the list or to introduce issues that they felt to be relevant. The issues raised in these interviews are situated in the G-S policy analysis framework. Finally, the implications of the results of the policy analysis are discussed.

Research, Development, and Evaluation Commission. The interview with a senior representative of the RDEC revealed the agency's perspective on a number of significant issues

with regard to the testing policy. First, the CEFR Common Reference Levels are considered to be appropriate for civil servants in Taiwan. Second, individual agencies now have the power to decide for themselves what proficiency levels are suitable for their employees. Third, it is recognized that the relatively heavy weighting of English proficiency as a promotion scoring criterion could have negative consequences for more senior employees with heavier workloads and less time to devote to improving their English. However, testing has traditionally been a means of upward mobility for civil servants, and in Taiwanese society as a whole, and tests are therefore likely to remain a feature of the civil service promotion scoring system. Fourth, developing English proficiency is seen as the responsibility of individuals, so the government will not set further targets for either the level of English proficiency that civil servants should achieve, nor the rate at which the percentage of civil servants with English proficiency should increase. Finally, the development of professional skills is of critical importance, and agencies will be encouraged to provide English training that is directly relevant to the needs of the individual agencies and their employees. The RDEC's Plan for Enhancing National English Proficiency (RDEC, 2009) outlines measures that public agencies will adopt to further develop the English proficiency of their employees.

Examination Yuan. The interview with a senior officer of the Examination Yuan, Ministry of Examination focused on a number of critical issues with regard to the testing policy. First, the English proficiency of civil servants is seen as making a major contribution to Taiwan's regional and global competitiveness; however, employees in the public sector are believed to lag behind those in the private sector, and the Ministry has adopted upgrading the English proficiency of civil servants as a key goal. Taiwan's President Ma has reportedly instructed the Ministry to ensure that it adopts measures to ensure that the English proficiency of civil servants

is enhanced. Second, the Ministry has a goal of aligning English proficiency levels with the civil service grades, but has encountered difficulty in gathering sufficient evidence to justify decisions about the appropriate level of English proficiency for civil servants at different ranks and grades. Third, efforts to integrate English proficiency testing within the civil service recruitment examinations have been undertaken, and English is now a required subject in virtually all civil service examinations. The difficulty of developing examinations with sufficient ability to discriminate English proficiency at different levels is significant, but the Ministry is endeavoring to overcome this obstacle. Finally, the current practice of recognizing multiple English proficiency tests that civil servants may take to demonstrate their English proficiency with regard to promotion scoring is seen as a positive move because it allows individuals to determine for themselves which test is most appropriate for them.

Central Personnel Administration. The discussion with a mid-ranking officer of the Central Personnel Administration covered a number of significant issues regarding the English testing policy. First, when English proficiency was adopted as a promotion scoring criterion, the differing needs of the individual agencies were not taken into consideration. In the interest of encouraging civil servants to improve their English proficiency, relatively high promotion scoring values were initially assigned. The potentially disruptive consequences of this action were later recognized, and individual agencies were subsequently empowered to establish promotion scoring values for the different English proficiency levels according to their own needs. Secondly, concern was expressed that the Ministry of Examination's English tests may not adequately assess language proficiency and could overemphasize grammar knowledge and listening ability at the expense of other language skills. Thirdly, the CPA is aware of the need to gather evidence to validate the decision to adopt the CEFR Common Reference Levels as a

proficiency yardstick, but has so far been unable to accomplish this. The strategy of having individual agencies take responsibility for determining suitable English proficiency levels is seen by the CPA as potentially ameliorating the consequences of the adoption of the CEFR proficiency standards. Similarly, the CPA expects that the adoption in March 2011 of new promotion scoring standards will result in substantial benefit. Individual agencies now have the authority to recognize proficiency in different languages, including English, other foreign languages, as well as Taiwan's local languages as a promotion scoring criterion.

Ministry of Education. The conversation with a mid-ranking officer of the MOE covered a number of important issues related to the policy to recognize English proficiency as a promotion scoring criterion for Taiwan's civil servants. First, the Ministry believes that the GEPT could be particularly well suited to testing the English proficiency of English learners in Taiwan since it was developed for that purpose with the support of the MOE. However, several factors are seen as obstacles to the wider adoption of the GEPT as the preferred test of English proficiency by government agencies. According to the informant, these include the limited recognition of GEPT scores by overseas test users and the concern by overseas English proficiency test developers that their tests might be excluded from competing in the Taiwanese testing market.

Secondly, the CEFR Common Reference Levels were adopted as a language proficiency yardstick on the advice of local English teaching experts who pointed out that the CEFR levels had gained wide international acceptance. The MOE acknowledged that little evidence to justify the validity of that decision had been gathered, and expressed concern that the English examination promotion scoring table had not been updated to reflect changes in the availability of the different English tests included on it. Furthermore, it was pointed out by the informant

that the scoring table still currently listed the TOEFL CBT as a recognized English proficiency test, despite the fact that it had been replaced by the TOEFL iBT in 2006. This point highlights an apparent lack of communication between government agencies with shared responsibility for managing the testing policy.

The MOE was authorized to recognize English proficiency tests that had been aligned with the CEFR Common Reference Levels, and it carried out this function in 2005, prior to the creation of the English examination promotion scoring table by the CPA in that same year. However, according to the CPA, the English examination promotion scoring table was withdrawn in 2006, and is no longer considered a relevant document for determining how scores on the various recognized English tests related to the CEFR Common Reference Levels. According to official documents, the Executive Yuan authorized individual agencies to assign promotion scoring values with regard to language proficiency according to their own needs as of March 2011. It is possible that a decision with regard to the recognition of English proficiency tests as a promotion scoring criterion had been made prior to that date, but this current research failed to uncover evidence of such an action. Given that the consequences of the use of language tests are substantial, greater transparency with regard to the testing policy would likely be of value to stakeholders in the testing policy, including test developers, test-takers, test users, and language testing researchers.

Central Bank of China (Taiwan), Ministry of Economic Affairs, and Ministry of Finance. Interviews with mid-ranking representatives of the CBC and MOEA and a high-ranking officer of the MOF were conducted in order to gather information about how the SELPT policy was being implemented in the respective agencies. In the discussions with the officers of the CBC and the MOEA, it was acknowledged that employees in positions in which English

proficiency was less important tended to be less satisfied with the policy of recognizing English proficiency as a promotion scoring criterion. The representative of the MOEA expressed the agency's belief that the inclusion of English proficiency as a subject in civil service recruitment examinations was a positive development that would have the effect of increasing the English ability of civil servants. The MOEA has for many years maintained a foreign language training program for its employees, and it is believed that requiring prospective civil servants to attain English proficiency prior to their recruitment would lessen the need for such a training effort. The CBC, on the other hand, does sponsor English courses for its employees, but these are administered by individual departments on an ad hoc basis.

While it was not possible to interview an MOF employee about the Ministry's overall efforts to encourage its employees to improve their English proficiency, one senior officer did make himself available on two occasions to discuss the use of English proficiency as a promotion scoring criterion. In the first discussion, he expressed his opinion that English proficiency could have a major impact on an employee's promotion prospects. Because civil servants at the same grade level often have nearly identical scores on the basis of the promotion scoring standard, the weight of English proficiency relative to other promotion scoring criteria is a key factor in awarding promotions. He qualified this statement by noting the fact that this would particularly be the case for positions in which there is a greater need for English ability. The decision to award a promotion to a qualified candidate is made in consideration of a number of factors, and the total promotion score is significant in this effort, but it is not the sole determinant. This point demonstrates that individual agencies, and the personnel within them, enjoy a certain measure of leeway in awarding promotions based on the specific needs of the agencies and positions within them.

The researcher was fortunate to have a final opportunity to discuss the testing policy with the MOF officer following the March 2011 adoption of the revised promotion scoring standards that recognized proficiency in languages in addition to English. This individual confirmed that to his knowledge, the Ministry had not significantly altered its valuation of language proficiency as a promotion scoring criterion. English is the main foreign language in which Ministry employees are expected to have adequate proficiency for carrying out their duties. He stated that English is the only foreign language in which proficiency would be considered a promotion scoring criterion. Furthermore, as of the date of the meeting, the promotion scoring value of different levels of English proficiency remained aligned with the values expressed in the English examination promotion scoring table produced by the CPA in 2005. Namely, English proficiency equivalent to the CEFR A2 level retains a value of 2 points; proficiency at the B1 level has a value of 4 points; and proficiency at the B2 level is valued at 5.5 points. This information suggests that while the CPA no longer acknowledges a policy of awarding standardized promotion scoring points for the English proficiency levels aligned with the CEFR Common Reference Levels, this practice remains in effect in at least one of the three agencies that were the focus of this research project.

Stages of the policy cycle. The interviews with the representatives of the government agencies helped to identify the key features of the testing policy cycle. The extended G-S policy analysis framework was used to clarify the roles of the different agencies in the policy cycle. Discussion of the results of the interviews is presented within the context of the six stages of the policy analysis framework. The key elements of the conceptual and institutional framework that comprise the context for the policy are also discussed.

Goals. In the G-S framework, broad-based goals are understood to be products of the first stage of the policy cycle. Goals are ordinarily established at the higher levels of the government hierarchy. In Taiwan, such levels would include the Office of the President and Premier of the Executive Yuan.

The government's policy to recognize English proficiency as a criterion for civil service promotion scoring had as its basis the goals that were expressed during the 1999 to 2001 period. In 1999, the MOE, as a cabinet-level agency of the Executive Yuan, offered its support to the LTTC's development of a test of general English proficiency. English was seen to play a key role in increasing international competitiveness, and it was believed that a national English proficiency test would promote lifelong learning among Taiwanese citizens.

Between 2002 and 2007, the goal of increasing Taiwan's national English proficiency was formalized with its inclusion in the Executive Yuan's Challenge 2008 development plan. The goal of improving English ability was seen to support the wider goal of improving the quality of the workforce. It was during this period that the goal of upgrading the English proficiency of civil servants was adopted. The GEPT was initially seen as a benchmark against which English proficiency could be measured.

Following the end of the Challenge 2008 plan, the government's goals with regard to developing English proficiency shifted their focus. The 2009 PENEP communicated the twin goals of creating an environment for learning English by using it and creating an internationally friendly environment. Within PENEP, plans for improving the English proficiency of civil servants were communicated, but the testing of English proficiency was not explicitly a feature of this effort.

Objectives. Objectives are understood to be “operationalized goals” in the G-S policy analysis framework. While objectives are still goals, and thus a product of higher levels of the government, they are formulated by an agency or agencies at a level subordinate to those which establish broader goals. In the context of the testing policy cycle in Taiwan, the RDEC would take responsibility for coordinating both the establishment of objectives and the implementation of measures to achieve them.

Building on the goals of promoting the development of Taiwan's national English proficiency during the period of 1999 to 2002, the government advocated improving the English ability of the general public and students. Beginning in 2002, the Challenge 2008 plan explicitly called for creating incentives to encourage civil servants to improve their English proficiency. At the same time, efforts were made to remove problems that prevented achievement of the government's goals and to establish an environment that would facilitate the development of English proficiency.

In 2009, specific objectives with regard to developing English proficiency were focused on facilitating learning. These specified the number of civil servants taking part in English training activities. No objectives with regard to English proficiency testing played a role in the PENEP. However, international comparisons based on scores on English proficiency tests were cited as justification for the continued focus on developing Taiwan's national English proficiency as a means of promoting national development and international competitiveness.

Methods/Instruments. The selection of methods/instruments for achieving policy objectives occurs at a level of government subordinate to that at which objectives are established. In the context of the policy cycle in Taiwan with regard to the policy of encouraging civil

servants to improve their English, these agencies included the Examination Yuan's MOEX and MOCS in a consultative role, the MOE, and the CPA.

The MOE's 1999 decision to support the LTTC's effort to create a national English proficiency test was one of the first indications that the government had selected the methods that it would use to meet its objectives of encouraging civil servants to upgrade their English ability. From 2002 to 2004, additional measures were selected, including the use of courses and websites that would train civil servants to improve their English proficiency. The MOCS and the CPA would oversee these efforts. The use of promotion scoring standards based on a criterion of English proficiency was also selected as a method for achieving the government's objective.

In 2009, added emphasis was placed on the use of English training courses to improve civil servants' English language proficiency. In that same year, the inclusion of English as a subject on virtually all civil service recruitment examinations was selected as a method for achieving the goals of improving civil servants' English proficiency.

Activities. Within the G-S policy analysis framework, activities are undertaken within the context of the methods/instruments selected at the previous level of the policy cycle. The key activities within this stage of the policy cycle occurred over an eleven-year period, beginning with the LTTC's introduction of the GEPT Intermediate level test in 2000 and followed by the launch of the Elementary and High-Intermediate levels in 2001.

In 2002, the MOE recognized the GEPT Elementary as the basic proficiency level for civil servants in Taiwan and began to actively encourage them to pass it. In 2004, the CPA adopted a standard promotion score of 2 points for civil servants who had passed the GEPT Elementary. In that same year, annual targets were set for the percentage of civil servants who were expected to have passed this test. For 2005, 2006, and 2007, the targets were set at 10%,

30%, and 50%, respectively. Additionally, administrative agencies were authorized to subsidize test fees for civil servants who passed an English proficiency test.

In 2005, the MOE adopted the CEFR Common Reference Levels and recognized a number of tests that had demonstrated that they were aligned to these levels. The CPA then assigned standard promotion scoring values for the different CEFR levels, awarding 2 and 4 points for the civil servants who passed a recognized English test at the A2 and B1 levels, respectively. Additionally, administrative agencies were authorized to set their own promotion scoring values for civil servants who achieved the CEFR B2 and C1 proficiency levels. The 2006 and 2007 targets were lowered from 30% and 50% to 12% and 18%, respectively. Finally, that year also saw the establishment of annual incremental growth targets of 5% and 10% based on agencies' achievement of the previous year's target.

In 2006, the CPA directed administrative agencies to set the promotion scoring values linked to the CEFR proficiency levels themselves. Furthermore, a program to commend the subordinate units within administrative agencies that had achieved the CPA's annual English proficiency targets was initiated. In that same year, the practice of subsidizing test fees was also extended to include all civil servants who had taken an English proficiency test.

In 2008, the targets instituted under Challenge 2008 were allowed to expire. However, Administrative agencies apparently maintained the promotion scoring values that had been linked to the CEFR Common Reference Levels. Then in 2009, an English subject examination began to be phased into nearly all civil service recruitment exams. The development was completed by 2011. In that year, the promotion scoring standards were revised to remove specific reference to English language proficiency, and a more general "language proficiency"

criterion was included instead. Administrative agencies were directed to assign promotion scoring values for language proficiency based on their specific needs.

Performances. The fifth stage of the policy cycle in the G-S policy analysis framework is performances, and this stage of the analysis focuses on the behavior that the testing policy was designed to produce. Again, starting from the introduction of the GEPT in 2000, it was apparent that English proficiency testing was a widely accepted practice in Taiwan. In just the first two years of the GEPT's administration, the number of test-takers exceeded 250,000. In 2002, it was estimated that just 1 to 2% of Taiwan's approximately 230,000 civil servants had some English ability (Personal communication with CPA official, March 2011).

By 2004, that number was estimated to have risen to 10%, with approximately 6% having passed the GEPT at or above the Elementary level. Following the adoption of English proficiency as a promotion scoring criterion in 2004, the 2005 figure remained at approximately the same level. In 2005, as mentioned above, the targets for 2006 and 2007 were lowered to 12% and 18%, respectively. According to CPA data, 12% of civil servants achieved English proficiency at or above the CEFR A2 level in 2006; and in 2007, this figure had risen to 18.9%, exceeding the target set in 2005. The most recent figure available indicates that 25% of civil servants had demonstrated English proficiency at the A2 level or above by the end of 2010. In fact, the CPA offers figures that demonstrate 18% of civil servants have achieved English proficiency at the A2 level, 4.5% have reached the B1 level, and 2.1% the B2 level. These figures apply to all permanent civil servants, excepting teachers and soldiers, at the Elementary, Junior, and Senior ranks.

In comparison, MOEA data about its employees reveals a different scenario. The MOEA maintains a workforce of 4,506 employees working in 29 different staff units and administrative

agencies. As of December 2010, the total number of individuals who earned a mark on an English proficiency test at or above the CEFR A2 level was 356, accounting for 7.88% of the total. According to this data, 2.57% of MOEA employees have demonstrated English proficiency at the CEFR A2 level; 2.24% have reached the B1 level; 2.59% are at the B2 level; and 0.48% the C1 level. Of those MOEA employees who have passed an English proficiency test, 32.58% achieved the A2 level, 28.37% reached the B1 level, 32.86% are at the B2 level, and 6.17% attained the C1 level.

The MOEA data also includes figures specifying which English proficiency tests had been taken by those of its employees who earned marks at the A2 level or above. The data reveals that 40.73% of those with English proficiency passed the FLPT, 35.39% passed the GEPT, 16.85% passed TOEFL CBT, and a further 7.02% passed other recognized English proficiency tests. Please see Appendix H for the complete data, including the percentage of employees who have passed an English proficiency test in each of the MOEA staff units and administrative agencies.

Evaluation. The sixth and final stage of the policy cycle represented in the G-S policy analysis framework is evaluation. In this stage of the cycle, performances of the policy stakeholders are evaluated. As this research of the testing policy is being conducted externally, i.e. carried out by a researcher without full access to the data collected by the various agencies, the results of this stage of the analysis are necessarily limited. However, conclusions can be drawn from a review of the documents provided by the different stakeholding agencies and revisions to the policy measures that occurred throughout the policy cycle.

First, the fact that the targets for the percentage of civil servants with English proficiency were lowered in 2005 is a significant indication that the initial difficulties encountered in

meeting the 2004 targets resulted in a reappraisal of the ability of the testing policy to encourage civil servants to improve their English. The new targets established in 2005 were met, however, and this would suggest that the testing policy was effective in achieving its revised goals. The continued increase in the percentage of civil servants in the following years demonstrates that the policy did produce satisfactory results in terms of quantitative measures. It should be noted that the CPA requested that administrative agencies submit six-monthly reports detailing their performance outcomes.

Second, the maintenance of the initial target for English proficiency level at CEFR A2 suggests that this level was considered adequate for civil servants. The decentralization of the authority to set target proficiency levels and promotion scoring values from the CPA to the administrative agencies suggests that standard proficiency levels and promotion scoring values were judged to be unsuitable for the civil service as a whole. The decision to authorize individual administrative agencies to determine these factors from 2006 onwards was likely considered to be an effective action judging from the fact that there was no reversal of this decision in the following years.

Third, the focus on facilitating the development of civil servants' English proficiency through training programs rather than a reliance on the targets associated with English proficiency tests suggests that shortcoming may have been identified with the test-centered approach launched in 2002. The adoption of the CEFR proficiency levels in 2005 and the publication of the English examination promotion scoring table suggest that the initial reliance on proficiency levels based on the GEPT may not have met the expectations of the higher-level, policy-making agencies. The subsequent withdrawal of the English examination promotion scoring table in 2006 may imply that the 2005 action was considered inadvisable. However, that

the administrative agencies seem to have maintained their use of the CEFR levels following the 2006 action suggests that those proficiency levels were of practical value, and with no ready alternative to them, their use has been sustained.

Finally, the 2009 policy to incorporate English subject examinations into the civil service recruitment exams suggests that English testing is still considered to be of value, both for determining the English proficiency level of civil servants and encouraging civil servants to improve their English ability. The examination system is an integral component of Taiwan's civil service system, and it is extremely likely that this system will continue to play a role in the future. It is apparent that the power of testing, both of language and professional skills and knowledge, to produce both positive and negative consequences is acknowledged, and future revisions to the elements of the civil service testing system will be implemented with the intention of producing more satisfactory results.

Extended G-S policy analysis framework. The extended G-S policy analysis framework incorporates two additional frameworks that are intended to contribute to an understanding of how policies will produce the intended outcomes. The conceptual framework introduces a systematic way of including factors such as theories and assumptions, definitions, ideology (norms and values), and attitudes and behaviors. The institutional framework includes the political, social, and economic setting; the institutional and legal setting; and stakeholders. The inclusion of these frameworks in the analysis is intended to overcome one of the weaknesses of using models for policy analysis: models may overly simplify and neglect crucial factors. These additional frameworks could help to explain why the actual effects of a policy differ from the intended effects.

Conceptual framework. This framework is intended to account for the theories and assumptions, influenced by geography, culture, and history, upon which policy makers based their policies. It would not be possible in a research project to include all such assumptions, but it is certainly practical to include several such factors that could have an influence on the consequences of the testing policy. Three such factors will here be explained, and their possible effect on policy outcomes will be discussed.

First, assumptions about the process of policy making and implementation influence the consequences of the testing policy. The policy process is driven by goals that are established at the highest levels of the government. These goals serve as development targets within the policy process, and there is an assumption by policy makers that that net benefit will result when the targets are met. The Challenge 2008 national development plan is one in a long line of such plans, and Taiwan's government rigorously sets out medium- and long-term development goals and objectives, identifies performance benchmarks, allocates budgets, and evaluates performance. Policy makers take for granted that policy implementers will accept that policies are reasonable and goals are attainable. Furthermore, it is assumed that incentives will sufficiently motivate their intended targets, and that if unintended consequences result, these may be rectified through revisions to the policy. This view is supported by the evidence that was uncovered during this research, including the downward revision of the targets for the English proficiency of civil servants, the adoption and later withdrawal of the English examination comparison table, and the decentralization of proficiency scoring standards from the CPA to individual agencies.

Secondly, assumptions and beliefs about the value of English proficiency also have a fundamental influence on the formation of goals and objectives, the selection of methods and activities, and the evaluation of performances. English ability is seen as useful for enhancing the

effectiveness of public institutions in an absolute sense, and also relative to other local agencies and to foreign governments. It is commonly acknowledged in Taiwan that English proficiency is linked to the competitiveness of Singapore and Hong Kong, and Taiwan's government seeks to promote national development through enhancing the English language environment in Taiwan. In addition, the English language has a high status in Taiwan, and it is seen as both a signal of elite status for individuals, as well as a channel through which one may acquire such status. English is also associated with academic achievement and individual merit, and these characteristics are understood to contribute to national development.

Thirdly, testing is assumed to be a fair and unbiased method of identifying those who have distinguished themselves through meritorious achievement. This belief is certainly related to the history of the use of competitive examinations in China and neighboring Asian regions. It is difficult to overestimate the influence that examinations have had on Asian cultures. The use of examinations by powerful institutions to select qualified individuals is almost unquestioned in Taiwan, and is the primary basis for access to higher education and vocational opportunity in the government. While the potential difficulties of developing and administering appropriate examinations are recognized, and there have even been moves within the academic domain to adopt non-testing methods for granting access to worthy individuals, it is nearly inconceivable that examinations will cease to be employed. More likely is that examination systems will be revised in order to minimize unintended consequences. This appears to be the case in the use of English proficiency examinations as a promotion scoring criterion for civil servants in Taiwan. Even when problems are noted within the examination system, proposed solutions are based on adopting modifications, such as the use of English as a subject on all recruitment exams, rather than the development of non-testing methods.

The conceptual factors introduced above are hypothesized to have an impact on the outcomes attributed to the testing policy analyzed in this research. The testing policy is a feature of language planning in Taiwan, and such planning is associated with the government's involvement in setting goals for national development, its selection of methods used to achieve objectives, and its evaluation of policy consequences. While the discussion of the three conceptual factors may give the impression that the government is monolithic in its approach to policy making with regard to English testing policy, the fact is that there are a great number of separate agencies with their own interests, and these shall be briefly discussed in the next section.

Institutional framework. In the same way that the conceptual framework enlarges the scope of factors that can be considered as having an influence on the testing policy, so the institutional framework brings in varied sources of potential impact on policy consequences. The extended G-S policy framework suggests the inclusion of political, social, economic, and institutional setting; and stakeholders. Thus far, the policy analysis has focused on the role of the various agencies in the cycle of making and implementing policy, the formal institutions that make up one portion of the policy stakeholders. In the following discussion, other institutional factors will be briefly considered so that a greater range of influential factors can be accounted for within the analysis of the testing policy.

Political setting. The policy that has been the focus of this research is not only a product of Taiwan's political system; it is also a policy that is intended to have a direct influence on the institutions of the government. In that sense, the political setting of the policy should be seen as having a significant impact on the formulation of goals and objectives. Challenge 2008 was introduced in 2002, and observers of Taiwan's political development will recognize that this places it firmly within the first DPP-led administration of Taiwan's central government.

Language has long played a role in Taiwan's politics, and while the politics of language in Taiwan have mainly been related to local languages, including Mandarin, Taiwanese Southern Min, Hakka, and the Austronesian languages spoken by Taiwan's indigenous peoples, it must be recognized that English and other foreign languages are also rooted in a political context. The return to power of the Chinese Nationalist Party, the Kuomintang or KMT, in 2008 coincided with the establishment of the Vision 2015 6-year development plan and the Plan for Enhancement of National English Proficiency. While it is certain that the broader political goals of the two parties differ, both the Chen-led DPP administration and Ma's KMT government recognized the importance of English proficiency to Taiwan's global competitiveness. The explicit goals and objectives, those of enhancing Taiwan's national English proficiency and upgrading the English ability of civil servants, did not change appreciably during the transition from one ruling party to the other. The alterations in the methods and activities that were used to promote the achievement of development goals could be ascribed more to the experience that was gained in carrying out the policy goals than to the adoption of entirely different goals altogether.

Social setting. Taiwan has a population of over 23 million, and while the approximately 125,000 civil servants who are affected by the English proficiency testing policy make up but a small portion of the total, they are an influential segment of the society. Civil servants enjoy a relatively high status in society, are predominantly well educated, and their actions have a direct bearing on the government's ability to meet the goals it sets for itself. The recruitment of qualified individuals to the civil service is thus of crucial importance, and one entire branch of the government, the Examination Yuan, is dedicated to managing public employees. The agencies that were included in this research are all located in Taipei, the political and economic

capital of Taiwan, and civil servants make up a significant portion of the workforce in Taipei. The influence of the English proficiency testing policy has thus probably played a larger role in this city than in other urban centers of Taiwan. Furthermore, Taipei is the most cosmopolitan of Taiwan's cities, and the trend of globalization within which the use of English in Taiwan is rooted has arguably had a greater impact in Taipei than elsewhere on the island. The use of English in Taiwan is also related to Taiwan's efforts to play a greater role in the international community, and the promotion of the English proficiency of civil servants is integral to this. The English proficiency testing policy can thus be seen as occupying a significant position in the Taiwan's social development, both nationally and internationally.

Economic setting. Taiwan's economic development is often seen as playing an influential role in many features of policy making in Taiwan, with relevance to social and political issues as much as economic ones. One rationale for upgrading Taiwan's workforce, a goal of the Challenge 2008 plan, is enhancing the nation's ability to develop its economy. In this era of globalization, the ability of Taiwan's government agencies to effectively communicate with foreign businesses, investors, and professionals living in Taiwan or abroad is an important factor motivating policy makers to encourage civil servants to upgrade their English skills. The economic impact of the incentives offered to civil servants is also significant. Those government employees who attain promotions will see their incomes rise, and this will directly benefit the civil servants themselves and their family members. One of the chief attractions of government employment, a career that generally lasts the entire working life of civil servants, is the stability it offers. The use of English proficiency as a promotion scoring criterion thus has an impact on the nation's ability to improve its economy and on individuals' own financial security.

One final feature of the economic setting for the testing policy is the fact that the English proficiency tests developed by foreign testing agencies are commercial products that bring profit not just to the overseas testing boards, but also to their local agents. A policy that financially benefits one or another testing institution will result in controversy, and regardless of the advantages that a policy might have for the government, these consequences must be taken into consideration. In the early stages of the civil service English testing policy, the government encouraged its employees to demonstrate their English proficiency by taking the GEPT. Then under the influence of comments from foreign test developers who may have felt they were being unfairly excluded from a lucrative market, the government adopted a scheme that recognized a number of English tests based on their alignment with the CEFR proficiency levels. While the economic impetus for this decision may not have been the sole determinant, it likely played a significant role. Thus, the economic setting, like the political and social settings, had an important effect on the elements of the English proficiency testing policy and the policy's eventual outcome.

Institutional setting. The roles played by the various institutions that have a stake in the testing policy have previously been discussed, but some mention will here be made of the relations between these agencies and the structure of their workforces. These agencies may be divided into three groups based on their role in the policy process. The first group includes the RDEC, the CPA, and the MOE. The RDEC coordinates policy measures to be carried out by the various agencies that are subordinate to the Executive Yuan. The CPA has responsibility for policies relating to the management of personnel within the agencies subordinate to the Executive Yuan. Its staff directly supervises the personnel departments within the agencies that make up this branch of the government. This institutional framework results in a uniformity of

personnel policy and compliance with central directives. The Ministry of Education is responsible for matters related to education, and thus was responsible for overseeing the selection of the CEFR Common Reference Levels as an English proficiency yardstick against which English proficiency tests would be aligned. The publication of the English Examination Comparison Table was carried out by the CPA on the basis of the recommendation of the Ministry of Education. These agencies all participated in formulating objectives and selecting measures.

The second group includes the Ministry of Examination and the Ministry of Civil Service, which are subordinate to the Examination Yuan. These agencies were not directly involved with formulating or carrying out the testing policy, but they participated in a consultative role with regard to the policy's formulation. Clearly, testing and education are related fields, as are personnel management and civil service testing. The MOEX is in charge of developing and administering the civil service recruitment and promotion examinations, but it was not directly responsible for the formulation of the scheme for recognizing the English proficiency examinations. Similarly, the role of the MOCS overlaps with the duties of the CPA, but the former agency administers personnel matters for employees within agencies not subordinate to the Executive Yuan, which include the Judicial Yuan, the Examination Yuan, the Control Yuan, and the Legislative Yuan.

The third group includes those agencies that had the task of implementing the testing policy and were responsible for achieving the goals set for them by the RDEC and CPA with reference to the methods selected by the Ministry of Education under consultation with the related agencies. The members of this group include the CBC, the MOF, and the MOEA. These latter two ministries are directly subordinate to the Executive Yuan, and their chief ministers are

cabinet members within the Executive Yuan. The CBC is nominally an independent cabinet-level agency whose Board of Directors and Board of Governors include the Minister of Economic Affairs and the Minister of Finance, along with representatives nominated by the Executive Yuan and approved by the president. Within these three separate agencies, different methods are employed to achieve similar objectives with regard to policies related to personnel development, including the English testing policy. While all three agencies are concerned with economics and financial matters, their specific responsibilities are separate and they employ staff with various specialties.

Conclusion

In this chapter, the results of this multi-method study have been presented and discussed. The questionnaire produced a wealth of data about the 282 employees of the CBC, the MOF, and the MOEA who participated in the study. With information about the English proficiency of these civil servants, their need for English in their jobs, their experience with English proficiency tests, and their English study, the impact of the English testing policy on employees of the three agencies could be investigated to an unprecedented degree. The participants' responses to the items in the test impression and policy effectiveness sections of the questionnaire helped to evaluate the validity of hypotheses about the testing policy's impact on the English proficiency of civil servants, their motivation to continue studying English, and their satisfaction with the use of scores on English tests as incentives to promote the development of English language skills.

Chapter 6

Conclusion

General Aims of the Study

This aim of this thesis was to answer the three research questions about the use of English language proficiency tests as a promotion scoring criterion for Taiwanese civil servants.

Implications of the Study

The first research question asked, “What are the impacts of the policy encouraging civil servants to pass an English language proficiency test in order to qualify for promotion?” It was hypothesized that the policy would influence civil servants’ English proficiency, their study of English, the use of English at work, and their interest in and motivation for improving their English. The results of the questionnaire indicate that the participants were more likely to study English after taking an English test than before taking a test, and also that just over one-third had attended a test-preparation course prior to taking an English test. The fact that more participants reported studying English after taking a proficiency test than before suggests the test policy motivated civil servants to study. However, while approximately two-thirds studied after taking a test, fewer than one-half were currently studying when they completed the questionnaire. This may imply that the testing policy’s influence on studying had only a temporary effect. Overall, the majority of participants were of the opinion that studying to take an English proficiency test would have a positive impact on their English ability, and that the policy motivated them to improve their English.

The results also revealed that approximately two-thirds of the participants had taken at least one English proficiency test. This figure is much higher than the either the 25% figure reported by the CPA or the 7.88% acknowledged by the MOEA. Just over one-half of those who

reported taking an English proficiency test cited work as the major reason for this, and this suggests that the testing policy has had an impact on the number of civil servants who have taken an English test. Of those who have not taken an English test, approximately one-third reported that they were not yet ready to do so, implying that they could be expected to take a test in the future. Another 40% of those who have not taken a test report that taking a test is unnecessary, and it is assumed that these participants work in jobs in which English proficiency is of little importance. Given that over 80% of the participants reported that test anxiety would influence their scores on an English test, nervousness might also play a role in limiting the number of participants who have taken a test.

The second research question asked, "Does this policy achieve the aims for which it was created?" As the preceding discussion suggests, the policy did influence the participants to study English, and most of the respondents said that studying for a proficiency test would result in an improvement of their English ability. On these bases, the policy did achieve its aims. The testing policy not only influenced the participants to upgrade their English, but it also motivated them to continue studying after taking a test. It is likely that this motivation was related to the incentives associated with the promotion scoring points that civil servants can receive if the score they receive on one of the recognized English proficiency tests is at the CEFR A2 level or above. However, most of the participants reported that the testing policy did not increase their personal interest in studying English. It should also be noted that most participants did not agree that improving their English proficiency would have a positive impact on their job skills. This result is somewhat surprising given the fact that approximately half of the participants said that English was of some importance to their job. It may be that while studying for an English proficiency test does improve civil servants' English ability, this ability is not directly applicable to the use

of English for their jobs. This could suggest that the content of the English proficiency tests taken by the civil servants is not highly correlated to the English skills that they need for their work. Overall, the results of the questionnaire indicate that the testing policy did increase the number of civil servants who have taken an English proficiency test, and that preparing to take an English test does have a positive influence on English ability, but questions remain as to the value of the increase in English proficiency as measured by the English tests relative to the English skills that civil servants believe to be useful for their work.

The answer to the final research question, “What factors have a bearing on the achievement of the SELPT policy aims?” may be arrived at by examining both the questionnaire results and the data gathered through the interviews with the representatives of relevant government agencies. It was hypothesized that the participants’ demographics, the need for English in different agencies and positions, the tests that were recognized, the participants’ impressions of the English proficiency tests and their interest in learning English, and finally, the incentives associated with the promotion scoring points would all potentially have an impact on the achievement of the policy’s aims. As discussed previously, the testing policy resulted in an increase in the number of participants who took an English proficiency test, and it is concluded that the promotion scoring incentives played a major role in this outcome based on the participants’ responses to the question about the appropriateness of the incentives to motivate them to study. The value of the promotion scores associated with the proficiency levels achieved by civil servants is likely to have had a significant influence on the willingness of civil servants to prepare for and take the recognized English proficiency tests. The fact that the CPA introduced standard promotion scoring values for the different CEFR proficiency levels in 2005 indicates that policy makers likely also attributed importance to these values. However, the fact

that the standard values were later rescinded and the authority to set those values was granted to the administrative agencies suggests that the optimum values may differ according to the conditions within the separate agencies. The questionnaire results revealed that a majority of the participants felt that the testing policy was unfair, and it may be reasonable to infer that dissatisfaction with the promotion scoring values was a significant factor. Comments made by the participants in response to the qualitative questionnaire item stated that some government employees devoted an unreasonable proportion of their work time to improving their English ability and neglected their other duties. It may be inferred that promotion scoring values that are set excessively high relative to other criteria may reduce the positive consequences associated with using English proficiency tests as incentives to motivate civil servants to improve their English ability.

Another factor that is likely to have a significant impact on the achievement of the testing policy's aims is the actual need for English in different agencies and for different positions. Promotion scores are not the only criteria for awarding a civil servant a promotion, and if English proficiency has a relatively low value for a particular job, higher English proficiency may do little to influence the likelihood of a civil servant qualifying for a promotion. The government employees who took part in this study work in fields related to economics and finance, and their education and professional abilities are likely to be the key determinants of their suitability for their jobs. Promotions are determined by evaluating a number of factors, and if English is of relatively low importance, it is unlikely that English ability will have a great bearing on promotion opportunities. This suggests that the promotion scoring values associated with English proficiency are best determined by the individual agencies rather than imposing standard values on all agencies. In fact, this is the current practice, and the promotion scoring

values associated with language proficiency, for English as well as other foreign and local languages, are now set by the different administrative agencies according to their own needs. From this perspective, the evolution of the testing policy can be viewed in a positive light. In those agencies with a higher or lower need for English, the promotion scoring values for English proficiency can be set at an appropriate level to act as a reasonable incentive for the employees of those agencies. It is expected that this approach should minimize the potentially negative consequences of establishing a standard policy for all agencies regardless of their actual need for English.

A further development that may have had an influence on the impact associated with the testing policy was the 2005 publication of the English examination comparison table following the adoption of the CEFR Common Reference Levels as an English proficiency yardstick. The use of the table suggested an equivalency among the recognized English proficiency tests that probably had an influence on the civil servants' choice of the test that they would take in order to qualify for promotion points. As demonstrated in the questionnaire results, the majority of the participants indicated that they would select an English proficiency test based on their expectation of its difficulty. As a consequence, civil servants would likely be motivated to choose a test that they felt would be the easiest to pass rather than the one that might test those English skills that would be the most useful for their work. The adoption of proficiency levels linked to the CEFR was a practical solution to the challenges that the government faced when it was implementing the testing policy. The subsequent recognition of those English proficiency tests that could demonstrate their alignment with the CEFR levels appeared to offer advantages, but the consequences of that decision may have resulted in other problems. In the five years since the examination comparison table was recalled, there appears to have been little progress in

sufficiently addressing the issue of test score equivalency. In fact, this is an issue which requires significant research in order to address appropriately, but the practical need for the use of English tests to measure the English proficiency of civil servants is ongoing.

It would seem that Taiwan's government has decided that rather than carrying out costly, time-consuming research that may not produce convincing evidence of the greater value of one test over another, it has decided to pursue an alternative approach. The decision to include English as a subject in all civil service recruitment examinations will lessen the need for external tests of the English proficiency of civil servants. With all incoming civil servants being expected to demonstrate achievement of a predetermined level of English proficiency, the importance of English proficiency as a promotion scoring criterion may be lessened. It is probable that individual agencies will continue to make an effort to assess the English ability of their employees with relation to their actual need for English, but the use of English proficiency tests that was encouraged in response to the goals associated with Challenge 2008 may decline in importance. The next challenge for the government in relation to the testing of the English proficiency of civil servants may be to gather evidence of the validity of the English subject test developed and administered by the Ministry of Examination.

Limitations of the Study

An attempt was made in the conduct of this study to collect information that would reveal the consequences of the policy of encouraging Taiwanese civil servants to improve their English proficiency by recognizing scores on English tests as a promotion criterion. It was decided in the planning of this research to distribute the questionnaire to employees of government agencies that deal with finance and economics in which it was expected that there would be a relatively high need for English due to the international nature of their work. While this approach offered

the advantage of producing relatively rich data about the consequences of the testing policy with respect to a narrow range of civil servants, the results may not be representative of the wider population. If the questionnaire had been distributed to employees of government agencies dealing with other fields, such as health care, education, or law enforcement, or those working in local government offices outside of Taiwan's capital of Taipei, the results may have differed significantly.

Another limitation of this study was the relatively unsophisticated statistical analysis carried out on the questionnaire results. This shortcoming is partly responsible for the lack of association of demographic features of the study population with trends in the results. It was hypothesized that younger participants, those with better English proficiency, and those working in jobs that required English would respond more positively to the testing policy than older participants and those with more limited English proficiency, and also that those in positions that had less need for English would be relatively less willing to commit to taking English proficiency tests in order to increase their promotion scores. Although preliminary analysis suggested that this hypothesis might have been confirmed, and several of the informants expressed their belief that these factors may have been significant in influencing the participants' response to the policy, the difficulties of conducting valid statistical analysis precluded the confirmation of these hypotheses.

A further limitation of this study was the lack of substantial qualitative data, or more specific quantitative data, from the respondents to the questionnaire. It would have been advantageous to invite respondents to the questionnaire to participate in interviews which could have produced additional data about the reasons behind their responses to items on the questionnaire. For example, a majority of the participants reported that the policy was not fair,

but their reasons for this opinion may only be speculated about. Similarly, the questionnaire items about English study were few in number, and this limits the inferences that can be made about the washback resulting from the testing policy.

Future Research Directions

The present study has identified several consequences of using scores on English proficiency tests as a promotion scoring criterion, including an increase in the percentage of civil servants who have taken English proficiency tests, increased willingness to study English, and increased confidence in using English in the workplace. The results also suggest that factors such as the adoption of the CEFR proficiency levels, the recognition of multiple English proficiency tests, the weight of the promotion scores, and the civil servants' estimation of the fairness of the testing policy affect the consequences that the test policy produces. While the results of this study are significant, further research is needed to investigate these topics in greater detail and also to explore issues that may enhance understanding of the English proficiency testing policy.

The results of the current study suggest that the weighting of promotion scoring values has a significant effect as an incentive to motivate civil servants to improve their English proficiency. Additional research could be carried out on the promotion scoring values to measure the effect of this incentive system on civil servants. The values associated with different proficiency levels are now determined by individual agencies, and it could be productive to investigate the relationship between the weighting of English proficiency as a promotion scoring criterion and the need for English at different civil service ranks and grades and within specific agencies and departments. The results of an English needs analysis could help to determine not only the appropriate weighting of promotion scoring values, but may also

identify the language skills that civil servants could develop to benefit the agencies in which they work. Such studies could utilize interviews of civil servants in both supervisory and non-supervisory positions, as well as observations of government employees as they carry out their duties.

The suitability of the CEFR proficiency levels is another subject in which further research is needed. Initial research into this issue could focus on validation of the adoption of the CEFR A2 level as the basic proficiency level for civil servants. Additional research might investigate whether it would be appropriate to associate the different CEFR proficiency levels with different ranks or grades within the civil service system. A study with this aim would also benefit the Ministry of Examination's efforts to develop and administer English subject tests as part of the civil service recruitment examinations.

A comparison of the different English tests that are recognized on their basis of their alignment with the CEFR Common Reference Levels may also be of value. Such a study might investigate the relation of test content, task types, and required cognitive abilities to the proficiency levels that civil servants at different grades and ranks are expected to develop.

References

- Alderson, J.C. (2007). The CEFR and the need for more research. *Modern Language Journal* 91(4), 659-663.
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics* 14(2), 115-129.
- Bachman, L. (2005). Building and supporting a case for test use. *Language Assessment Quarterly* 2(1), 1-34.
- . (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Baldauf Jr., R.B. et al. (2007). Successes and failures in language planning for European languages in Asian Nations. Paper presented at 5th Nitobe Symposium, Tokyo.
- Bandura, A. (1989). Self-regulation of motivation and action through internal standards and goal systems. In A. Pervin, *Goal concepts in personality and social psychology* (pp. 19-85). Hillsdale, NJ: Lawrence Erlbaum Associates.
- . (1991). Self-regulation of motivation through anticipatory and self-reactive mechanisms. In R. A. Dienstbier (Ed.), *Nebraska symposium on motivation 1990: Perspectives on motivation* (pp. 69-164). Lincoln: University of Nebraska Press.
- Boggs, D. L. (1981). Philosophies at issue. In B.W. Kreitlow (Ed.), *Examining controversies in adult education*. San Francisco: Jossey-Bass.
- Business Language Testing Service (BULATS). (2011). BULATS. <http://www.bulats.org/Bulats/Overview.html>.
- Byrnes, H. (2007). Developing national language education policies: Reflections of the CEFR. *Modern Language Journal* 91(4), 679-685.
- Chalhoub-Deville, M. (2009). The intersection of test impact, validation, and educational reform policy. *Annual Review of Applied Linguistics* 29, 118-131.
- Chen, J., Warden, C., & Chang, H.T. (2005). Motivators that do not motivate: The case of Chinese EFL learners and the influence of culture on motivation. *TESOL Quarterly* 39(4), 609-633.
- Cheng, L. (2010). The history of examinations: Why, how, what and whom to select? In Liying Cheng and Andy Curtis (Eds.), *English language assessment and the Chinese learner* (pp. 13-25). New York: Routledge.
- Clement, R., Gardner, R.C., & Smythe, P.C. (1980). Social and individual factors in second language acquisition. *Canadian Journal of Behavioural Science* 12, 293-302.

- Cooper, R.L. (1989). *Language planning and social change*. Cambridge: Cambridge University Press.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: learning, teaching, and assessment*. Cambridge: Cambridge University Press.
- . (2009). *Relating language examinations to the Common European Framework of References for Languages: Learning, teaching, assessment: A Manual*. Strasbourg: Council of Europe.
- Cropley, A. (1980). Lifelong learning and systems of education: an overview. In A. Cropley (Ed.), *Towards a system of lifelong education: Some practical considerations*. Oxford, Pergamon Press.
- Deci, E.L. & Ryan, R.M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.
- Dörnyei, Z. & Otto I. (1998). Motivation in action: A process model of L2 motivation. In *Working papers in applied linguistics 4*. London: Thames Valley University.
- Educational Testing Service (ETS). (2007) *TOEIC examinee handbook*. Princeton: Educational Testing Service.
- . (2011a). TOEFL iBT test content. <http://www.ets.org/toefl/ibt/about/content/>.
- . (2011b). TOEIC home. <http://www.ets.org/toeic>.
- . (2011c). 2010 Taiwan TOEIC results report. <http://www.toeic.com.tw/file/11063020.pdf>.
- Entwistle, N. (1987). Motivation to learn: conceptualizations and practicalities. *British Journal of Educational Studies* 35(2), 129-148.
- Fulcher, G. (2005). Deluded by artifices? The Common European Framework and harmonization. *Language Assessment Quarterly* 1(4), 253-266.
- Fulcher, G. & Davidson, F. (2007). *Language testing and assessment; An advanced resource book*. Oxford: Routledge.
- Gardner, R.C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London: Edward Arnold.
- Gardner, R.C. & Lambert, W.E. (1972). *Attitudes and motivation in second language learning*. Rowley, MA: Newbury.
- Gardner, R.C. & McIntyre, P.D. (1991). A instrumental motivation in language study: Who says it isn't effective? *Studies in Second Language Acquisition* 13, 57-72.

- Hamp-Lyons, L. (2000). Fairness in language testing. In A.J. Kunnan (Ed.), *Studies in Language Testing: Fairness and validation in language assessment*. Cambridge: Cambridge University Press.
- Harlen, W., & Crick, R.D. (2003). Testing and motivation for learning. *Assessment in Education* 10(2), 169-203.
- Hawkey, R. (2006). *Studies in Language Testing 24: Impact theory and practice*. Cambridge: Cambridge University Press.
- Hsu, H.F. (2009). The impact of implementing English proficiency tests as a graduation requirement at Taiwanese universities of technology. (Doctoral dissertation, University of York). Retrieved from etheses.whiterose.ac.uk/576/1/PhD_thesis.pdf.
- Hughes, A. (2003). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Hughes, C. & Tight, M. (1995). The myth of the learning society. *British Journal of Educational Studies* 43(3), 290-304.
- Hyatt, D. & Brooks, G. (2009). Investigating stakeholders' perceptions of IELTS as an entry requirement for higher education in the UK. In *IELTS Research Reports Volume 10*. London: IELTS Australia and British Council.
- International English Language Testing Service (IELTS). (2009). *Guide for stakeholders*. Cambridge: Cambridge ESOL Examinations.
[www.ielts.org/PDF/IELTS Guide for Stakeholders March 2009.pdf](http://www.ielts.org/PDF/IELTS%20Guide%20for%20Stakeholders%20March%202009.pdf).
- Jones, N. (2009). A comparative approach to constructing a multilingual proficiency framework: Constraining the role of standard setting. *Cambridge ESOL Research Notes* 3, 6-9.
- Jones, N. & Saville, N. (2009). European Language Policy: Assessment, learning, and the CEFR. *Annual Review of Applied Linguistics* 29, 51-63.
- Kellaghan, T., Madaus, G., & Raczek, A. (1996). *The use of external examinations to improve student motivation*. Washington, DC: AERA.
- Khalifa, H. & French, A. (2008). Aligning Cambridge ESOL Examinations to the CEFR: Issues and practice. Paper presented at 34th Annual Conference, International Association for Educational Assessment.
- Knowles, M.S. (1980). *The modern practice of adult education*. New York: The Adult Education Company.

- Kunnan, A.J. (2005). "Language assessment from a wider context. In Eli Hinkel (Ed.), *Handbook of research in second language teaching and learning*. London: Lawrence Erlbaum Associates.
- Kunnan, A. J. & Wu, J. (2009). Language Training and Training Center, Taiwan: Past, present and future. In L. Cheng & A. Curtis (Ed.), *Language Assessment for Chinese Learners* (pp. 77-91). Philadelphia, PA: Routledge.
- Language Training and Testing Center (LTTC). (2009). *2009 annual report*.
<http://www.lttc.ntu.edu.tw/Annualreportfiles/2009annualreport.pdf>.
- . (2011a). Foreign Language Proficiency Test.
http://www.lttc.ntu.edu.tw/flpt/Foreign_version/FLPTE.htm.
- . (2011b). Language Testing. http://www.lttc.ntu.edu.tw/E_LTTC/languagetesting.htm.
- . (2011c). GEPT-CEFR alignment.
http://www.lttc.ntu.edu.tw/E_LTTC/E_GEPT/alignment.htm.
- . (2011d). GEPT level descriptors: Superior.
http://www.lttc.ntu.edu.tw/E_LTTC/E_GEPT/superior.htm.
- Little, D. (2007). The Common European Framework of reference for languages: Perspectives on the making of supranational policy. *Modern Language Journal* 91(4), 645-655.
- McNamara, T., and Roever, C. (2006). *Language testing: The social dimension*. Oxford: Blackwell Publishing.
- Medel-Añonuevo, C., Ohsako, T., & Mauch, W. (2001). *Revisiting lifelong learning for the 21st century*. New York: UNESCO Institute for Education.
- Messick, S. (1989). Validity," In R.L. Linn (Ed.), *Educational Measurement* (pp. 13-103). New York: Macmillan/American Council on Education.
- North, B. (2006). The Common European Framework of Reference: Development, theoretical, and practical issues. Paper presented at the Osaka University of Foreign Studies, March 5, 2006.
- . (2007). The CEFR illustrative descriptor scales. *Modern Language Journal* 91(4), 656-659.
- Norton Pierce, B. (1995). Social identity, investment, and language learning. *TESOL Quarterly* 29(1), 9-32.
- Powers, D.E., Kim, H.J. & Weng, V. (2008). The redesigned TOEIC (Listening and Reading) Test: Relations to test-taker perceptions of proficiency in English. Princeton:

Educational Testing Service.

Ross, S.J. (2008). Language testing in Asia: Evolution, innovation, and policy changes. *Language Testing* 25(1), 5-13.

Saville, N. (2003). The process of test development and revision within Cambridge EFL. In C. Weir & M. Milanovic (Eds.), *Continuity and innovation: Revising the Cambridge Proficiency in English Examination 1913-2002*. Cambridge: Cambridge ESOL/Cambridge University Press.

Schouwstra, M. & Ellman, M. (2006). *A new explanatory model for policy analysis and evaluation*. Tinbergen Institute Discussion Paper. Amsterdam: Tinbergen Institute.

Shohamy, E. (1982). Affective considerations in language testing. *The Modern Language Journal* 66(1), 13-17.

Shohamy, E. (2000). The social responsibility of language testers. In R. Cooper, E. Shohamy, & J. Walters (Eds.), *New perspectives in educational language policy*. Philadelphia: John Benjamins Publishing.

Shohamy, E. (2001). *The power of tests. A critical perspective on the uses of language tests*. Harlow, Essex: Pearson Education.

Shohamy, E. (2008). Language policy and language assessment: The relationship. *Language Planning* 9(3), 363-373.

Sloane, F.C. & Kelly, A.E. (2003). Issues in high-stakes testing programs. *Theory into Practice* 42(1), 12-17.

Spolsky, B. (1995). *Measured Words: The development of objective language testing*. Oxford: Oxford University Press.

Suen, H.K. & Yu, L. (2006). Chronic consequences of high-stakes testing? Lessons from the Chinese civil service exam. *Comparative Education Review* 50(1), 46-65.

Taiwan Central Personnel Administration. (2005) English examination promotion scoring table. <http://www.ejob.gov.tw/official/english.htm>.

---. (2009). Statistics: Analysis of Public Servants Manpower (Executive Yuan and Subordinate Administrative Agencies and Schools, 2009 2nd Quarter) <http://www.cpa.gov.tw/%5C/ct.asp?xItem=7276&ctNode=321&mp=10>.

Taiwan Executive Yuan. (2005) Action plan for creating English-friendly living environment. http://www.i-taiwan.nat.gov.tw/en/index.php?option=com_content&view=article&id=142:action-

- plan&catid=31:creating-an-english-friendly-living-environment&Itemid=91.
- Taiwan Government Information Office. (2005). Challenge 2008: The six-year national development plan. <http://www.gio-gov.tw/taiwan-website/4-0a/20020521/2002052101.htm>.
- Taiwan Ministry of Civil Service. (2008). Civil Service Employment Act. http://www.mocs.gov.tw/english/英譯法規/任用法_英_-修3.pdf.
- . (n.d.) Introduction to the civil service legal system of Republic of China. <http://www.mocs.gov.tw/english/lis.asp>.
- Taiwan Ministry of Education. (n.d.) ROC (Taiwan) students in the U.S.A. (1950-2004). <http://english.moe.gov.tw/public/Data/72817222271.jpg>.
- . (2005). E-generation national manpower cultivation plan. <http://english.moe.gov.tw/ct.asp?xItem=7043&ctNode=784&mp=3#I>.
- Taiwan Ministry of Examination. (2008) English to be required as common subject on civil service examinations. <http://www.moex.gov.tw>.
- Taylor, L. (2004). IELTS, Cambridge ESOL examinations and the common European framework. *Cambridge ESOL Research Notes* 18, 2-3.
- . (2005). Washback and impact. *ELT Journal* 59(2), 154-155.
- Tight, M. (1998) Lifelong learning: Opportunity or compulsion? *British Journal of Educational Studies* 46(3), 251-63.
- Tremblay, P.F. & Gardner, R.C. (1995). Expanding the motivation construct in language learning. *Modern Language Journal* 79, 505-18.
- Wall, D. (2005). *Studies in Language Testing 22: The impact of high-stakes examinations on classroom teaching*. Cambridge: Cambridge University Press.
- Weir, C. (2005). *Language testing and validation; An evidence-based approach*. New York: Palgrave and Macmillan.
- . (2005). Limitations of the Common European Framework for developing comparable examinations and tests. *Language Testing* 22, 281.
- Wu, J.R.W. & Wu, R.Y.F. (2010). Relating the GEPT comprehension tests to the CEFR. *Studies in Language Testing 33: Aligning tests with the CEFR: Reflections on using the Council of Europe's draft manual*. Cambridge: Cambridge University Press.

Appendix A

CEFR Common Reference Levels: Global Scale

Proficient User	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.
	C1	Can understand a wide range of demanding, longer texts, and recognize 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.
Independent User	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.
	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 travelling 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.
Basic User	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.
	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.

Note. This table is adapted from Common European Framework for Reference. Council of Europe 2001 http://www.coe.int/t/dg4/linguistic/CADRE_EN.asp.

Appendix B

Civil Service English Examination Scoring Table

Promotion scoring points		2	4	* ^a	*	*
CEF		A2	B1	B2	C1	C2
Common Reference Levels		Waystage	Threshold	Vantage	Effective Operational Proficiency	Mastery
Cambridge Main Suite		KET	PET	FCE	CAE	CPE
BULATS		ALTE 1	ALTE 2	ALTE 3	ALTE 4	ALTE 5
FLPT	Listening / Reading	150	195	240	315	--- ^b
	Speaking	S-1+	S-2	S-2+	S-3	---
GEPT		Elementary	Intermediate	High-intermediate	Advanced	Superior
TOEFL	PBT	390	457	527	560	630
	CBT	90	137	197	220	267
TOEIC		350	550	750	880	950
CSEPT	Level 1	170	230	---	---	---
	Level 2	---	240	330	---	---
IELTS		3	4	5.5	6.5	7.5

Note. This table is adapted from Civil Service English Examination Scoring Table. It was promulgated on September 28, 2005 by the Executive Yuan. It is located on the Internet at <http://www.ejob.gov.tw/official/english.htm>.

^aValue is assigned by individual agencies.

^bTest score is not associated with this CEF level.

[illegible]

英語能力自我評估 (各技能請擇一數字圈選)	極佳	佳	中等	差	極差
1. 聽	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. 說	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. 讀	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. 寫	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

英語在工作上的重要性	非常重要	重要	普通	不太重要	完全不重要
5. 聽	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. 說	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. 讀	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. 寫	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

請說明在您工作上會用到英語的場合或工作內容

9. 請提供您曾報考過的標準化英語檢測之相關資訊；若您曾多次報考過同一測驗，請提供最近一次測驗之相關資訊。若您尚未收到成績通知，則分數欄請留白。(註：標準化英語檢測之測驗種類不包含 GRE)

測驗名稱	測驗	報考原因				成績/分數
		學業	工作/職場	興趣	其他	
托福 <input type="checkbox"/> 網路測驗 <input type="checkbox"/> 紙筆測驗 <input type="checkbox"/> 電腦化測驗		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____分
雅思考試國際英語測驗		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____分
多益英語測驗		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____分
劍橋博思職場英語檢測		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____分
外語能力測驗		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____分
全民英檢(初級)		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> 通過 <input type="checkbox"/> 不通過
全民英檢(中級)		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> 通過 <input type="checkbox"/> 不通過
全民英檢(中高級)		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> 通過 <input type="checkbox"/> 不通過
全民英檢(高級)		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> 通過 <input type="checkbox"/> 不通過
其他_____		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____

10. 我從未報考過任何一種標準化英語檢測

是 ☐ 請提供一項原因_____

否 ☐

11. 我目前有參加英語課程或自我進修 是 ☐ 否 ☐

12. 我在報考標準化英語測驗前就習慣性地學習或練習英語 是 ☐ 否 ☐

13. 我在報考標準化英語測驗前曾經參加過相關的準備課程 是 ☐ 否 ☐

14. 達到升遷標準後我仍舊持續學習或練習英語 是 ☐ 否 ☐

以下請在 1-5 上選擇一項

非常同意	同意	沒意見	不同意	非常不同意
------	----	-----	-----	-------

15. 在報考標準化英語測驗前我就有自信我的
英語能力足以讓我通過一定的成績標準 ☐ ☐ ☐ ☐ ☐

16. 我在報考標準化英語測驗前就已經對該測驗
方式有充分的了解 ☐ ☐ ☐ ☐ ☐

17. 台灣本土開發之標準化英語測驗比國外研發的
測驗更適用於台灣當地的公務人員 ☐ ☐ ☐ ☐ ☐

18. 我認為標準化英語測驗分數，能反映出我實際
的英語能力 ☐ ☐ ☐ ☐ ☐

19. 為達到升遷標準，我會選擇最容易得到高分的
標準化測驗 ☐ ☐ ☐ ☐ ☐

20. 我認為參加標準化英語測驗時的緊張情緒對成
績有負面的影響 ☐ ☐ ☐ ☐ ☐

21. 將標準化英語測驗成績作為升遷參考的政策,影
響我使用或學習英語的方式 ☐ ☐ ☐ ☐ ☐

請簡短說明該政策對你的影響

	非常同意	同意	沒意見	不同意	非常不同意
22. 將標準化英語測驗作為升遷參考的政策，提升了 我個人對英語學習的興趣	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
23. 若標準化英語測驗未被視為升遷標準，我 就不會學習英語	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
24. 我比較希望標準化英語測驗成績不被視為升 遷的標準	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25. 我認為準備標準化英語測驗的同時，我的英 語能力也逐步提升	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
26. 利用標準化英語測驗通過與否作為升遷標準 的政策，能有效提升公務人員增進英語能力的動機	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
27. 利用標準化英語測驗通過與否作為升遷標準對 所有公務人員皆具公平性	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
28. 通過標準化英語測驗能增加我在工作上使 用英語聽、說、讀、寫的自信	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
29. 通過標準化英語測驗對我的工作能力及表 現沒有任何影響	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
30. 整體而言，利用標準化英語測驗作為升遷 標準的政策對於公務人員的專業水準有正面影響	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

若您對以上問題有任何其他的意見或補充說明，請書寫在本頁背面。謝謝您！

[illegible]

Rate your English ability (Check one box for each skill)	Very good 5	Good 4	Not bad 3	Poor 2	Very poor 1
1. Listening	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Speaking	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Reading	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Writing	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Importance of English for job	Very important 5	Important 4	No opinion 3	Not very important 2	No importance 1
5. Listening	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Speaking	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Reading	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Writing	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Please comment on the overall importance of English for your job.

9. Please provide information about the SELPT you have taken. If you have taken the same test more than once, provide information about the most recent attempt. If you are still waiting for your results, leave "Score" blank. (Note: The GRE is not an English test.)

Test name	Reason				Score
	Study	Work	Interest	Other	
TOEFL iBT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____ Points
TOEFL PBT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____ Points
TOEFL CBT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____ Points
IELTS	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____ Points
TOEIC	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____ Points
BULATS	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____ Points
FLPT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____ Points
GEPT Elementary	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> Pass <input type="checkbox"/> Fail
GEPT Intermediate	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> Pass <input type="checkbox"/> Fail
GEPT High-intermediate	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> Pass <input type="checkbox"/> Fail
GEPT Advanced	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> Pass <input type="checkbox"/> Fail
Other: _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	_____

10. I have NEVER taken an SELPT.

Yes ☐ Please give a reason. _____

No ☐

11. I currently attend an English course or study on my own. Yes ☐ No ☐

12. I studied or practiced English in my free time before I began to prepare to take an SELPT as a promotion criterion. Yes ☐ No ☐

13. I attended a test preparation course prior taking an SELPT as a promotion criterion. Yes ☐ No ☐

14. I continued to study or practice English regularly after passing an SELPT as a promotion criterion. Yes ☐ No ☐

Answer the following questions by checking the appropriate box.

15. I was confident that my English proficiency was sufficient to allow me to pass the SELPT before I took it.

Strongly agree ☐ Agree ☐ No opinion ☐ Disagree ☐ Strongly disagree ☐

16. Before I took an SELPT, I had adequate knowledge about its test format and content.

☐ ☐ ☐ ☐ ☐

17. A locally produced SELPT is more appropriate for civil servants in Taiwan than an overseas one.

☐ ☐ ☐ ☐ ☐

18. I believe that the score that I receive on an SELPT accurately reflects my English proficiency.

☐ ☐ ☐ ☐ ☐

19. I would choose an SELPT because it was the easiest one on which to score high marks.

☐ ☐ ☐ ☐ ☐

20. My nervousness while taking the SELPT would have a negative impact on my score.

☐ ☐ ☐ ☐ ☐

21. The SELPT policy has influenced the way I use or study English.

☐ ☐ ☐ ☐ ☐

Please comment on the influence it has had.

	Strongly agree	Agree	No opinion	Disagree	Strongly disagree
22. The SELPT policy has increased my personal interest in learning English.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
23. I would not study English if passing an SELPT were not a promotion criterion.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
24. I would prefer that passing an SELPT were not a criterion for promotion.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25. Preparing to take an SELPT would improve my overall English ability.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
26. The policy to recognize a passing score on an SELPT as a promotion criterion is an effective way to motivate civil servants to improve their English.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
27. The policy to recognize a passing score on an SELPT as a promotion criterion is fair to all civil servants.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
28. Passing an SELPT would increase my confidence in using English in my job.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
29. Passing an SELPT would have no impact on my ability to perform my job.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
30. The SELPT promotion criterion policy has an overall positive impact on the professional standards of civil servants.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

On the back of this page, please provide further comments on any of your above answers.

Include the number for each item you are commenting on. Thank you very much!

Appendix E

Interview Questions

For use with personnel department staff at Central Bank of the Republic of China (Taiwan) and the Ministry of Economic Affairs.

1. What percentage of the employees in this agency have taken a SELPT as a promotion qualification?
本機關公務人員曾報考過標準化英語檢測的比例為何？
2. What percentage of the employees in this agency had taken an SELPT before it was recognized as a promotion qualification?
本機關公務人員在標準化英語檢測被視為陞遷標準之前即報考過相關測驗的比例為何？
3. What are the key features of the policy to recognize SELPTs as a promotion qualification, and how does this policy affect your work?
公務人員英語檢測陞任辦法的為何？該辦法對你工作內容的影響為何？
4. In your opinion, what are the major impacts of the policy on employees of this agency?
你認為公務人員英語檢測陞任辦法的實施對於本機關人員最大的影響為何？
5. How did the employees in this agency react when the policy was introduced?
該辦法剛推出時，本機關人員多半有何反應？
6. How do employees of this agency currently view the policy?
本機關人員目前對該辦法多半持何種看法？
7. How important is English ability to employees of this agency?
英語能力對本機關人員的重要性為何？
8. How influential is passing an SELPT on an employee's chance of promotion?
標準化英語檢測通過與否，對於一般公務人員時晉升有多大影響？
9. Does this agency recommend that its employees take a particular SELPT? Why or why not?
本機關是否有建議公務人員報考某種特定的標準化英語檢測？為什麼(不)？
10. How does your department communicate the details of this policy to the employees of your agency?
貴部門是如何跟機關內部人員討論該辦法的細節？
11. Have there been any problems associated with the policy at this agency?
公務人員英語檢測陞任辦法實施後，是否對本機關造成任何問題？
12. How could the policy be improved?
該辦法有何改善空間？
13. Would you say that the test policy is an effective way to encourage civil servants to improve their English?
你認為英語檢測陞任辦法的實施，是否能有效鼓勵公務人員提升個人的英語能力？
14. Do you think that methods in addition to or instead of a test policy could produce an equal or greater benefit?
你認為是否有其他的辦法，也能帶來相同的效益？

For use with Executive Yuan, Central Personnel Administration staff

1. What are the key features of the policy to recognize standardized English language proficiency tests (SELPT) as a promotion qualification?
公務人員英語檢測陞任辦法的要點為何？
2. What is the CPA's role in the formulation, implementation, and administration of this policy?
行政院人事行政局在該辦法的規劃、執行及監督各方面扮演的角色為何？
3. What percentage of civil servants have taken an SELPT?
公務人員報考過標準化英語檢測的比例為何？
4. What were the initial goals of the policy?
該辦法所訂定的初期目標為何？
5. Has the policy had the expected impact?
該辦法是否已達到預期之成效？
6. What are the major benefits of the policy?
該辦法的主要效益為何？
7. What are its drawbacks?
該辦法的缺點為何？
8. Were there any difficulties in implementing the policy?
該辦法在執行階段時曾遭遇過何種困難？
9. Have the details of the policy been amended since it was first introduced?
該辦法在推行後是否曾修正過相關細節？
10. Are further revisions expected?
未來是否將會推出修訂版本？
11. How is the policy monitored and evaluated?
該辦法如何被監督及評估？
12. What information related to this policy does the CPA provide to other agencies?
行政院人事行政局如何提供各地方機關該辦法的相關資訊？
13. Were any non-testing measures such as regular, annual attendance of English courses considered as alternatives or supplements to recognizing an SELPT as a promotion qualification for civil servants?
除了標準化英語檢測，是否有其他管道，例如每年定期參加英語課程，可用來做為公務人員陞遷的參考依據？
14. Do individual agencies have the option of modifying the terms of the policy to suit their specific needs?
各地方機關是否可以視實際情況而適度修改該辦法的內容，以符合該機關之個別需求？

For use with Ministry of Education staff

1. What was the role of the Ministry of Education in establishing the standardized English language proficiency test (SELPT) policy in regards to its use as a promotion criterion for civil servants?
教育部在公務人員英語檢測陞任辦法的規劃階段中所扮演的角色為何？
2. What were the key steps in the process of formulating this policy?

該辦法規劃的主要步驟為何？

3. What were the key considerations when this policy was being formulated?
該辦法在規劃時的主要考量為何？
4. How were the SELPT selected?
本辦法如何挑選所採認的英語測驗？
5. Why was the CEFR chosen as a standard?
CEF 語言能力參考指標為何被選為參考標準？
6. Is the MOE likely to support the development of a localized foreign language proficiency standard in the future?
教育部日後是否有可能支持研發在地化的語言能力參考標準？
7. Was an effort made to bring a wide variety of viewpoints into discussions on the uses of SELPT as a promotion criterion? Were dissenting opinions raised, and how were these dealt with?
公務人員英語檢測陞任辦法實施前，是否曾廣納各方意見及看法並加以討論？是否出現過反對聲浪？那些異議又是如何解決的？
8. What has been the MOE's role in monitoring and evaluating this policy since it was implemented?
該辦法實施後，教育部在監督及評估方面扮演何種角色？

For use with Executive Yuan, Research, Development, and Evaluation Commission staff

1. Why is it important for civil servants to have the ability to communicate in English?
英語溝通能力對於公務人員的重要性為何？
2. How important is the E-Generation Manpower Cultivation Plan to Taiwan's social and economic development and to its global competitiveness?
E 世代人才培育計畫對於台灣社會經濟發展以及國際競爭力的重要性為何？
3. Did the RDEC have a role in determining that passing a standardized English language proficiency test would be used as a promotion criterion for civil servants? What was the nature of that role?
在公務人員英語檢測陞任辦法規畫時期，行政院研究發展考核委員會是否曾扮演過具有決策性的角色？該角色的本質為何？
4. How well does the standardized English language proficiency test policy support the goals of the Manpower Cultivation Plan?
公務人員英語檢測陞任辦法是否有助於人才培育計畫目標的實現？
5. How does the RDEC monitor the effectiveness of the individual components of the E-Generation Manpower Cultivation Plan, and how is feedback used?
行政院研究發展考核委員會如何監督 E 世代人才培育計畫各項目的成效？如何運用監督評估的結果？

Appendix F

Questionnaire Results

Item A: Employer

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	CBC	107	37.9	37.9	37.9
	MOEA	77	27.3	27.3	65.2
	MOF	98	34.8	34.8	100.0
	Total	282	100.0	100.0	

Item B: Grade

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Elementary A	3	1.1	1.8	1.8
	Elementary B	15	5.3	8.8	10.5
	Junior A	62	22.0	36.3	46.8
	Junior B	81	28.7	47.4	94.2
	Senior	10	3.5	5.8	100.0
	Total	171	60.6	100.0	
Missing	System	111	39.4		
Total		282	100.0		

Item C: Department

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	CBC Foreign exchange	31	11.0	12.3	12.3
	CBC Various	20	7.1	7.9	20.2
	CBC Financial stability	22	7.8	8.7	29.0
	CBC Economic research	29	10.3	11.5	40.5
	MOF Local tax agencies	25	8.9	9.9	50.4
	MOF National level agencies	34	12.1	13.5	63.9
	MOF Customs	23	8.2	9.1	73.0
	MOEA Taiwan Power Company	27	9.6	10.7	83.7
	MOEA Staff units	20	7.1	7.9	91.7
	MOEA Administrative agencies	21	7.4	8.3	100.0
	Total	252	89.4	100.0	
Missing	System	30	10.6		
Total		282	100.0		

Item E: Years of Service

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	28	9.9	10.9	10.9
	2	24	8.5	9.3	20.2
	3	33	11.7	12.8	32.9
	4	13	4.6	5.0	38.0
	5	7	2.5	2.7	40.7
	6	4	1.4	1.6	42.2
	7	6	2.1	2.3	44.6
	8	5	1.8	1.9	46.5
	9	2	.7	.8	47.3
	10	13	4.6	5.0	52.3
	11	3	1.1	1.2	53.5
	12	6	2.1	2.3	55.8
	13	2	.7	.8	56.6
	14	2	.7	.8	57.4
	15	7	2.5	2.7	60.1
	16	2	.7	.8	60.9
	17	7	2.5	2.7	63.6
	18	6	2.1	2.3	65.9
	19	8	2.8	3.1	69.0
	20	11	3.9	4.3	73.3
	21	7	2.5	2.7	76.0
	22	6	2.1	2.3	78.3
	23	9	3.2	3.5	81.8
	24	2	.7	.8	82.6
	25	2	.7	.8	83.3
	26	6	2.1	2.3	85.7
	27	2	.7	.8	86.4
	28	5	1.8	1.9	88.4
	29	6	2.1	2.3	90.7
	30	11	3.9	4.3	95.0
	31	4	1.4	1.6	96.5
	32	3	1.1	1.2	97.7
	33	1	.4	.4	98.1
	35	2	.7	.8	98.8
	37	3	1.1	1.2	100.0
	Total	258	91.5	100.0	
Missing	System	24	8.5		
Total		282	100.0		

Item F: Education level

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	College graduate	20	7.1	7.2	7.2
	University undergraduate	136	48.2	49.3	56.5
	University graduate level	115	40.8	41.7	98.2
	University post-graduate level	5	1.8	1.8	100.0
	Total	276	97.9	100.0	
Missing	System	6	2.1		
Total		282	100.0		

Item G: Major

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Arts and literature	9	3.2	3.3	3.3
	Law/administration/sociology	33	11.7	12.2	15.5
	Finance / economics	151	53.5	55.7	71.2
	Physical science	32	11.3	11.8	83.0
	Other	46	16.3	17.0	100.0
	Total	271	96.1	100.0	
Missing	System	11	3.9		
Total		282	100.0		

Item H: Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	120	42.6	43.3	43.3
	Female	157	55.7	56.7	100.0
	Total	277	98.2	100.0	
Missing	System	5	1.8		
Total		282	100.0		

Item I: Age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	25-29	43	15.2	15.3	15.3
	30-34	67	23.8	23.8	39.1
	35-39	34	12.1	12.1	51.2
	40-44	44	15.6	15.7	66.9
	45-49	37	13.1	13.2	80.1
	50-54	32	11.3	11.4	91.5
	55-59	20	7.1	7.1	98.6
	60+	4	1.4	1.4	100.0
	Total	281	99.6	100.0	
Missing	System	1	.4		
Total		282	100.0		

Section 2: English

Item 1: Self-assessed listening

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very good	10	3.5	3.6	3.6
	Good	81	28.7	28.8	32.4
	Not bad	145	51.4	51.6	84.0
	Poor	43	15.2	15.3	99.3
	Very poor	2	.7	.7	100.0
	Total	281	99.6	100.0	
Missing	System	1	.4		
Total		282	100.0		

Item 2: Self-assessed speaking

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very good	14	5.0	5.0	5.0
	Good	107	37.9	38.1	43.1
	Not bad	127	45.0	45.2	88.3
	Poor	32	11.3	11.4	99.6
	Very poor	1	.4	.4	100.0
	Total	281	99.6	100.0	
Missing	System	1	.4		
Total		282	100.0		

Item 3: Self-assessed reading

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very good	6	2.1	2.1	2.1
	Good	51	18.1	18.1	20.3
	Not bad	147	52.1	52.3	72.6
	Poor	74	26.2	26.3	98.9
	Very poor	3	1.1	1.1	100.0
	Total	281	99.6	100.0	
Missing	System	1	.4		
Total		282	100.0		

Item 4: Self-assessed writing

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very good	14	5.0	5.0	5.0
	Good	92	32.6	32.7	37.7
	Not bad	139	49.3	49.5	87.2
	Poor	35	12.4	12.5	99.6
	Very poor	1	.4	.4	100.0
	Total	281	99.6	100.0	
Missing	System	1	.4		
Total		282	100.0		

Self-assessed average

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	6	2.1	2.1	2.1
	1.25	2	.7	.7	2.8
	1.50	2	.7	.7	3.6
	1.75	5	1.8	1.8	5.3
	2.00	38	13.5	13.5	18.9
	2.25	17	6.0	6.0	24.9
	2.50	35	12.4	12.5	37.4
	2.75	28	9.9	10.0	47.3
	3.00	80	28.4	28.5	75.8
	3.25	19	6.7	6.8	82.6
	3.50	17	6.0	6.0	88.6
	3.75	6	2.1	2.1	90.7
	4.00	24	8.5	8.5	99.3
	4.50	2	.7	.7	100.0
	Total	281	99.6	100.0	
Missing	System	1	.4		
Total		282	100.0		

Item 5: Importance listening

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very important	20	7.1	7.1	7.1
	Important	49	17.4	17.4	24.5
	No opinion	75	26.6	26.6	51.1
	Not important	106	37.6	37.6	88.7
	No importance	32	11.3	11.3	100.0
	Total	282	100.0	100.0	

Item 6: Importance speaking

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very important	21	7.4	7.4	7.4
	Important	53	18.8	18.8	26.2
	No opinion	73	25.9	25.9	52.1
	Not important	104	36.9	36.9	89.0
	No importance	31	11.0	11.0	100.0
	Total	282	100.0	100.0	

Item 7: Importance reading

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very important	17	6.0	6.0	6.0
	Important	41	14.5	14.6	20.6
	No opinion	61	21.6	21.7	42.3
	Not important	112	39.7	39.9	82.2
	No importance	50	17.7	17.8	100.0
	Total	281	99.6	100.0	
Missing	System	1	.4		
Total		282	100.0		

Item 8: Importance writing

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very important	20	7.1	7.1	7.1
	Important	55	19.5	19.5	26.6
	No opinion	80	28.4	28.4	55.0
	Not important	97	34.4	34.4	89.4
	No importance	30	10.6	10.6	100.0
	Total	282	100.0	100.0	

Importance average

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	16	5.7	5.7	5.7
	1.25	3	1.1	1.1	6.7
	1.50	1	.4	.4	7.1
	2.00	35	12.4	12.4	19.5
	2.25	7	2.5	2.5	22.0
	2.50	7	2.5	2.5	24.5
	2.75	6	2.1	2.1	26.6
	3.00	42	14.9	14.9	41.5
	3.25	14	5.0	5.0	46.5
	3.50	20	7.1	7.1	53.5
	3.75	16	5.7	5.7	59.2
	4.00	72	25.5	25.5	84.8
	4.25	8	2.8	2.8	87.6
	4.50	9	3.2	3.2	90.8
	4.75	3	1.1	1.1	91.8
	5.00	23	8.2	8.2	100.0
	Total	282	100.0	100.0	

Item 9a: SELPT taken

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	TOEFL iBT	2	.7	1.1	1.1
	TOEFL PBT	22	7.8	12.1	13.2
	TOEFL CBT	9	3.2	4.9	18.1
	IELTS	7	2.5	3.8	22.0
	TOEIC	89	31.6	48.9	70.9
	BULATS	4	1.4	2.2	73.1
	FLPT	7	2.5	3.8	76.9
	GEPT-E	22	7.8	12.1	89.0
	GEPT-I	13	4.6	7.1	96.2
	GEPT-HI	7	2.5	3.8	100.0
	Total	182	64.5	100.0	
Missing	System	100	35.5		
Total		282	100.0		

Item 9b: Reasons for taking an SELPT

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Study	36	12.8	20.1	20.1
	Work	92	32.6	51.4	71.5
	Interest	29	10.3	16.2	87.7
	Other	6	2.1	3.4	91.1
	Study + work	8	2.8	4.5	95.5
	Work + interest	8	2.8	4.5	100.0
	Total	179	63.5	100.0	
Missing	System	103	36.5		
Total		282	100.0		

CEFR level

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	1	.4	.8	.8
	A2	24	8.5	19.4	20.2
	B1	51	18.1	41.1	61.3
	B2	40	14.2	32.3	93.5
	C1	8	2.8	6.5	100.0
	Total	124	44.0	100.0	
Missing	System	158	56.0		
Total		282	100.0		

Item 10a: I have NEVER taken an SELPT.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	105	37.2	53.8	53.8
	Yes	90	31.9	46.2	100.0
	Total	195	69.1	100.0	
Missing	System	87	30.9		
Total		282	100.0		

Item 10b: Reason for not taking an SELPT

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No need	25	8.9	43.1	43.1
	No interest	6	2.1	10.3	53.4
	Not ready	17	6.0	29.3	82.8
	No time	7	2.5	12.1	94.8
	Too expensive	2	.7	3.4	98.3
	Deaf	1	.4	1.7	100.0
	Total	58	20.6	100.0	
Missing	System	224	79.4		
Total		282	100.0		

Item 11: I currently attend an English course or study on my own.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	147	52.1	53.6	53.6
	Yes	127	45.0	46.4	100.0
	Total	274	97.2	100.0	
Missing	System	8	2.8		
Total		282	100.0		

Item 12: I studied or practiced English in my free time before I began to prepare to take an SELPT as a promotion criterion.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	128	45.4	47.2	47.2
	Yes	143	50.7	52.8	100.0
	Total	271	96.1	100.0	
Missing	System	11	3.9		
Total		282	100.0		

Item 13: I attended a test preparation course prior taking an SELPT as a promotion criterion.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	171	60.6	63.3	63.3
	Yes	99	35.1	36.7	100.0
	Total	270	95.7	100.0	
Missing	System	12	4.3		
Total		282	100.0		

Item 14: I continued to study or practice English regularly after passing an SELPT as a promotion criterion.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	89	31.6	33.3	33.3
	Yes	178	63.1	66.7	100.0
	Total	267	94.7	100.0	
Missing	System	15	5.3		
Total		282	100.0		

Section3: Test impression

Item 15: I was confident that my English proficiency was sufficient to allow me to pass the SELPT before I took it.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	10	3.5	3.6	3.6
	Disagree	70	24.8	25.3	28.9
	No opinion	101	35.8	36.5	65.3
	Agree	93	33.0	33.6	98.9
	Strongly agree	3	1.1	1.1	100.0
	Total	277	98.2	100.0	
Missing	System	5	1.8		
Total		282	100.0		

Item 16: Before I took an SELPT, I had adequate knowledge about its test format and content.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	11	3.9	4.0	4.0
	Disagree	43	15.2	15.5	19.4
	No opinion	83	29.4	29.9	49.3
	Agree	133	47.2	47.8	97.1
	Strongly agree	8	2.8	2.9	100.0
	Total	278	98.6	100.0	
Missing	System	4	1.4		
Total		282	100.0		

Item 17: A locally produced SELPT is more appropriate for civil servants in Taiwan than an overseas one.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	8	2.8	2.9	2.9
	Disagree	39	13.8	14.0	16.8
	No opinion	181	64.2	64.9	81.7
	Agree	48	17.0	17.2	98.9
	Strongly agree	3	1.1	1.1	100.0
	Total	279	98.9	100.0	
Missing	System	3	1.1		
Total		282	100.0		

Item 18: I believe that the score that I receive on an SELPT accurately reflects my English proficiency.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	11	3.9	4.0	4.0
	Disagree	48	17.0	17.3	21.2
	No opinion	111	39.4	39.9	61.2
	Agree	104	36.9	37.4	98.6
	Strongly agree	4	1.4	1.4	100.0
	Total	278	98.6	100.0	
Missing	System	4	1.4		
Total		282	100.0		

Item 19: I would choose an SELPT because it was the easiest one on which to score high marks.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	9	3.2	3.2	3.2
	Disagree	27	9.6	9.7	12.9
	No opinion	80	28.4	28.8	41.7
	Agree	135	47.9	48.6	90.3
	Strongly agree	27	9.6	9.7	100.0
	Total	278	98.6	100.0	
Missing	System	4	1.4		
Total		282	100.0		

Item 20: My nervousness while taking the SELPT would have a negative impact on my score.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	2	.7	.7	.7
	Disagree	19	6.7	6.8	7.5
	No opinion	64	22.7	22.9	30.5
	Agree	164	58.2	58.8	89.2
	Strongly agree	30	10.6	10.8	100.0
	Total	279	98.9	100.0	
Missing	System	3	1.1		
Total		282	100.0		

Item 21: The SELPT policy has influenced the way I use or study English.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	3	1.1	1.1	1.1
	Disagree	41	14.5	14.7	15.8
	No opinion	98	34.8	35.1	50.9
	Agree	117	41.5	41.9	92.8
	Strongly agree	20	7.1	7.2	100.0
	Total	279	98.9	100.0	
Missing	System	3	1.1		
Total		282	100.0		

Section 4: Policy effectiveness

Item 22: The SELPT policy has increased my personal interest in learning English.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	20	7.1	7.4	7.4
	Disagree	98	34.8	36.0	43.4
	No opinion	80	28.4	29.4	72.8
	Agree	69	24.5	25.4	98.2
	Strongly agree	5	1.8	1.8	100.0
	Total	272	96.5	100.0	
Missing	System	10	3.5		
Total		282	100.0		

Item 23: I would not study English if passing an SELPT were not a promotion criterion.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	36	12.8	13.2	13.2
	Disagree	166	58.9	61.0	74.3
	No opinion	50	17.7	18.4	92.6
	Agree	19	6.7	7.0	99.6
	Strongly agree	1	.4	.4	100.0
	Total	272	96.5	100.0	
Missing	System	10	3.5		
Total		282	100.0		

Item 24: I would prefer that passing an SELPT were not a criterion for promotion.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	3	1.1	1.1	1.1
	Disagree	33	11.7	12.3	13.4
	No opinion	107	37.9	39.8	53.2
	Agree	96	34.0	35.7	88.8
	Strongly agree	30	10.6	11.2	100.0
	Total	269	95.4	100.0	
Missing	System	13	4.6		
Total		282	100.0		

Item 25: Preparing to take an SELPT would improve my overall English ability.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	6	2.1	2.2	2.2
	Disagree	20	7.1	7.4	9.6
	No opinion	79	28.0	29.0	38.6
	Agree	155	55.0	57.0	95.6
	Strongly agree	12	4.3	4.4	100.0
	Total	272	96.5	100.0	
Missing	System	10	3.5		
Total		282	100.0		

Item 26: The policy to recognize a passing score on an SELPT as a promotion criterion is an effective way to motivate civil servants to improve their English.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	13	4.6	4.8	4.8
	Disagree	42	14.9	15.4	20.2
	No opinion	66	23.4	24.3	44.5
	Agree	141	50.0	51.8	96.3
	Strongly agree	10	3.5	3.7	100.0
	Total	272	96.5	100.0	
Missing	System	10	3.5		
Total		282	100.0		

Item 27: The policy to recognize a passing score on an SELPT as a promotion criterion is fair to all civil servants.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	30	10.6	11.0	11.0
	Disagree	97	34.4	35.7	46.7
	No opinion	91	32.3	33.5	80.1
	Agree	47	16.7	17.3	97.4
	Strongly agree	7	2.5	2.6	100.0
	Total	272	96.5	100.0	
Missing	System	10	3.5		
Total		282	100.0		

Item 28: Passing an SELPT would increase my confidence in using English in my job.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	13	4.6	4.8	4.8
	Disagree	45	16.0	16.6	21.4
	No opinion	101	35.8	37.3	58.7
	Agree	104	36.9	38.4	97.0
	Strongly agree	8	2.8	3.0	100.0
	Total	271	96.1	100.0	
Missing	System	11	3.9		
Total		282	100.0		

Item 29: Passing an SELPT would have no impact on my ability to perform my job.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	5	1.8	1.8	1.8
	Disagree	44	15.6	16.2	18.0
	No opinion	108	38.3	39.7	57.7
	Agree	92	32.6	33.8	91.5
	Strongly agree	23	8.2	8.5	100.0
	Total	272	96.5	100.0	
Missing	System	10	3.5		
Total		282	100.0		

Item 30: The SELPT promotion criterion policy has an overall positive impact on the professional standards of civil servants.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly disagree	19	6.7	7.0	7.0
	Disagree	29	10.3	10.7	17.6
	No opinion	104	36.9	38.2	55.9
	Agree	109	38.7	40.1	96.0
	Strongly agree	11	3.9	4.0	100.0
	Total	272	96.5	100.0	
Missing	System	10	3.5		
Total		282	100.0		

Appendix G

Pearson Correlation Among Test Impression and Policy Effectiveness Variables

	KNO	LOC	ACC	ESY	NRV	INF	INT	NOT	PRF	ABL	MOT	FAR	CON	JBK	IMP
PRC	.374**	.078	.076	.029	-.075	-.036	.161**	.019	.188**	.224**	.110	.256**	.181**	-.018	.284**
KNO	--	.198**	.248**	.160**	.096	-.055	.200**	-.014	.216**	.248**	.235**	.217**	.236**	.055	.242**
LOC		--	.357*	.187**	.168**	-.055	.237**	-.107	.103	.224**	.237**	.225**	.301**	.140**	.242**
ACC			--	.160**	-.001	-.100	.311**	.216**	.188**	.312**	.356**	.279**	.356**	.265**	.317**
ESY				--	.262*	.196*	.209*	-.120*	-.015	.213*	.148*	.109	.149*	-.005	.170**
NRV					--	.098	.001	-.146*	-.081	.082	-.009	-.094	-.006	.075	-.033
INF						--	.054	-.204**	.204**	.058	.027	-.016	-.018	-.034	-.020
INT							--	-.272**	.409**	.373**	.548**	.575**	.405**	.264**	.551**
NOT								--	.080	.209**	.440**	.480**	-.155**	.080	.462**
PRF									--	-.209**	-.440**	-.480**	.296**	-.347**	-.347**
ABL										--	.435**	.346**	.488**	.173**	.369**
MOT											--	.542**	.473**	.223**	.585**
FAR												--	.467**	.277**	.593**
CON													--	.256**	.490**
JBK														--	.322**
IMP															--

Note. Dependent variables from questionnaire items 15 to 30. Higher values indicate greater correlation between variables. PRC = Confidence in English ability prior to taking test; KNO = Test knowledge before test; LOC = Local test superiority; ACC = Tests are accurate; ESY = Choose easier test; NRV = Anxiety affects test score; INF = Test preparation influences English use; INT = Policy increases English interest; NOT = Would not study English without policy; PRF = Prefer no test policy; ABL = Test preparation improves English ability; MOT = Policy increases motivation to study English; FAR = Policy is fair; CON = Policy increases confidence in English use; JBK = policy has no influence on job skills; IMP = Policy has overall positive impact.

* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

Appendix H

Ministry of Economic Affairs, December 2010 English Proficiency Test Data

Unit name	Staff #	SELPT passed					CEFR level			
		GEPT	FLPT	TOEFL CBT	Other	Total %	A2	B1	B2	C1
Secretariat	24	1	1	2	0	16.7%	1	2	1	0
Commerce	57	3	3	0	1	12.3%	4	2	1	0
Department of Mines	16	0	0	0	2	12.5%	0	0	0	2
General Affairs	46	1	0	0	0	2.2%	0	0	1	0
International Cooperation	27	0	3	0	1	14.8%	0	2	1	1
Investment Services	35	0	4	2	1	20%	0	2	4	1
Technical Department	24	0	5	3	1	37.5%	0	5	3	1
Research and Development	23	2	3	0	0	21.7%	1	1	3	0
Human Resources	29	1	0	0	0	3.4%	1	0	0	0
Ethics	11	0	0	0	0	0.0%	0	0	0	0
Accounting	30	1	0	0	0	3.3%	1	0	0	0
Statistics	24	1	1	0	0	8.3%	0	1	1	0
Legal Affairs	13	0	2	0	0	15.4%	0	2	0	0
Committee of Appeals	26	0	5	0	0	19.2%	2	1	1	1
Information Center	3	0	0	2	0	66.7%	0	0	2	0
Central Region Office	121	0	0	1	0	0.8%	0	1	0	0
Industrial Development Bureau	214	3	9	8	0	9.3%	1	4	14	1
Bureau of Foreign Trade	300	7	1	7	2	3.7%	2	4	4	7
Intellectual Property Office	419	15	31	12	6	15.3%	9	27	25	3
Bureaus of Standards, Metrology, and Inspection	880	17	21	10	8	6.4%	19	19	17	1
Water Resources Agency	1461	54	10	2	1	4.6%	56	7	4	0
Export Processing Zone Management	247	12	3	2	1	7.3%	12	3	3	0
Investment Commission	53	2	1	3	0	11.3%	1	2	1	2
Central Geological	75	1	6	1	0	10.7%	0	4	4	0

Survey										
Small and Medium Enterprise Administration	75	1	4	3	1	12%	2	2	5	0
Bureau of Energy	97	1	18	1	0	20.6%	2	5	11	2
Bureau of Mine	151	2	7	1	0	6.6%	2	2	6	0
International Trade Commission	25	1	7	0	0	3.2%	0	3	5	0
Total Number	4506	126	145	60	25	(356)	116	101	117	22
Percentage of SELPT		35.39%	40.73%	16.85%	7.02%		32.58%	28.37%	32.86%	6.17%
Percentage of staff						7.88%	2.57%	2.24%	2.59%	0.48%

Note. This table is adapted from a document provided by the Ministry of Economic Affairs.

