The Washback Effects of an English Exit Exam on Teachers and Learners in a Korean University English Program

Submitted By

Jason Adam Di Gennaro

to the University of Exeter as a thesis for the degree of Doctor of Education in TESOL

in June 2017

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other university.

Abstract

Studies have shown that language tests can and often do have powerful influences on teaching, learning, and the creation and dissemination of educational materials, such as textbooks, in addition to the formation and implementation of language education policies (Au, 2007; Alderson & Wall, 1993; Bailey, 1996, 1999; Cheng, 2008). While the literature provides evidence for this influence, collectively described as 'impact', or more specifically, 'washback', the form and intensity in which it occurs differ greatly across contexts, due to the dynamic and complex nature of washback phenomena. This case study investigated the washback effects of an English language speaking test, the GMATE (General Multimedia Assisted Test of English), used as an exit examination in a large university in Seoul, South Korea. Developed from the Washback Hypotheses (Alderson & Wall, 1993, p.120-121), there were two main research questions answered through this study: 1) What are the perceived washback effects of the GMATE on teachers' teaching? 2) What are the perceived washback effects of the GMATE on students' studying? To answer these and related sub-questions, a mixed-methods approach was taken, including questionnaires, interviews, and classroom observations. This provided a clear picture of what was occurring in this particular context, while offering a voice to the 459 students and 17 teachers who participated in the study. The findings of this study showed that the GMATE indeed had washback effects on the teachers and student participants, and that these effects varied depending on students' proficiency level, year in school, and term of study. Furthermore, these results supported the notion that washback is highly contextual (Cheng et al., 2014; Cheng, Sun, & Ma, 2015; Cheng, Watanabe, & Curtis, 2004), as this thesis highlighted the importance of bearing in mind sociocultural factors that may contribute to washback effects in this and other unique research contexts.

Acknowledgement

I would like to offer my sincerest gratitude to my amazing and supportive wife, Sookyung, who was always close by to offer words of encouragement (or appropriate scolding, during inevitable periods of procrastination). A language professional and educator yourself, you inspire me every day to be a better teacher, researcher, and person.

I must, of course, mention my energetic twins, Juneau and Geno, whose appearance near the end of this program lit a new fire within me and motivated me to push through to the end. Your happiness and joy helped to keep me smiling during exhausting editing sessions, and gave me a welcome break from my studies when most needed. I look forward to being able to spend more time watching you grow with the conclusion of this chapter in my academic journey.

Finally, incredible thanks go to my team of supervisors. Dr. Li and Dr. Riley, you brought such a wonderful mix of personality and expertise, and offered much-needed support combined with informative guidance throughout this long process. As a student away from Exeter, I appreciated your timely communication and accommodation, without which this project would not have been possible.

Table of Contents

Abstract ... 2
Acknowledgement ... 3
Chapter One: Introduction ... 7
1.1 Background and Rationale ... 7
1.2 Research Questions and Sub-Questions ... 11
1.3 Organization and Structure ... 12
Chapter Two: Context ... 14
2.1 Culture of Learning in Korea ... 14
2.2 Inequality in Education ... 15
2.3 Education as Competition ... 17
2.4 Leveling the Field ... 18
2.5 Educators in Korea ... 19
2.6 English Teaching and Learning in Korea ... 21
2.7 High-Stakes English Exams in Korea ... 22
2.8 The General Multimedia Assisted Test of English (GMATE) ... 24
Chapter Three: Literature Review ... 26
3.1 Critical Theory ... 26
3.2 Critical Applied Linguistics ... 28
3.3 Critical Language Testing ... 31
3.4 Washback Hypotheses and Models ... 34
3.5 Washback Research ... 39
3.5.1 Exam Stakes and Motivation to Succeed ... 43
3.5.2 Language Testing Policy and Use ... 46
3.5.3 Entrance and Exit Examinations ... 50
3.5.4 Washback on Teaching and Learning ... 52
3.5.5 Washback Studies in South Korea ... 59
3.5.6 Studies Finding Limited Washback Effects ... 61
3.5.7 Methodological Considerations ... 62
3.6 Summary ... 64
Chapter Four: Methodology ... 65
4.1 Theoretical Framework and Methodology ... 65
4.2 The Case Study ... 66
4.3 Research Questions ... 69
4.4 Data Collection Methods and Instrumentation ... 71
4.4.1 Questionnaires ... 71
4.4.2 Interviews ... 75
4.4.3 Classroom Observations ... 76
4.5 Participants and Selection ... 78

4.5.1 Teacher Participants 80
4.5.2 Student Participants 81
4.6 Data Collection Procedures 84
4.7 Data Analysis 87
4.7.1 Quantitative Analysis 88
4.7.2 Qualitative Analysis 89
4.8 Ethical Considerations 91
4.9 Methodological Limitations and Obstacles 92
4.9.1 Observation Difficulties 92
4.9.2 Researcher Bias 93
4.9.3 Interview Bias 94
4.9.4 Power Issues 94
4.9.5 Student Participant Voice 95
4.10 Summary 96
Chapter Five: Findings 99
5.1 GMATE Washback on Teachers 99
5.1.1 Washback on Teaching Based on Level 99
5.1.2 Washback on Teaching Based on Exam Stakes 108
5.1.3 Washback Effects Observed in Classes 115
5.2 GMATE Washback on Students 121
5.2.1 General Perceptions of English Exams 122
5.2.2 Student Knowledge and Perceptions of the GMATE 123
5.2.3 Student Test Preparation for the GMATE 126
5.2.4 Student Emotions Related to the GMATE 128
5.2.5 Student Differences Based on Level 130
5.2.6 Student Differences Based on Year 134
5.2.7 Student Differences Based on Seasonality 136
5.2.8 Differences Based on Type of Course 140
Chapter Six: Discussion 146
6.1 Summary of Key Findings 146
6.2 Exam Stakes and Perceptions 147
6.3 Teaching and Learning Processes 151
6.4 Contextual Factors 156
Chapter Seven: Conclusion 161
7.1 Key Findings and Conclusions 161
7.2 Implications for Stakeholders 163
7.3 Contributions to the Field 164
7.4 Recommendations for Further Research 165
7.5 Limitations of the Study 167

Appendices 169
A1 Student Questionnaire Korean 169
A2 Student Questionnaire English (translated) 170
B Teacher Questionnaire 171
C1 Interview Guide First Meeting Example 174
C2 Interview Guide Subsequent Meeting Example 175
C3 Coding Schemes 176
C4 Interview Coding Example 177
D1 Observation Scheme (completed) 178
D2 Observation Scheme (blank) 179
D3 Observation Scheme Item Definitions 180
E1 Bilingual Consent Form 181
E2 Teacher Information Sheet 183
E3 Student Information Sheet (Korean) 184
E4 Student Information Sheet (English) 185
F1 GMATE Item Examples 186
F2 GMATE Rating Level Descriptions 187
Bibliography 188

Chapter One: Introduction

1.1 Background and Rationale

Educational testing has become an important topic that has garnered a considerable amount of researchers' attention in the past few decades. The power that tests have to shape the ways that students study, teachers teach, and publishers create books is unparalleled in the realm of education (Bailey, 1996). The broader effects that are observed in changes to teaching, learning, materials creation, and policy are often collectively referred to as 'impact', while 'washback' is used to indicate those effects that are more directly involved in shaping the learning environment (Alderson & Wall, 1993; Wall, 2000). South Korea has a long and vivid history of examinations driving education practices, both in and outside of classrooms (Seth, 2005). In addition to the historic exam taken to become a government official or employee in South Korea (Seth, 2002), standardized English tests, too, such as the Test of English for International Communication (TOEIC), Test of English as a Foreign Language (TOEFL), and International English Language Testing System (IELTS) have become increasingly popular in Korea, with numerous Koreans taking these tests each year (Choi, 2008).

High-stakes tests are now used as gatekeepers to many stages of Koreans' lives. Important assessments are taken at different phases of their public education, with each serving to rank and track students as they progress through the system. None of these is more important than the KSAT (Korean Scholastic Aptitude Test), however, which is used by universities as a major deciding factor in granting admissions (Cho, 2007). For many South Koreans, success on this exam can mean incredible academic and subsequent employment opportunities, and thus the KSAT is often viewed as the pinnacle of one's academic endeavors. Recently, though, academic inflation has occurred in South Korea (Schmid, 2013), as more and more Koreans enter and graduate university, creating a well-educated, yet jobless pool of young people, and this has not gone unnoticed by news media around the world (Kim, 2011; Yoon, 2014; Hiatt, 2011). This academic inflation in South Korea has led some universities to offer

more certificate and diploma programs in order to help potential job seekers gain an advantage in the competitive Korean job market.

Another way some institutions have done this has been to create standardized exit exams required of all graduating students. Some of these assessment tools have been well received by major companies in Korea. Most notably, the Test of English Proficiency (TEPS), created by Seoul National University (Choi, 1999), has gained acceptance at large South Korean employers such as Samsung, and is now seen by students and companies as a viable alternative to the more prominent English exams available today, such as the TOEIC, TOEFL, or IELTS. Seoul National University is very highly regarded by Koreans as the top university in the nation, and indeed, this is supported by recent university rankings, showing SNU as the number one university in Korea (Quacquarelli Symonds and U.S. News rankings display SNU as first in South Korea, while Times Higher Education has SNU at second, in the 2017 tables). Thus, many universities make administrative policy decisions based on SNU's successes and failures (Abelmann, Kim, & Park, 2005).

An example of this may be shown by the fairly recent creation of similar language education policies and exit exams at other top universities in South Korea. In this testocracy, the additional requirement of further testing comes as little surprise to learners and educators, but the actual effects that these additional examinations have on the learning environment have not been thoroughly investigated. Stakeholders in this examination environment have become somewhat complacent with the current education system, but it may be that some assessments are being needlessly thrust upon teachers and students, costing all involved monetary investments as well as invaluable study and teaching time.

Much of language testing washback research done in Korea in the past has focused on the KSAT, and how the English portion of this exam affects learners and educators throughout the public school system (Cho, 2007; Hwang, 2003; Kim & O, 2002). These researchers have shown that public school English tests have a powerfully negative influence on teachers and students in South Korea, yet exams are still held in high regard as the only way to

provide an equal and fair opportunity for academic success to all students. Choi (2008) examined the impact of language testing in Korea, and specifically addressed the need for further research on the effects and influence of these types of exams.

Through triangulation by use of questionnaires, interviews, and classroom observations, this study applied ideas proposed in the Washback Hypothesis (Wall & Alderson, 1993) to the General Multimedia Assisted Test of English (GMATE), a high-stakes English exit exam, in order to examine to what extent positive and negative washback were occurring within the teacher and student populations. I hope that this thesis can assist in closing the research gap that exists regarding English exit examinations in Korean universities, and that the findings of this study may help to steer future inquiry and language testing creation and implementation not only in this research context, but in those similar contexts to which the findings and conclusions may be appropriately generalized.

As is supported by the breadth of literature on washback studies of entrance exams, there is much attention paid to assessments that act as gatekeepers to academic and career opportunities, but far fewer studies have investigated exams that must be passed before *leaving* university. This research was undertaken with specific goals of telling stories. The stories needing telling came directly from the teachers and learners in the research context, and were communicated through questionnaires and observations in the case of student participants, and by a combination of interviews and observations for the teachers. I problematized the current testing situation at a university in South Korea, as I believed the examination policy surrounding an English exit test there was disempowering both students and teachers, while also affecting them in other negative ways beyond those related to their studies or teaching.

I was specifically interested in the extent to which this test, the GMATE, affected students' perceptions and behaviors related to English study and test preparation, and how these factors may have changed based on students' English ability or test outcome needs. On the teacher side of the study, I wanted to know how and to what extent teachers were affected by

the test, as I suspected them to think and behave in similar ways to those teachers' stories presented in the literature.

However, this study provided some very interesting opportunities for unique contributions to the literature. First of all, this study was done in a university setting in South Korea, and focused on an English exit exam. A large number of studies in the literature focus on exit exams at the high school level, and these tests were generally regarded as integral to the university admissions process, making them important gatekeepers for the student stakeholders in those contexts. Another handful of papers looked at washback effects at the tertiary level, but in these cases the exams concerned were focused on reading and listening comprehension, with little to no attention paid to speaking abilities, particularly here in South Korea. In the current study, though, the GMATE speaking test tests *only* students' ability to speak, and does so holistically, whereas other speaking tests such as the IELTS yield individual scores in various criteria, which are then averaged and weighted.

Lastly, the current context provides a unique addition to the growing body of knowledge regarding washback effects, as the teacher participants in this study were of different backgrounds than their students. Washback studies are often done in situations in which the teachers and students share an L1, which may contribute to similarities across contexts with regard to how teachers explain and prepare for test items. In the current context, though, I was able to see how native English speakers taught Korean speakers of English without the aid of their students' L1, and how the education culture here in South Korea may or may not have influenced the ways in which washback effects manifested.

1.2 Research Questions and Sub-Questions

1. What are the perceived washback effects of the GMATE on teachers' teaching?

    1.1 To what extent do teachers feel differently about the GMATE, and how does their teaching and test preparation vary depending on the English proficiencies of their students?

    1.2 To what extent do teachers feel differently about the GMATE, and how does their teaching and test preparation vary when their classes include students for whom the GMATE is a high-stakes exam?

    1.3 What are the observed washback effects of the GMATE on the classroom environment?

2. What are the perceived washback effects of the GMATE on students' studying?

    2.1 To what extent do students of different English proficiencies feel and act differently towards English courses and the GMATE?

    2.2 To what extent do students for whom the GMATE is a high-stakes exam feel differently towards English courses and the GMATE than those students for whom the GMATE is low-stakes?

    2.3 To what extent does the length of term influence students' approach to English language learning and test preparation?

1.3 Organization and Structure

This study is presented in seven distinct chapters. In this first chapter, I present a rationale for undertaking the study while also offering the reader some background information related to washback and language testing. The goal of this section of the paper is to familiarize the reader with terms that will frequently appear throughout the study, as well as to offer the reader an explanation of the significance of the study. A brief historical overview of the current educational and testing situation in Korea is given and problematized here, and research questions are outlined for the reader.

Chapter two presents important characteristics of the research context. It is crucial that the reader understand the intricate details surrounds the study, as cultural, historical, and even physical factors that exist in Korea have a profound impact on the behaviors and beliefs of those living and operating within its boundaries. Korea has a very long history of education shaped by forces both foreign and domestic, with the resulting system representing a complex amalgamation of traditions and research. Additionally, this section of the paper provides the reader with some background regarding the teacher and learner populations in Korean universities, and ends by detailing the English exit examination with which this study is concerned.

The third chapter of this paper examines the literature related to the current study in depth, and provides criticism of methods and findings. The early sections of this chapter present the reader with the philosophical framework on which the research is conducted, and include explanations of pertinent theories relevant to the current study. A detailed account of the Washback Hypotheses (Wall & Alderson, 1993) is provided here, as these hypotheses form the foundation of the research questions, and helped in part to mold the questions included in the questionnaires and interviews. The later sections of chapter three focus on washback studies that have been conducted in the past, and by identifying strengths and weaknesses in these papers, I attempted to pull from those studies the most sound research ideas, while avoiding those notions that may not have enhanced this study.

Chapter four informs the reader of the methodology and research methods used in the study, as well as provides strong evidence for the choice of each tool or technique. This chapter also helps readers understand the paradigmatic framework under which this study is conducted, while tracing the connections between research paradigm and effective use of qualitative and quantitative research methods. Following the methodology, chapter five presents the findings of the research, allowing the reader to see and read the results of the questionnaires, interviews, and observations, along with explanations regarding how to understand the data provided.

Chapter six provides the reader with a discussion of the research data. While the previous chapter will have presented findings in a more raw form, accompanied by superficial explanation, chapter six allows for more in-depth interpretation and translation of the information collected throughout the study. This section draws specific connections with the literature, relative theories, and finally, the wider realm of educational practice, while noting incongruences between this study's findings and those of previous studies. Finally, the seventh and final chapter draws conclusions from the previous chapters, presents implications these findings have for stakeholders, and offers direction for future research, while also acknowledging limitations of the study.

Chapter Two: Context

2.1 Culture of Learning in Korea

Korean education and society have been profoundly influenced by Confucian philosophy and values, which place great significance on learning and academic achievement (Shin & Koh, 2005). Confucianism, which was considered to be a hindrance to the practical activities required for entrepreneurship and investment, as well as to modernization itself, has enjoyed a strong presence in Korean society for the greater part of two millennia, when it was used to create systematic education for the ruling elite, thus laying the groundwork for a society in which education was not simply the means to gain higher social status through employment or financial gain, but the mark of achievement in and of itself (Seth, 2005; Sorensen, 1994).

The collective Korean society and its members do not, however, seek this upward mobility as individuals, but instead define their success by the various groups to which they belong. Indeed, in addition to hierarchical societal constructs, Koreans are taught that filial piety underscores all matters, and these loyalties present themselves throughout Koreans' lives, as Koreans express their concern about "family matters such as family name, the prosperity of family, and mutual support among family members" while celebrating family members' successes as their own (Shin & Koh, 2005).

Confucianism further emphasizes its traditional values over the development of new ideas (Rhee, 1995). These tenets of Confucian thought are thus a lasting part of the nation's ability to create a culture and society in which citizens behave according to expectations built on these foundations, which extend into every aspect of Korean society, including home, school, community, and politics (Oh-Hwang, 1993). The degree to which Confucian ideas are a concrete influence on educational thoughts and practices in particular is shown by the rigid social class organization of Korean society, and by the king-like reverence of educators as dispensers of knowledge and molders of character (Siu, 1992).

Modern Confucian traditions do not exist in mutual exclusion with many educational practices, however. While some have attacked aspects of Confucianism in education, decrying its mistaken tendency to force students into passive submission and rote learning behaviors, other scholars argue that the differences between Western students and their East Asian counterparts are not nearly so pronounced (Shi, 2006). This evolution of Confucianism in education may put today's Korean students further at odds with the instructional choices made by their teachers, who were likely trained to teach in ways perhaps less conducive to creating a democratic, communicative learning atmosphere.

2.2 Inequality in Education

In the past, teachers were regarded as an indispensable part of the educational system created by and for the elite for its own edification and culture, provided to the wealthiest and most powerful families to allow these aristocrats the opportunity to effectively cement their status in society (Abelmann, Kim, & Park, 2005). An individual of proper family background could thus become "virtuous through the study of ethically oriented Confucian classics. He could then play an informal role as a moral exemplar and as a teacher and advisor to others, thus enhancing his status and influence in society" (Seth, 2005, p. 5).

In later years, there was a marked increase in literacy in males due to the expansion of private academies that provided education to the elite *yangban* class (Choe, 1987). This fervor for educational achievement, and the social success that accompanied it, led the wealthy elite to devote "a great deal of energy and expense on education and examination preparation. In this way they behaved much like modern South Korean families" (Seth, 2005, p. 6). While Confucianism formed the basis for much of Korea's education system for several centuries, the rigid examination system used as a gatekeeper for upward class mobility remained through the arrival of foreign missionaries and the Japanese occupation, regarded as a spark to Korea's modernization (Shin, 1978).

Japan's occupation of Korea during the latter part of the nineteenth and beginning of the twentieth centuries saw sweeping reforms in Korean education policy. These changes were not welcomed by many Koreans, many of whom were likely members of the upper classes, as the Educational Ordinance set forth by the Japanese sought to bring western-style democracy to Korean education while dispelling the established Confucian system of discrimination based on social class and academic achievements (Lee, 2002).

Ironically, the Japanese used this education system as a way to *create* inequality between the Japanese and Koreans residing in Korea. Quite apparent was this goal of disempowerment, as the Japanese set up "separate and unequal" tracks of schooling for Korean and Japanese children, with the latter receiving much more intensive academic attention, while the former was restricted to fewer years of less-focused instruction (Lee, 1984). Seth (2005, p. 6) expands on these activities:

> "The dominant view among Japanese policymakers was that Korea was a backward society, and that this backward society should occupy a subordinate position in the empire. Japanese wartime policies after 1938 further limited the number of higher education institutions, and redirected the curriculum away from literary pursuits and towards less prestigious technical education and vocational training. As a result, many middle class families became frustrated by the limited access to educational opportunities."

Because education was being used primarily as vehicle for cultural assimilation and political dissemination, only the select few that pledged loyalty to the Japanese empire, often already members of the Korean elite, were given the opportunity to attend institutes of higher education (Lee, 2002). This additional example of the state's exploitation of society led to a long-lasting relationship of antagonism that added fuel to Koreans' dissatisfaction with the lack of advancement in education and administration (Koo, 1993).

2.3 Education as Competition

The growing resentment felt by Koreans towards unstable government education policies carried over into the period immediately following the Korean War, as did many of the institutionalized influences the Japanese occupation had instilled into the Korean education system (Seth, 2002). Some legacies of the Japanese colonization of Korea still persist today, including a strict and regimented classroom environment, centralized government control of curricula, teacher training, and textbook development, and perhaps the strongest influences of all – the implementation of highly competitive examinations at various levels throughout the system and the shifting of educational funding costs from the government to the citizens themselves (Koo, 1993; Yoo, 1973).

The Korean government soon did away with the Japanese-enforced practice of detailed tracking and sorting of students, but this egalitarian approach quickly created a state in which all students "could and soon (most) did seek to advance to higher levels resulting in fierce competition" (Seth, 2005, p. 6). This newly realized environment of intense ambition to attain academic success gave rise to the formation of hundreds of new schools across the country, comprised of a mix of private, public, and those funded by outside governments or organizations, but these efforts were mere attempts to keep up with the public's insatiable appetite for education (Sorenson, 1994).

Seats in the best classrooms and schools were extremely sought after from elementary school on, and thus admissions to various schools were decided by a combination of interviews, teacher reports, and standardized exams (Sorenson, 1994). Intense pressure to perform well on examinations had been part of Korean education for hundreds of years, when the *yangban* elite and a few lucky bourgeoisie males would prepare for tests to become local and state officials (Seth, 2002). However, in the years following the war, socioeconomic status had become less readily identifiable, while remaining paramount to Koreans' measures of happiness and success (Robinson, 1994). This dissolution of status based on hereditary lines created a more level playing field in which anyone could achieve success through their

efforts alone, and therefore, although most Koreans disagreed with the stresses placed on families and children to prepare for these high-stakes tests, the inevitable alternatives of tuition hikes or other exclusionary practices would have likely been met with very little support from the general public (Seth, 2005).

Lim (2005) provides a detailed account of the different movements affecting the government's education decisions throughout history, but it remains clear that despite decades of intermittent, reactionary policy changes by the Ministry of Education in Seoul, private schooling's role as a crucial part of the overall education system has continued to be secured by the persistent focus on the university entrance admissions test, as well as the public's growing obsession with English ability and test scores (Seth, 2005). Today's education system in Korea attempts at every turn to level the playing field in the hunt for educational opportunity, but these activities, which run counter to a culture eager to assign rank and position whenever possible, have yet to accomplish the goals set forth by the Ministry (Lee, 2004).

2.4 Leveling the Field

The "education equalization" policies of the Ministry of Education, in addition to strict control of textbooks and curricula, include such actions as first assigning students to schools based on a lottery system, and then again assigning students to classrooms in a manner that creates a non-skewed normal distribution of academic achievement based on prior standardized tests (Kang & Hong, 2008). It is believed this would prevent certain classes receiving higher or lower levels of instruction based on their student makeup. It is not only the students who are circulated throughout schools, however. Teachers, as well, are rotated to schools with a virtually equal probability of teaching at any institution within a specified region or province, and are given incentives to teach in low-opportunity areas such as rural communities or districts with poor economic conditions (Kang & Hong, 2008). Society's perception of the quality of public education in providing adequate test preparation has

diminished the achievements of these leveling policies, and in some cases, the public has been right in being skeptical of the returns on their children's public educational investment.

English education policies set forth by the government in Korea are watered down when explained to lawmakers and the public with little regard to students' linguistic rights. What was originally an elective second language has become a mandatory part of every child's academic life, with no concern offered towards the actual advantages that English proficiency may or may not offer certain students. This kind of language policy creation is not uncommon in public schools, universities, and even entire nations (Shohamy, 2003). It is thus taken for granted that all students will benefit from attaining partial or total proficiency in the language of majority, though we can see in actuality that this is far from the truth.

2.5 Educators in Korea

Over the past forty years, Korean education has seen a transformation in the types of education available, as well as changes in how and to whom education is delivered. Education equalization policies have changed classroom dynamics and teaching practices by attempting to provide a more level academic playing field across the nation, but this has in turn shifted many families' and students' focus from what is taught in public school classrooms to that which is available in the booming private education sector (Seth, 2002). The importance of public school teachers, therefore, has waned, while the public perception of private education opportunities has transformed from one considering such classes as supplementary to a realization that the formal education students receive outside public school classrooms is most important to remain competitive.

Though there still exists an increased attraction towards becoming a teacher, with all of the stability and prestige that accompanies the position of instructor or educator in Confucian cultures, recruiting and retaining quality educators in Korea, like many countries, continues to be a difficult endeavor (Ambrosie & Haley, 1988). Attracting individuals to the teaching profession is more challenging as South Korea continues to develop economically. Relative

to other fields, teachers' salaries have stagnated in many countries, and many teachers are more concerned with what appears in their bank accounts each month, rather than what other intrinsic rewards their career choices may offer (OECD, 2005; Kang & Hong, 2008)

Although teachers in Korea have developed their professional identities through years of extensive training and education, and by successfully passing rigorous certifying assessments, their graduate degrees and broad understanding of their respective knowledge bases have been forgotten by the public and policy makers as the attention of South Korea's education system has been paid to high-stakes exams. The effects of standardized tests on teachers' ability to teach effectively, to adhere to that for which they morally strive as educators, and to develop professionally, have been investigated for many years (Smith, 1991).

In South Korea today, better language skills, and especially higher English abilities, are equated with an improvement in efficacy on the global competitive stage. This often places the education spotlight on teachers of English, and allows some to assign blame to English educators for falling university rankings or reduced student placement rates at international conglomerates (Chang, 2008). As outdated language education curricula based on grammar-translation approaches produced English language teachers with a firm grasp of the language, but who were unable to successfully teach or speak in English, schools and universities looked to native English speakers to help conduct more communicative English lessons, while this also served to help achieve the government's goal of reducing spending abroad on English education (Jeon & Lee, 2006).

In modern day classrooms, native speakers of English often teach alongside Korean English teachers in public schools, but in most major universities, native speakers alone teach the majority of conversation and writing courses. This may be due in part to students' preference for language instructors that are native speakers, based on students' valuation of their teachers' accents (Ling & Braine, 2007). Myriad socio-cultural factors affect how these university lecturers form their identities and conduct themselves as language educators and

professionals in Korean institutes of higher education, and this study comments on the extent to which required standardized examinations exert influence on these teachers, too.

2.6 English Teaching and Learning in Korea

Modern Korean education has thrust English education into the forefront, allowing the place of English language skills in Korean society to dominate debates on future curriculum and testing policy, while also allowing those who can prove their English skills (most often through the use of standardized test scores, but more recently also through English interviews) to gain a distinct advantage in competitive decisions such as admission, employment, and promotion. All of these aforementioned complexities that make up Korea's education history influence the current state of affairs regarding English education, as public and private school curricula clash, and students from more affluent families effectively seek out high-priced tutors to maintain their children's competitive advantage in classrooms (Kang, 2007). The goals of English education in Korea are consistently questioned, and the age at which English language studies should and shouldn't begin is always a topic of contention. It would not be unfair to say that English language studies, then, take a dominant position in the educational focus of Korean students and their parents.

The age at which English is taught in Korean public schools is now just the third grade of elementary school, although many parents spend astronomical sums to send their children to English kindergartens long before those students set foot in a public school classroom. And while not all younger children receive this head start in English learning, a large enough number enter their first public school English lessons at a proficient enough level that much covered in their government-mandated textbooks are of little to no real use, thus forcing teachers to attempt to adapt materials for the range of students that stretches from true beginners to those who may at only eight years old have had years of formal English instruction.

English education itself in Korea has undertaken several major transitions over the past decades, being pulled into more communicative teaching practices by younger teachers fresh from graduate schools, as well as advocates in universities and teacher training programs (Magno, 2010). Furthermore, the English learning experiences of the language teachers themselves can have a profound influence on the teaching methods that Korean teachers of English choose to employ in their classrooms (Moodie, 2016). However, the strength of the incumbent teachers and their unions has been such that only the most minor of shifts in curricula and teaching practices have made their way to the classrooms, as such a major change as moving from a grammar-translation approach to one in which teachers must attempt to conduct classes in English would require the replacement or incredible retraining of those currently holding positions in public school English language classrooms (Nunan, 2003). Therefore, English classes and their place in Korean public schools and universities remains as a talking point for presidential candidates, who all promise positive change in one form or another. History shows, though, that these changes are much easier said than done.

2.7 High-Stakes English Exams in Korea

At the center of students' lives in Korea, tests are simultaneously feared and revered for the power they wield in terms of changing the course of one's future. An integral part of Korea's extensive education history, exams have long served to divide and rank individuals in the ultra-competitive Korean society, while also serving as gatekeepers to better institutions, careers, and indeed, lifestyles. In particular, standardized English exams have emerged as force that have fueled a billion-dollar cram-school industry, helping students to remain competitive with others whose families make the financial sacrifice to send their children to institutes focused on English exam preparation, and while the bulk of the effects of high-stakes English testing occur in the private sector, there are surely changes to teachers' methods and materials based on what these exams set forth.

Today, a large part of this matter lies within the rankings of Korean colleges and universities. Students' performance on the College Scholastic Aptitude Test is widely held to carry the

most weight concerning university admissions, and thus years of study and thousands of dollars in private education culminate in this one high-stakes examination (Abelmann & Park, 2005). Korean society's recognition of affiliation with particular universities as inherently superior, despite the field of study undertaken, has led citizens to assume that students of institutions such as Seoul National University, Korea University, or Yonsei University are worthy of the enhanced employment opportunities that often await them upon graduation (Koo, 2007). The winner-take-all situation that has materialized in Korean education has grown more prominent with each passing admissions cycle, and the enormous amount of stress placed on students to perform well on the national university admissions test has resulted in annual protests and suicides in response to scores that were deemed unsatisfactory by parents or students themselves (Seth, 2002).

Although many students in the past were able to avoid the burdens of high-stakes English exams after gaining entrance to university, standardized English exams are becoming more common at some of the top universities in Korea. Universities are looking for ways to standardize their curricula while also giving students something to include on another line in their resumes, or "spec," as qualifications, certifications, and experiences are referred to here. According to data from a recent OECD report on education, Korea is sending far more students to university than most other developed nations, and 98 percent of Koreans between the ages of 25 and 34 have graduated from some form of tertiary education (OECD, 2010). This has created an overabundance of educated Koreans, and it now seems that education is no longer enough to set oneself apart from the pack when searching for gainful employment.

One reason for universities' shift in focus towards standardized English exams is the attempt to offer students a viable alternative to the external English tests already in place. Choi (2008) describes the major English exams currently offered in South Korea, and examines the impact these tests have had on the educational landscape there. He writes that "virtually all universities in Korea adopt a policy where students are to fulfill graduation requirements by obtaining a certain range of scores on some of the standardized EFL tests," and that "this practice is controversial among students and professors as well, as these tests may not serve

the purpose of graduation requirements adequately" (p.57). While students and faculty members have voiced their concerns regarding the effects of these tests on students' studying practices, little research has been done that actually investigates this influence and how it manifests within and outside of university classrooms.

High-stakes tests are not unique to Korea, and have been examined extensively in the literature due to their application to situations and systems in which inequality pervades (Solorzano, 2008). From an early age, students are subjected to these often government-mandated exams that track students into their future classes and ultimately, their careers. While they continue to be a convenient choice for ranking and separating students around the world, the physical and mental stresses associated with participating in these examinations has been shown to be extensive (Shohamy, 1982). It is an incredible injustice that is done to these students when they are forced to undergo the anguish of both preparing and sitting for these tests, and the discrimination continues as those students who perform poorly (often due to the stress itself, rather than a lack of ability or knowledge) are steered towards paths that offer them limited upward academic and occupational mobility. Academic success offers Korean students little respite from the pressures of standardized English exams, however, as the competitive job market now requires university graduates to prove their worth through the reporting of high-stakes English test scores.

2.8 The General Multimedia Assisted Test of English (GMATE)

The Multimedia Assisted Test of English (MATE) was first administered in its current form in 2006, after many years of research and development. It includes more than twice as many separate items as the GMATE, and covers a much wider range of proficiency levels. While both the MATE and GMATE are available as writing and speaking tests, only the GMATE speaking test was examined in this thesis. The GMATE speaking test consists of four items, three of which are rated by general English faculty members to achieve one of four possible holistic scores reflecting students' proficiency levels (see Appendix F). The test is administered via computer in a controlled setting, after which time test recordings are made

available to faculty members to be rated. Each recording is sent to two faculty members, and in the case that their individual ratings do not match, the recording is made available to a more experienced third rater in order to achieve better interrater reliability and assign an accurate rating. The four GMATE items are described in Table 2.1 below.

Table 2.1 GMATE Item Description

Item Number	Recording Time (Preparation Time)	Function
1	10 (0) seconds	Self-introduction Personal Information
2	60 (15) seconds	Picture Description
3	70 (20) seconds	Process Explanation
4	80 (20) seconds	Opinion Advantages/Disadvantages

While some English speaking tests, such as the International English Language Testing System (IELTS), use a more interactive, interview style speaking tool to assess test-takers' English ability, the GMATE presents a pre-recorded English prompt, after which time students have a specific amount of time to think (but not write notes) before speaking into a microphone. Specific examples of GMATE items and a detailed explanation of the rating levels are found in Appendix F.

Chapter Three: Literature Review

This chapter presents key theories and constructs upon which the current study was conducted. In the first section of this chapter, Critical Theory is outlined for the reader. Through better understanding of the history and development of this research framework, one can gain a better understanding of how I approached and problematized the current language testing situation surrounding the GMATE. Within Critical Studies is the field of Critical Applied Linguistics, which has given rise to Critical Language Testing (CLT). CLT has provided researchers with a framework to investigate intricate aspects of language testing from a problematizing perspective, and allows us to view assessment influence from multiple angles. After a review of the aforementioned theories, the Washback Hypotheses is outlined and explained as they relate to the current study. This important piece of research had great influence on the research questions and design of the current study, thus a detailed account and explanation is needed. Finally, recent and relevant washback studies are reviewed for the reader, and shortcomings as well as connections to the current study within the literature are addressed. I present these studies in such a way that I hope the reader is able to move from a more generalized view of washback effects in contexts around the world, towards the more focused contexts of East Asia, which share some cultural, historical, and education-related elements with the context of the current study. Finally, I look at washback research more specifically related to South Korea. In this way, I hope that the reader is able to more clearly identify the similarities and differences between different contexts, as I revisit the importance of context in washback throughout this thesis.

3.1 Critical Theory

The groundwork for critical theory was first laid at the Institute of Social Research at the University of Frankfurt in the 1920s (Creswell, 2009). Influenced by the work of Marx, Kant, and Weber, scholars such as Max Horkheimer and Herbert Marcuse were in agreement that the world was filled with injustice, and that it was this inequality in power relationships that should be the starting point for social research (Kincheloe & McLaren, 2002). Moreover, they

felt that “inquiry needs to be intertwined with politics and a political agenda [for reform] that may change the lives of the participants, in the institutions in which the individuals work or live, and the researcher’s life” (Creswell, 2009). Indeed, the terms *emancipation* and *emancipatory* are frequently used to describe and guide this theoretical worldview (Ponterotto, 2005).

As postpositivism arose from the positivist paradigm due to the lack of certain assumptions postpositivist researchers believed were crucial to the research process, so did the critical paradigm come to be in response to what a select group of scholars perceived as an inadequate attempt to centralize values “to the task, purpose, and methods of research” in other research paradigms (Ponterotto, 2005, p. 129). Mertens (2014) supports this, writing that the critical, or *transformative* paradigm originated in the 1980s and 1990s as researchers recognized the fact that many of the sociological theories forming the bases of the dominant paradigms had been created from perspectives of and through research on those in positions of power (i.e., white, male, etc.).

Like constructivists, critical researchers “advocate a reality that is constructed within a social-historical context (Ponterotto, 2005, p. 130). The language education and testing situation existing in modern day South Korea is representative of this reality, built through centuries of influence from those in power. This ontological view is also referred to as *historical realism*, and assumes that reality is shaped over time by “social, political, cultural, economic, ethnic, and gender factors, and then reified into a series of structures that are now inappropriately taken as… natural and immutable” (Guba & Lincoln, 1994, p. 110). Just as reality is viewed as being subjectively constructed over time, so is knowledge seen as being transactional and *value-mediated*. That is, the researcher and participants are assumed to be linked through their interaction and thus it is their values that inevitably influence the inquiry and subsequently, its outcomes (Guba & Lincoln, 1994). The impacts of the researcher’s proactive values are marked in critical theory, which distinguishes it from the value-free paradigms of positivism and postpositivism (Ponterotto, 2005).

The combination of simultaneously addressing history, society, economy, and culture as strong influences on present day realities, along with the notion that there exist injustices from which participants can find some reprieve, drove the current study. I have more than a decade of experience holding a variety of stakeholder roles, and rather than seeking to remove subjective ideas regarding language testing and language education in South Korea, I instead sought to harness those experiences to help focus and steer this investigation. There must exist a delicate balance in the Korean context, however, between providing sufficient critical evidence to elicit positive change, and appearing as though one is attacking the system in its current form.

The confluence of long-held cultural beliefs in hierarchy and *chaemyeon*, or face-saving, requires critical researchers to operate with both an artful cognizance of institutional expectations, and upon professional and ethical foundations that guarantee the research itself is sound. As Horkheimer writes about Critical Studies, “It is simply not possible to have admissions of this sort in the research report of an Institute that exists on the research funds of this shackling society” (Crotty, 2010, p. 140). While the papers mentioned by Horkheimer were ultimately published by an entity other than the Frankfurt Institute, the notion that one must work within present systems to be most effective in realizing positive change is a reality that existed then, and is still true today. The current study was thus undertaken with these goals in mind, while I strived to avoid excessive influence that may have undermined the ultimate objectives of the research. Pennycook sums this up beautifully in writing that it is “a compassion grounded in a sharp critique of inequality that grounds our work” as critical researchers (2010, p.7).

3.2 Critical Applied Linguistics

Developed from Critical Theory is Critical Applied Linguistics (CALx). The current study relied heavily on major principles of CALx, as I sought to explore and define the power differences and inequality that may have been occurring as a result of the recent language education and testing policies at a Korean university. CALx is a field of inquiry that includes and overlaps

with varying areas of research including Critical Literacy and Critical Bilingualism, as well as others that were drawn upon in this study, such as Critical Language Policy, Critical Pedagogy, and perhaps most crucial to this paper, Critical Language Testing (Pennycook, 2010). CALx was born out of the need to move beyond applied linguistics research that observed and described existing phenomena into a realm in which the status quo is effectively questioned and problematized, as language policies and pedagogies are scrutinized for their powerful ability to create and perpetuate power differences and inequality in classrooms and throughout societies (Pennycook, 1990). As outlined in chapter two, the South Korean education system provides numerous situations in which inequality occurs, and while some evidence of injustice may be deeply-rooted in historical or cultural practices (Dewey, 1929), others, such as language testing policies, are more recent developments, and are ripe for change.

As is the case many developed countries around the world, the Korean education system is run through a top-down approach, in which the government and related politics play a central role. The Ministry of Education in Korea maintains critical influence over almost every aspect of schools' curricula, hiring, testing, and pedagogical practices, and thus elections that place this power in the hands of officials are incredibly significant in the lives of Korean families. That instances of disempowerment or injustice may occur as a direct result of political choices, rather than despite them, is a major principle upon which CALx is formed. Pennycook (2010, p.7) outlines this idea in writing that "Critical applied linguistics is not about developing a set of skills that will make the doing of applied linguistics more objective, but is about making applied linguistics more politically accountable." In this regard, CALx doesn't only work in opposition to political beliefs, but instead is essentially forced to espouse those political stances that are in line with bringing attention to injustice and inequality.

Further to its somewhat awkward partnership with politics and political views, CALx relies heavily on self-reflexivity in order to avoid presenting ideas that simply replace long-held notions of what is just and right, and goes further in its attempt at transformative work by avoiding "grandiose modernist utopian" objectives, and instead strive for "alternative futures"

grounded in ethical principles (Pennycook, 2010). As the Korean context allows for change only by working within the present system, this focus on realistic, alternative situations presents attainable goals towards which to work. That is, the current study does not wish to imply that there exists a Korean education system in which all students and teachers are truly safe from the negative influences of testing practices, but that there must certainly be ways of approaching these issues that can eliminate some of the inequality ingrained in the system itself.

Phillipson (2012) agrees with the importance of political connectivity in writing that it can be considered a major weakness of EFL programs and institutions that they are commonly uninformed as to the local politics and culture surrounding language and education both, and that this ignorance can stem from a propensity to "ignore general educational research" (p. 250). He continues that it is precisely because English language education practices and policies have strived to remain apolitical that they have failed in doing just this. It is the outward dismissal of the political connections that these policies have that allows them to lead an often dysfunctional and ineffective existence, in terms of the preservation and respect of local culture and language (Phillipson, 2012). In Korea, however, we are starting to see educators and policy-makers express their concerns regarding the immense sociopolitical implications of English education policy reform, especially in terms of the perpetuation of socioeconomic inequality and injustice. I conducted this study with the hope that I might be able to lend further credibility to the argument that political change is necessary in order to realize positive transformation at the classroom level.

Finally, there are most certainly limitations on the degree to which political powers can be relied upon to invoke effective change in educational settings. Garnishing support for the noble cause of stamping out endemic educational marginalization and injustice through rhetorical chants of "equality" and "opportunity" are often used to not only promote state language policy, but to help reinforce uncritical beliefs and assumptions held by language educators as they justify their educational practices as "commonsense" (Tollefson, 1995). When CALx researchers propose changes to the current system, there is a very real risk that

those changes may replace long-held notions as universal truths, and thus a new power relationship is created based on stakeholders' beliefs that the new system is being put into place with their best interests in mind (Fairclough, 1989).

Freire (1993) believed that education was a tool that served to perpetuate or challenge the status quo, and that all educational stances were based on ideological stances that must be considered when making recommendations for change. Thus, it is with this caveat in mind that the current study was conducted: while the education system in Korea has undergone several enormous political and pedagogical shifts at different points throughout history, it has been and will likely remain so that the changes effected in the name of eliminating some instances of injustice will create others, and therefore it is imperative that any results and recommendations yielded from this paper are accepted with a high degree of skepticism and reflexive thought towards the many alternative futures that may exist as outcomes of such transformation.

3.3 Critical Language Testing

CALx provided a substantially more focused, critical approach to problematizing issues surrounding language teaching and learning, and from within that came Critical Language Testing (CLT). Just as CALx evolved out of the need for a more socially and politically responsible approach to applied linguistics research, so did CLT come into being as researchers began to question the social and value-oriented consequences of language testing policies. Shohamy (2001), for example, writes of posing questions regarding test construction and administration (p. xiii):

> Why was the test being given at the first place?
> What was the agenda that drove the introduction of the test?
> What were the politics of the test?
> Who was to gain and who was to lose?
> What was the political motive?

What was the relationship among the different bodies that administered it?
How would the results be used?
How would it affect teaching?
What did the test mean for the test takers, their parents, their schools?
What were the long- and short-range consequences of the test of the lives of the individuals?
What did the test do to the knowledge being assessed?

These questions drove early developments in CLT, and Critical Language Testing has held an integral place in education for many decades, and has served numerous purposes. While some researchers and educators have demonized tests and assessments as having a wholly negative effect on education (Madaus, 1988), others have supported the effective creation and implementation of exams in order to positively drive the development of quality teaching materials and methods (Popham, 1987). Tests have most often been used to measure the degree to which students have learned taught material, providing teachers with information to adjust future teaching practices, but they are also used to allow central governments to assert greater control over institutions (Cheng, 1999), to allocate students to selective places in higher education (Eckstein and Noah, 1993), and to innovate curriculum (Shohamy, 2001). Shohamy (1993) writes that exams may have a strong influence in test takers' lives in that (p. 2):

> "Results obtained from tests have serious consequences for individuals as well as programs, since many crucial decisions are made on the basis of test results. Among these are the placements of students in class levels, the granting of certificates or diplomas…the selection of students most suitable for higher-education institutions, and the acceptance of job applicants."

The stakes of the test were certainly very high in this regard, although for some student participants in the current study, the stakes had been lowered by recent policy changes made

in response to questions raised regarding the fairness and utility of the GMATE (see section 3.5.1 for more regarding exam stakes).

CLT helps researchers focus on the different social components of language assessment by providing specific aspects upon which to direct inquiry. As mentioned prior, the *use* of examinations within an institution or state are an important factor when investigating language tests and policy through a critical lens. Whereas traditional testing approaches assessments as individual, isolated events that can be focused upon with little attention paid to stakeholders, use-oriented testing notions raise concerns regarding the educational, social, and political implications that tests may have in the greater context (Shohamy, 2001). This view of testing takes into account important ideas of validity, fairness, voice, and power. Messick (1996), for example, argued that test validity is not something that can be measured unless one also brings into focus the effects that the test may have on curriculum, ethics, morals, politics, and knowledge, and also that values must be used to expand on the definition of testing validity as it currently exists.

The concept of voice in CLT is an important one, as it helps to determine and qualify the impact that a test may have on critical aspects of stakeholders' lives (Shohamy, 2010). Through stakeholders', and especially test-takers' own accounts, we can start to form a clearer picture of issues such as marginalization, stigmatization, negative effects on self-esteem and identity, and socioeconomic problems, in addition to other consequences that may negatively affect stakeholders' attitudes towards and perceptions of the classroom environment and education system in general.

Social values play a pivotal role in steering influential CLT research, and are a necessary force in directing education systems within societies (McNamara & Roever, 2006). These values may not necessarily be apparent, however, until policies or practices come under criticism or are subject to change (Akiyama, 2004). Furthermore, in discussing the interplay of society and exams, however, it becomes increasingly important that societal values and expectations and their influence on exams are considered. While this review of the literature

presents many papers on the effects that tests may have on society, Carlsen (2009) argues that more attention should be focused on the two-way relationship that exists between tests and society. Carlsen explains that in Norway's egalitarian society, high-stakes testing can be met with extreme backlash from the Norwegian parent and student stakeholders, and that traditionally there has been very little testing in schools when compared to other European countries (2009, p.349). In hierarchical South Korea, then, it is interesting to observe that society vehemently opposes high-stakes, standardized testing practices, while simultaneously relying on them to help create "equal" opportunities for students that may hope to succeed based on merit alone. As Carlsen (2009) writes, it is of upmost importance that education systems strive to create equality for students, even if this requires governments and institutions to allocate more resources to pupils disadvantaged socioeconomically. While efforts have made to achieve exactly that, it is an unfortunate reality that the modern education testocracy has succeeded in only widening the education and opportunity gap between the haves and have-nots (see chapter two).

3.4 Washback Hypotheses and Models

The influence and effects that exams, and in particular high-stakes tests have on learning environments and stakeholders are most often regarded as negative in nature, and can sometimes be incorrectly identified as testing 'impact' (Cheng, Watanabe, & Curtis, 2004). However, Wall (1997) draws a clear division between 'washback' and 'impact' in writing that *impact* refers to "any of the effects that a test may have on individuals, policies or practices, within the classroom, the school, the educational system or society as a whole," while *washback* pertains to the more focused "effects of tests on teaching and learning" (p. 291). The effects that tests and assessments have on teachers and students are collectively referred to as 'washback,' and will be described as such in this study. Bailey (1996) presents several definitions of washback (p.258, citing Buck):

> "…a natural tendency for both teachers and students to tailor their classroom activities to the demands of the test, especially when the test is very important

to the future of the students, and pass rates are used as a measure of teacher success…this washback effect can be either beneficial or harmful."

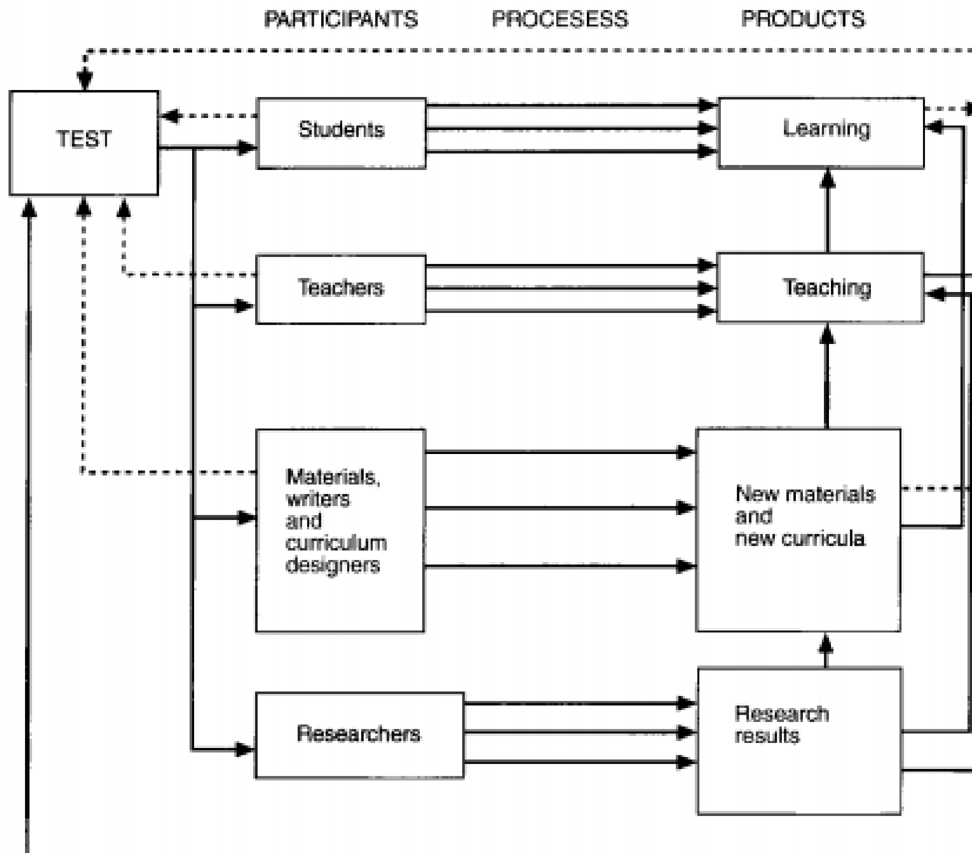
Through this definition we see that washback may not reside on either side of the dichotomy that papers by Madaus and Popham may create (Madaus, 1988; Popham, 1987). Another clear description of washback is provided as "the extent to which a test influences language teachers and learners to do things they would not necessarily otherwise do that promote or inhibit language learning" (Messick, 1996, p. 241). While this definition again seems to consider both the benefits and shortcomings presented by testing practices, Alderson and Hamp-Lyons (1996) identified four main areas in which the results of testing impact studies are concentrated, writing that findings showed a narrowing of curricula, loss of instructional time, a reduced emphasis on critical thinking or problem-solving, and a rise in test scores without a simultaneous increase in ability in the test construct (p. 281).

It is clear that, while many researchers acknowledge that assessments *can* be used to positively promote or reinforce teaching and learning, what is actually taking place in schools around the world may be at odds with these goals. Bailey (1996) summarizes and reinforces this point, reporting that "washback is generally defined as the influence of testing on teaching and learning, it is widely held to exist and to be important, but that relatively little empirical research has been done to document its exact nature…[while] there are concerns about what constitutes both positive and negative washback, as well as how to promote the former and inhibit the latter." (p. 259). In her extensive review of the available washback literature, Bailey draws four important conclusions (p. 275):

1) For a test to provide beneficial washback, stake-holders must understand the purpose of the test, what it is designed to measure, and how the results will be used.
2) Test results must be believable to test-takers and institutions, and must be made available in a timely and detailed manner.

3) Test-takers must assess the testing tool as credible and fair, as even the most transparent of exams will fail to elicit positive washback if it is judged otherwise.
4) The test must succeed in testing what a program is meant to teach. If these are clearly connected, administrators and teachers can more effectively interpret washback, be it positive or negative.

Bailey's model of washback is shown here (1996, p. 264):



In Bailey's model above, there are clear connections shown between exam stakeholders, materials, research results, and the exam itself, with arrows indicating the direction of influence. That is, arrows drawn from the test to teachers and students imply that the test is exerting washback effects on these participants. Bailey (1996) explains the dotted lines and

arrows as indicating possible “washforward” effects that may occur as the participants influence the test itself. As can be seen in the model, the only real influence currently acknowledged as having a strong effect on tests is research and the results yielded from such studies, but future inquiry may find evidence to support the dotted line effects shown above.

Cheng and Curtis (2004) review the varying views of washback as bidirectional, and emphasize that the researchers and participants, contexts, time and duration of testing practices, rationale, and different approaches used by participants will all influence washback in various ways, and that the acceptance of washback as occurring in both positive and negative ways should lead to methodological changes that strive to bring about positive change (p. 8). As mentioned previously, washback has traditionally been seen as a negative occurrence, with such pejorative phrases as “teaching to the test” being used to characterize inappropriate teaching practices as a result of testing policies. This facet of washback has been studied for decades, with particular attention paid to effects on the curriculum (Wiseman, 1961; Davies; 1968). However, just as there certainly exist contexts in which washback occurs in a negative way, there are other exams and testing environments connected to positive washback effects. Some researchers have proposed that testing be an integral part of curricula, and should then be used interchangeably with learning activities and tasks (Cheng & Curtis, 2004).

One factor affecting washback on which researchers seem to agree is that washback is an extremely complex and dynamic phenomenon, and can’t be observed or measured accurately without taking necessary steps to identify and describe the many aspects contributing to its existence. Shohamy (1993) raised concerns regarding washback effects on validity, and that these effects will come as a result of the interaction of numerous variables. Several of these variables are identified as The Trichotomy Washback Model (Bailey, 1996 citing Hughes, 1993). This model focuses on the following:

1) Participants: students, classroom teachers, administrators, materials developers, and publishers

2) Processes: any actions taken by participants related to the learning process
3) Products: what is learned, and the quality of learning taking place

In order to more clearly present the ways in which tests may affect education, Alderson and Wall (1993) investigated much more minute aspects of washback, and published their findings in a powerful piece in which they proposed the Washback Hypothesis (p. 120-121):

1) A test will influence teaching
2) A test will influence learning
3) A test will influence *what* teachers teach
4) A test will influence *how* teachers teach
5) A test will influence *what* learners learn
6) A test will influence *how* learners learn
7) A test will influence the *rate* and *sequence* of teaching
8) A test will influence the *rate* and *sequence* of learning
9) A test will influence the *degree* and *depth* of teaching
10) A test will influence the *degree* and *depth* of learning
11) A test will influence *attitudes* to content, method, etc. of teaching/learning
12) Tests that have important consequences will have washback
13) Tests that do not have important consequences will have no washback
14) Tests will have washback on *all* learners and teachers
15) Tests will have washback effects for *some* teachers and some learners, but *not* for others.

Many of these hypotheses were used to direct the formation of research questions as well as interview items for the current study, which are introduced in a later section of this paper. Some of these notions are more salient in the Korean context, as has been shown through a

growing number of studies in developing nations around the world (Wall, 2000, citing Eisemon, 1990; Heyneman and Ransom, 1990; Kellaghan and Greaney, 1992). In particular, that tests have important consequences, such as those examinations acting as gatekeepers to promotion or admission to higher education, has brought about the greatest degree of washback in the Korean context (Seth, 2002). This is supported in other contexts in an influential paper that introduced the notion that there are many factors that must be considered before washback can be effectively measured or predicted for a particular test in a particular context (Shohamy, 1993), and Alderson and Wall (1993) write that there are certainly other influences within a society or education system that would prevent washback from appearing, or that might change the nature of washback.

3.5 Washback Research

This section reviews empirical research studies into washback phenomena having occurred in a variety of contexts. As this thesis was conducted to investigate washback effects on both students and teachers, empirical studies into both teaching and learning are reviewed, and methodological considerations are noted and critiqued when appropriate. Certain aspects of students and teacher behaviors and attitudes influenced by high-stakes examinations have been explored in some research studies, and are presented in corresponding subsections. A summary of studies presented in this thesis are provided in Table 3.1.

Table 3.1 Summary of Empirical Washback Studies

Author(s)	Context	Washback Focus	Participants	Instruments
Alderson & Hamp-Lyons (1996)	USA	TOEFL preparation in specialized language institute classrooms	Students Teachers	Individual and group interviews Classroom observations
Andrews, Fullilove, & Wong (2002)	Hong Kong	Modifications made to Use of English (UE) exam to include speaking	Students	Video and discourse analysis of simulated oral examinations
Chen (2002)	Taiwan	Reformed Basic Competency Test (BCT) of English in Taiwanese junior high schools	Teachers	Questionnaires Group interviews
Cheng (1997, 1998, 1999, 2004)	Hong Kong	Revised Hong Kong Certificate of Education Exam (HKCEE)	Students Teachers	Questionnaires Teacher interviews Classroom observations
Cheng et al. (2014)	Canada China Taiwan	Comparison of washback in three different contexts (Canadian Academic English Language Assessment in Canada, College English Test in China, and General English Proficiency Test in Taiwan)	Students	Questionnaires Exam score analysis
Chu & Yeh (2017)	Taiwan	Perceptions of the GEPT and washback effects of the exam including various stakeholder participants at two universities.	Students Teachers Administrators	Questionnaires Interviews Classroom observations
East (2015)	New Zealand	Teacher views regarding newly-introduced high-stakes language testing policy in NZ high schools	Teachers	Questionnaires
Fox & Cheng (2016)	Canada	Comparison of successful and unsuccessful student experiences with TOEFL iBT	Students	Questionnaires Interviews Test score analysis
Green (2007)	United Kingdom	Comparison of test preparation on IELTS scores in three classrooms	Students Teachers Institute directors	Questionnaires Interviews Classroom observations Test score analysis
Hwang (2003)	South Korea	Analysis of textbooks used in Korean high schools for college admissions test preparation	N/A	Textbook analysis
Jung (2008)	South Korea	Washback on teachers from English listening section of Korean Scholastic Admissions Test (KSAT)	Teachers	Questionnaires Interviews Classroom observations Test question analysis
Kim & O (2002)	South Korea	KSAT influence on 12th grade high school English classes	Students Teachers	Interviews Classroom observations Textbook analysis
Munoz & Alvarez (2010)	Colombia	Washback of a newly introduced oral assessment system on EFL classrooms	Students Teachers	Questionnaires Classroom observations Student performance analysis
Pan (2014)	Taiwan	Comparison of washback effects from GEPT and TOEIC exams on groups of students subjected to different exit examination policies	Students	Questionnaires Test score analysis

Qi (2005, 2007)	China	Comparison of intended and resulting washback effects from the National Matriculation English Test (NMET)	Students Teachers Test creators	Questionnaires Interviews Classroom observations
Rodriguez & Arellano (2016)	USA	Effects of high-stakes testing on Latino students' higher education goals and subsequent enrollment in universities	Students	Questionnaires Academic reports Test score analysis
Shih (2007)	Taiwan	Comparison of two higher education institutes with differing use of the GEPT	Students Teachers Parents Administrators	Interviews Classroom observations
Shohamy (1993)	Israel	Comparison of washback effects of three exams of differing perceived and real stakes – one test of Arabic language skills, a test of English language skills, and an L1 reading comprehension exam	Students Teachers	Questionnaires Interviews Classroom observations Document analysis
Shohamy, Donitsa-Schmidt, & Ferman (1996)	Israel	Re-examination of the washback effects and impact shown in the findings from the 1993 Shohamy study, with focus on the Arabic and English language tests	Students Teachers Education inspectors	Questionnaires Interviews Document analysis
Stecher, Chun, & Barron (2004)	USA	Investigation into intended washback of a statewide writing test, with comparisons made based on teachers' perceptions and experiences	Teachers Principals	Questionnaires
Tsagari (2009)	Greece	Washback effects of the First Certificate in English (FCE) on teachers and students -- lending support to the notion of washback complexity	Students Teachers	Interviews Diaries Textbook analysis
Valli & Buese (2007	USA	Longitudinal study of changing tasks and roles of teachers due to high-stakes testing policies and administration	Teachers	Interviews
Wall & Alderson (1993)	Sri Lanka	Study of washback effects of a newly introduced English examination on teachers' reported and observed teaching practices	Teachers	Interviews Classroom observations
Watanabe (1996)	Japan	Comparison of two teachers' approaches to preparing their students for a Japanese university entrance exam	Teachers	Interviews Classroom observations
Xie (2015)	China	Washback effects based on how students perceive and place value on section of the CET, and how much faith they place in the exam's validity	Students	Questionnaires Test score analysis
Zhan & Andrews (2014)	China	Case study on CET4 influence on three students' language identity and out-of-class study and learning strategies	Students	Interviews Diaries
Zou & Xu (2017)	China	University administrators' perceptions of the Test for English Majors for Grade Eight (TEM)	Administrators	Questionnaires

The scores students achieve on the university admission exam in Korea are believed to be the most critical factor concerning university selection and placement, and thus years of focused study and thousands of dollars in private education culminate in this one high-stakes test (Abelmann & Park, 2005). High-stakes assessments are not unique to Korea, however, and have been examined extensively in the literature due to their prevalence in education systems in which inequality is commonplace (Solorzano, 2008). In Korea, from an early age, students are subjected to these often government-mandated exams that track students into their future classes and ultimately, their careers (Choi, 2008). While they continue to be a convenient choice for ranking and separating students around the world, the physical and mental stresses associated with participating in these examinations, for both teachers and learners, has been shown to be extensive (Shohamy, 1982).

Solorzano (2008) examined a large number of studies related to testing impact and addresses several issues accompanying high-stakes testing. He writes that tests are often used by governments to promote accountability of both schools and teachers; however, educators are simultaneously being asked to teach from resources they have often not chosen, but are held accountable for the outcomes of students on the high-stakes exams that follow their teaching (p. 263). It is at this point that many teachers choose to narrow the curriculum or create their own materials in order to satisfy the accountability expectations thrust upon them by administrators (Shohamy et al., 1996). Similarly, Au (2007) examined 49 research studies investigating high-stakes testing, and synthesized their findings, noting that "The primary effect of high-stakes testing is that curricular content is narrowed to tested subjects, subject area knowledge is fragmented into test-related pieces, and teachers increase the use of teacher-centered pedagogies" (p.258). Furthermore, Au found strong support for the notion that curricula were narrowed and fragmented as a direct result of assessment practices, and that non-test knowledge is often neglected in favor of test knowledge (2007). Finally, there was a strong incidence of pedagogical shift seen in these studies, towards a much more teacher-centered approach, while student-centered activities and teaching practices were shunned. These effects may provide some explanation for the situations in which some Korean teachers defend the use of high-stakes exams, as the

resultant shift to teacher-focused instruction may be more in-line with their teacher education and personal learning experiences.

3.5.1 Exam Stakes and Motivation to Succeed

Wise and DeMars (2005) investigated the effects of low-stakes exams on students, and found that learners will tend to put forth much less effort when faced with these types of exams, therefore underestimating proficiency, and subsequently invalidating the results. While these test results may have major consequences for teachers or institutions, making them in some ways high-stakes tests, the impact on the students is such that there exists little motivation for students to put forth their best efforts (Wise & DeMars, 2005).

Xie and Andrews applied Expectancy-value theory to a context in which they believed washback effects were occurring and collected questionnaire data from 800 university students preparing for a high-stakes English exam in China (2013). Although this study presented only questionnaire data, it provides an interesting connection between the value students placed on the exam, and how that translated to motivation to succeed, as well as influencing the learners' test preparation and study behaviors. The researchers in this study showed that students who valued test taking, positively viewed test design, and had higher expectations of success on the exam were more likely to engage in test preparation (Xie & Andrews, 2013). In this thesis, I looked at students valuations of English language testing in general, as well as of the GMATE in particular.

In another glimpse into the connection between student motivation and exam stakes, Xie (2015) used data from the aforementioned study to look at how university students showed washback effects, either positive or negative, based on their valuation of weighted tasks or test items. This study had the added strength of analyzing questionnaire data in conjunction with the actual exam scores (outcomes), which helped Xie present a stronger discussion of the findings. She showed that students valued the more heavily-weighted portions of the test, prepared more for those sections, and exhibited some positive washback effects, although

these occurred alongside negative effects, as well. These findings may seem somewhat commonsensical, but they provide further evidence for the influence that exam stakes, real in this case, as the test items carried true differences in weight, yet perceived in the case of the GMATE, as a great many students were very explicitly subjected to a lower-stakes exam.

In my study, there were many students who were retaking the course after having failed the GMATE in previous attempts, which may have affected how they perceived the exam, and their connected levels and sources of motivation to prepare for and achieve success on the test. Research done in an American college showed that adult learners in that context were demotivated by failure of a high-stakes exam, and that subsequent attempts provided lower outcomes (Rodriguez & Arellano, 2016).

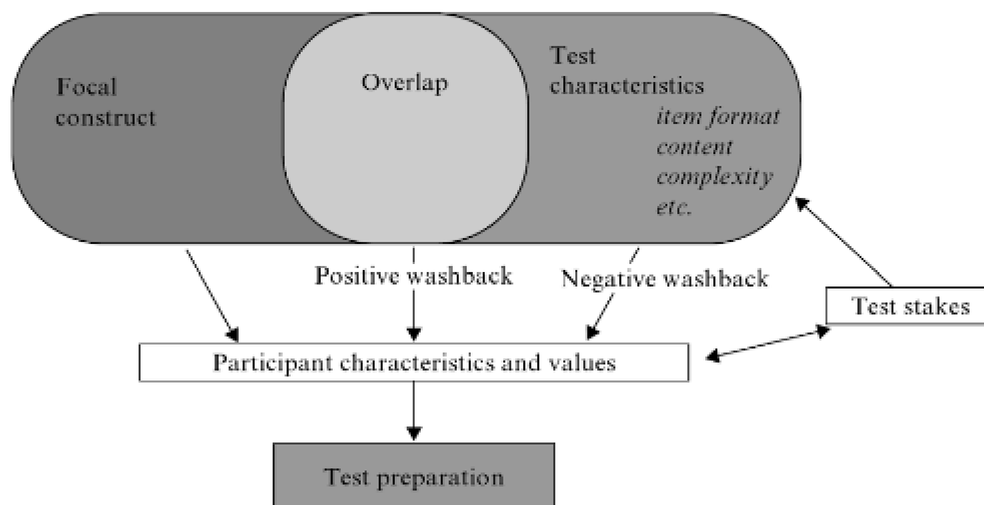
Shohamy, Donitsa-Schmidt, and Ferman (1996) observed great differences between the washback effects of a low-stakes test of the Arabic language, and a high-stakes English exam. While teachers reacted strongly to the English exam, there was virtually no change from teachers' usual methods or content in the case of classes related to the Arabic test (Shohamy et al., 1996, p. 304). They observed further that teachers preparing for the English exam simulated exam conditions and included many more activities directly related to test-taking techniques or strategies. The researchers in this study noted that tests were used in Israel to effect changes to the overall education system. Through the implementation of these tests, it was hoped that Arabic would become a more significant language that would attract more students, as well as that English would be taught and learned with more attention paid to speaking and listening skills. Unfortunately, the changes made through the exams were not accompanied by similar shifts in teacher education or curricula.

Prior to the aforementioned study, Shohamy (1993) looked at teaching and learning conditions related to several exams used in Israel schools. Through classroom observations, questionnaires, interviews, and the analysis of relevant documents, Shohamy's data allowed for effective triangulation in order to draw accurate conclusions. The researcher investigated the influence of a test of Arabic, a reading comprehension exam, and an assessment of oral

EFL skills, and found that each test had varying intensities of influence on the exams' stakeholders (Shohamy, 1993). In the case of the Arabic test, teachers claimed to have been affected in that they better knew what and how to teach, but that although the test seemed to raise the profile of Arabic in schools, it also created tension and pressure (Shohamy, 1993, p. 10). A narrowing of the curriculum was most apparent in the early adoption period of the test of Arabic, but the researchers observed that washback effects disappeared over time as the distinction between teaching and testing slowly blurred. Interestingly, classroom observations showed that teachers did not vary their teaching much at all, perhaps due to the fact that textbooks had been created specifically tailored to the exam material.

The GMATE assessment has been in use long enough that the textbooks and teaching materials created for the English program connected with it may have taken on a similar appearance to the 1993 Shohamy study mentioned above. By changing the stakes, however, it may have been that there was a shift in studying and teaching methods or materials used during classes. Green (2007) sums up the bi-directional influence that exam stakes and participants and their values have on one another, as well as how test stakes impact test design in a simple yet eloquent figure, shown below (Green, 2007, p.17):

Figure 3.1 Model of Washback Direction, Incorporating Test Stakes



As one can see in this depiction, test stakes exist on a continuum affected by the participants' perceptions and beliefs. Green writes that "participants set the stakes according to their awareness (or lack of awareness) of the uses to be made of the test results. The stakes associated with the test influence the behavior of those preparing for the test; high stakes encouraging greater adjustment on the part of participants. They also affect test design issues" (2007, p.17).

It should be noted, however, that participant values and beliefs are incredibly variable within and between contexts, again highlighting the importance of context in washback research. Cheng et al. (2014) conducted an exploratory study that looked at high-stakes testing situations in Taiwan, China, and Canada in order to investigate how social contextual factors influenced motivation, anxiety, and testing outcomes. They found further evidence that context, personal factors, and test use all contribute simultaneously to performance and motivation, writing "the differences in test importance to stakeholders and test purposes in each context, in turn, influenced test-takers' motivation and test anxiety within the specific testing context" (Cheng et al., 2014, p.325).

3.5.2 Language Testing Policy and Use

This thesis was conducted with a hope to change and positively influence the language testing policy regarding the GMATE. Important questions that must be posed regarding the creation and administration of language exams are, how is knowledge defined for the purposes of the exam, and who is making this crucial distinction (Knight, 2002)? Issues of validity are also appropriately raised here, as these are critically intertwined with concerns regarding the use and interpretation of tests, as well as the social consequences and fairness of examinations (Messick, 1989). An important moral issue is raised when addressing the definition of mastery, proficiency, or general language "knowledge" assessed through examinations. Johnston (2003, p.83) explains this:

> "The fundamental *immeasurability* of language proficiency lends a further moral dimension to our work in language assessment; the decisions we are forced to make about how competence will be assessed are always subjective and thus can only be rooted in our beliefs about what is right and good, beliefs which, we must always acknowledge, could be mistaken."

Shohamy (2000, p.15) writes that when tests are used to create policies (such as is the case for entrance/exit examinations, like the GMATE), the test often "becomes the single most influential pedagogical source and the *real* knowledge." In the current study, the GMATE exam presented several tasks as defining language knowledge (see section 2.8) and decision rules that determine at what level a student is determined to have achieved sufficient mastery of these tasks. These decision rules must also be scrutinized and compared to their real-world generalizability, and the specific contexts and target language should be considered when making inferences and interpretations based on exam scores (McNamara & Roever, 2006).

As another example of defining knowledge, in this case as oral English abilities, Shohamy (1993) examined the effects of a newly introduced oral EFL exam, and she found that teachers reported spending much more time on oral language skills, but also that teachers constrained their definition of "oral activities" to role-plays and interviews that reflected items included on the exam (p. 12). Furthermore, the researcher observed novice teachers to be less affected by the test changes than those teachers who had been teaching for a longer period of time, perhaps due to more extensive training in communicative tasks and teaching methods. As tests are used to define and value knowledge, so in some cases does the training and expertise of educators become more or less valuable based on their abilities regarding test preparation. In the Korean context, for example, many tests are skewed heavily towards assessing students' knowledge of grammar and vocabulary, and thus teachers that are better-versed in teaching through communicative methods, such as the novice teachers in the Shohamy study, may not be appreciated for this skill set, as it is rendered useless by the demands of the exams in place.

The use, or utilization of test scores is a very important aspect of CLT, and has appeared throughout the literature (Bachman, 1990; Bachman, 2004; Bachman, 2005, Kane, 2002; Lynch, 2001). While the use of tests and exams to introduce new policies has been argued as unjust (Shohamy, 2000), the continued development and dissemination of assessments in modern education is a reality with very concerning consequences. Shohamy presents one caveat for language testers seeking to create fair and just exams in writing that "language testers cannot remove themselves from the consequences and uses of tests and therefore must reject the notion of neutral language testing. Pretending it is neutral only allows those in power to misuse language tests…language testers must realize that much of the strength of tests is not their technical quality but their use in social and political dimensions" (p.15). If we cannot strive for a truly fair system of language assessment, then how high can our hopes be concerning the creation and implementation of more just language testing policies? That the use of exams must be very closely monitored is an idea that is very important to CLT, and is raised repeatedly in washback studies, as well.

Shohamy investigated the impact of an L1 reading assessment and writes that the use of the exam for purposes other than for which it was originally intended, to judge, punish, and reward teachers based on students' scores, made this a high-stakes exam (1993, p. 16). This may have intensified the washback effects based on teachers' perception of the importance of the test, shaping teachers' classroom activities and tasks directly based on what material appeared on the exam. A very interesting point raised by the findings of this study is that the stakes of a test are not always defined by the test-takers' situations, and that an exam may indeed be high-stakes if it carries powerful consequences for any of the stakeholders involved. A clear strength of this study is in the inclusion of observations, as well, which is considered by Wall and Alderson (1993) to be a crucial component of critical testing and washback studies, and helps to better recognize the shifting nature of washback effects when they are observed throughout the study. The current study also employs observations for this reason, and relies on that data in conjunction with questionnaire and interview information to discuss the context and washback effects that may be occurring.

An important aspect of testing and education policy, however, relates specifically to context. Shohamy presents a detailed insight into the use of language tests as not only education policy tools, but as instruments to affect and define language policies on larger scales (Shohamy, 2007). She argues that these policies can exist as either overt, as in the case of the GMATE policy being explicitly outlined for teachers and students, or more implicit, where these policies exist through the presence of "de facto practices" (p.119). Fulcher takes notions of language policy in context one step further, however, presenting the notion that language testing and policy stems from the political philosophy of a nation or region (2009). He asserts that the ways collectivism and individualism are perceived and espoused within a context can serve as a predicting measure regarding what kind of testing environment one will find. In South Korea, for example, a nation built on collectivism, one would expect to find testing practices that assign more absolute value to an exam to accurately determine a test-taker's aptitude or worth in cases of work or education, and it would be more likely that testing would be used to a gatekeeping effect in a collectivist state than in an individualistic state (Fulcher, 2009).

Language policies themselves are created and applied by those in power, and in most educational settings, this means that administrators are often those that ultimately decide the what and how of testing policy. Although most washback studies focus on teachers, students, or both of these stakeholder groups, some research has paid attention to or included administrators in the inquiry. Chu and Yeh (2017) interviewed administrators as part of their study into the washback effects of the GEPT in Taiwan, and reported that administrators generally viewed the exam favorably, particularly noting the test's ability to provide norm-referenced, quantitative data to describe the effectiveness of teaching and learning. The researchers in this paper wrote that the administrator participants in their study were fairly apathetic to the difficulties students had in passing the test in order to graduate, and instead focused on overall pass and failure rates of their respective student test-takers. Although the administrator data presented in this study was fairly scant, it did provide an important look at administrator perceptions and valuations of the exam in this study.

In China, Zou and Xu (2017) used questionnaires to inquire into university administrators' perceptions of a major exit exam for students majoring in English, the TEM8. Interestingly, in this study, too, the administrator participants viewed the exam favorably, and showed that they were knowledgeable regarding its form and related syllabus. However, although the participants in this study were less aware of the marking criteria attached to this test, they nonetheless believed in its ability to accurately reflect teachers' ability and to predict students' future performance, as well. Finally, the participants in this study believed that the test exerted positive washback effects on their university programs in general, and that the intended of use of the exam was being faithfully carried out, with only very minor exceptions (job employment, university ranking, etc.). This study gives us an excellent view into the perceptions of an oft-neglected group of stakeholders. To be able to juxtapose these beliefs with those of teachers and students is a unique opportunity in washback research, indeed.

Finally, McNamara (2000) highlights the inherent erraticism of washback, and the consequent dangers existing when governments strive for positive washback effects through assessment and policy reform (p.74):

> "Authorities responsible for assessment sometimes use assessment reform to drive curriculum reform, believing that the assessment can be designed to have positive washback on the curriculum. However, research both on the presumed negative washback of conservative test formats, and on the presumed positive washback of communicative assessment (assumed to be more progressive) has shown that washback is often unpredictable."

3.5.3 Entrance and Exit Examinations

Washback research has generally focused on two types of test: entrance examinations, often required to gain admissions to university, such as the KSAT in Korea or NMET in China; or exit exams, such as the CET or TEM in China, and the GMATE in the research context of this thesis. This section presents research studies that have focused on important gatekeeper

roles in students' desires to enter further education, or to graduate from school to pursue other goals.

Entrance examinations are often administered to students near or at the end of their secondary education, just before applying to universities. The TOEFL, TOEIC, and IELTS exams, for example, are widely used in English-speaking countries as admissions prerequisites for prospective students from countries in which English is not spoken predominantly, and thus have been studied fairly extensively for the effects those exams have on student and teacher test preparation processes (Alderson & Hamp-Lyons, 1996; Fox & Cheng, 2016; Green, 2007; Pan, 2014). But standardized entrance exams are not limited to these widely-used tests. For students wishing to study in their home countries, high-stakes exams created and implemented domestically are often waiting for them as they approach the end to secondary education.

In Hong Kong, the HKCEE is taken as students' complete their high school period of study, and success is crucial in order to pursue further tests or education opportunities. Cheng's dissertation (1997, 1998, 1999, 2004) focused on this test, and through a mixed-methods approach with student and teacher participants, shone a light on the testing situation there, while serving to help direct and drive future inquiry. In China, the NMET (Qi, 2005; 2007) is used to determine students' fates in higher education (and in many cases, their destiny beyond university, too). In Korea, the KSAT (Hwang, 2003; Jung, 2008; Kim & O, 2002) controls whether students enter university, and to which university they may gain admission. Finally, in Japan, there exist several forms of university entrance exams, although the largest number of students take the National Center Test for University Admissions (Watanabe, 1996; 2013).

Domestic tests such as these have a major impact on students' ability to enter university, but passing another form of exam is often required of them after their matriculation. While these exit examinations are similarly high-stakes in their role in separating students from their hopes of graduation and further education or career aspirations, they are less widespread around

the world, and thus research into exit examinations is slightly less pronounced in the literature. In China, students must pass different exit exams depending on their course of study. For students majoring in English, the TEM4 and TEM8, taken during the second and final years of university study, respectively, are required (Zou & Xu, 2017). These exams are much more rigorous than their counterparts which must be passed in order to graduate for non-English majors, the CET4 (Qian & Cumming, 2017). An interesting aspect of the CET4 is that the speaking component is only accessible to students who achieve high scores on the reading, listening, and writing components.

In Taiwan, universities require students to achieve certain scores on their choice of several examinations. From 2003 to 2012, the percentage of universities and technical colleges in Taiwan requiring students to pass English exit exams rose from just 5% to 90% (Pan, 2014). Exams available to the students there include the IELTS, TOEIC, TOEFL, GEPT, and CSEPT (College Student English Proficiency Test). This policy is somewhat similar in nature to the individual English exit examination policies here in South Korea, as students are sometimes given the option to meet their university's exit exam score requirement by submitting an corresponding score on the TOEIC or TOEFL. This is the case in the context of my study, too, although students often focus their attention on preparing for the GMATE, due to the fact that it is a graded component of the required English curriculum.

3.5.4 Washback on Teaching and Learning

An important notion raised by Shohamy (1993) is that washback "is complex, occurring in a number of directions, strongly dependent on the nature and purpose of the test, and changes over time" (p. 15). That is, it is still not yet possible for us to accurately identify and quantify washback effects due to the contextual nature, and the fact that they can exist in both positive and negative states. Furthermore, as this review of the literature has shown, these washback effects can grow stronger or dissipate throughout the course of teaching and learning, and present themselves very differently depending on the stakes and type of the examination involved.

Valli and Buese (2007) write that washback may have a stronger effect on teachers that previously regarded their roles as teachers as more certain or unchangeable, and that teachers were likely to alter the depth and pace of their instruction as a direct result of high-stakes exams (p. 531). They conducted a study over several years with principals, teachers, and specialists through individual and group interviews. Through these interviews and classroom visits, they observed that teachers' roles would be increased, intensified and expanded as a direct consequence of testing policy reform. That is, teachers' roles would be increased as shown by a greater number of tasks as well as the complexity of the tasks themselves, intensified by definition of the increased responsibility for teaching and learning according to new exams and policies, and expansion as exhibited by extra responsibilities outside the classroom (Valli & Buese, 2007). Valli and Buese's findings were taken into account when designing and conducting classroom observations and interviews, so that such role effects might be confirmed to exist in the context of the current study.

In Japan, Watanabe (2000) reported findings in accordance with the Washback Hypothesis, showing that teachers and their teaching practices would be affected by exams in different ways, although many of what he observed runs counter to the results of most other empirical studies on language testing washback. While teachers participating in the study showed themselves to be influenced by the format or content of the exam, several educators reported that they in fact tried to create a variety of materials to help prepare for a major English exam, while also showing innovation and a reluctance to adopt overt test-taking preparation techniques in their classrooms (Watanabe, 2000, p. 45).

Another speaking test underwent reform in New Zealand, as it was transformed from a summative teacher-student conversation to a collection of more formative, yet spontaneous instances of student-student peer interactions (East, 2015). In this study, 152 teachers were surveyed regarding their experiences with and perceptions of the old and new exams, and although they reported that they generally preferred the newer test, and considered it to be more useful, but that they believed in the strongly-held notion that examinations should be held in some way apart from the activity of teaching, which runs counter to those arguments

proposing the adoption of dynamic testing practices (East, 2015). Furthermore, similar to the concerns raised by Munoz and Alvarez (2010) regarding teacher and rater training for specific tests, procedures, and rubrics, Ducasse and Brown (2009) investigated raters' accounts of rating peer interaction as a speaking test, and found that it requires a very different set of rating criteria than traditional interview speaking exams, and that the rating scales used by raters should be clearly explained and consistently applied in order for the test to remain valid.

In Hong Kong, Cheng (1997) took a mixed-methods approach combing teacher and student questionnaires and classroom observations and found that teaching would be influenced by major tests most, and that washback that may be intended to occur in a top-down fashion may in fact coexist with unintended side effects or consequences. She observed that although the content or materials used often changed with alterations made to a major test, the methods employed by teachers generally remained the same (Cheng, 1997, p. 52). For example, although more communicative activities like role-plays were used in place of simply reading aloud, both activities were practiced in classrooms through drilling, which may have defeated the original aims of these activities. Cheng (1997, p. 51-52) recommends further research into the "superficiality" of washback effects, and cautions against over-interpreting the results of washback studies, writing:

> "Investigation should first be carried out to analyze the nature of the test type, be it a large scale public examination or classroom assessment since the function and/or the stake a particular test bears determines the degree of its influences and the area of its washback intensity. Then the investigation should direct attention to the investigation of that particular education context under which the examination is issued. Only by analyzing the particular educational context under which the examination works, would we be likely to find out how a particular examination might influence teaching and/or learning; how intensive the influence could be and the kind of washback effects, positive or negative on teaching and learning."

Alderson and Hamp-Lyons (1996) compared the washback effects of tests using two classrooms. One class was considered "normal," while the other was dedicated to preparing students for the TOEFL exam. This allowed researchers to more effectively observe and judge washback effects using what seemed to be control groups. Although there was teacher overlap between the classes, the researchers found further supporting evidence of strong washback effects on teaching methods, noting that teachers acted in very different ways concerning their usual classes in contrast to behaviors they showed when approaching TOEFL courses (Alderson & Hamp-Lyons, 1996). The researchers observed teachers spending more time talking, and the participants said that they were less likely to assign and collect homework or spend time planning lessons for their TOEFL classes (Alderson & Hamp-Lyons, 1996, p. 292). Although some participants in this study were strongly affected by the exams they were teaching to, others reported little or no effects, supporting other studies that showed washback effects to vary from teacher to teacher (Cheng, 1997; Cheng, 1999; Watanabe, 1996). The authors of this paper supported Cheng's (1997) idea that washback is a highly complex phenomenon that cannot be effectively defined by existing hypotheses.

Andrews (1994) conducted a study in Hong Kong that supported Spratt's (2005) suggestion that there exists an integral role that teachers play in connecting the classroom to intended washback effects set forth by exam creators. The researcher found that there was a major disconnect between the expected student motivation and study changes sought by those in charge of developing curricula and exams, and what was actually experienced and observed by teachers (Andrews, 1994). Andrews and Fullilove (1994) explained that these intended effects may not be fully realized due to "conflicting message(s) implicit in the tests, especially in those countries where examinations are highly important and yet where the examination format has been particularly resistant to change" (p. 58).

Andrews and others (2002) used a case-study carried out in two phases to investigate the efficacy of working to achieve washback in classrooms and found evidence to support Cheng's (1998) findings that washback effects are much easier to elicit in *what* teachers teach (content), rather than *how* teachers teach (methodology). The researchers in this study

also noted that washback is very complex, particularly because of the unpredictable nature of individual teacher and student differences (Andrews et al., 2002).

In the current study, some time was spent investigating both the content and methodology being used in classrooms. As teachers in the university were in charge of creating their own textbooks and supplemental content, interviews and questionnaires provided me with an accurate picture of how the test was affecting materials development and teaching practices in the English language classroom environment. The stakeholders in any given testing context will be affected by and will influence an exam, but beyond the stakes of the exam, one must also consider the format and delivery of a test as contributing to the complexity of washback. Two unique aspects of the GMATE are that it is administered via computer, and it is not an interview, as are many modern speaking tests. The characteristics of an exam, as well as the stakeholders' understanding of and experiences with those qualities, has been investigated for their ability to contribute to washback effects.

Munoz and Alvarez conducted a 2010 study in which they investigated the washback effects of an oral assessment system composed of various rubrics specifying unique levels of English speaking ability. The researchers found that teachers were able to elicit some degree of positive washback when those teachers were given feedback and support regarding the administration and objectives of the exam, (Munoz & Alvarez, 2010). Just as Qi (2005; 2007) found, it can be important for all stakeholders to be aware of the original objectives of a test in order to better achieve those goals in reality

Cheng's 1999 study in Hong Kong observed and compared three teachers' lessons using old and new test syllabi. Her findings supported Watanabe's study, which concluded that even if tests are created in order to elicit intended washback, the actual resulting effects on teaching methods and activities will vary, and what occurs in the classroom may not undergo much change when compared to adjustments made to the exams themselves. In Cheng's particular study, she found that teachers often changed their content or materials in response

to the exam, but that their methodologies, and the ways they interact with students, remained relatively constant, perhaps due to insufficient teacher training or a lack of ability for the tests to change the ways that teachers perceived teaching and learning (Cheng, 1999). This study lends further evidence that there are likely multiple variables affecting washback that should be explored beyond the tests themselves.

Daly (2011) writes of the negative washback that occurred due to the English testing and curriculum reform that took place in Taiwan. Due to the intense washback effects that have occurred there, he writes that there is a culture of “more teaching, less learning” that is a direct consequence of the English testing that is being done (p. 78). Another side effect of this testing and English education environment, Daly (2011) notes, is that students often seek the services of professional testing preparation institutions in order to best give themselves chances at success.

Another study focused on the effects of a public exam on English teaching in Taiwan (Chen, 2002). The researcher interviewed and surveyed 151 teachers of English at 11 junior high schools regarding their perceptions of a standardized English exam, and how it affected their teaching. The findings of this study showed that the introduction of examination changes strongly influenced teachers’ classroom planning and teaching, with most teachers explaining that they shifted their teaching focus from reading and writing to a more balanced approach, addressing all four language skills (Chen, 2002, p. 14). One explanation for this change in behavior is that the changes made to the test were accompanied by similar shifts in the textbook materials provided to teachers, which suddenly centered more on speaking and listening. This study provided some support for washback effects occurring as a result of examination change, but offered explanations based only on interview and questionnaire data. As washback has been shown to be a complex, multi-faceted phenomenon, it would have been very useful for Chen to use observations to triangulate his findings. The researcher briefly addresses this shortcoming, writing that “although most of the teachers interviewed claimed that they followed the new testing objectives by adherence to a more communication-

based syllabus, it may be interesting to observe to what extent the new syllabus is actually implemented in the classrooms" (Chen, 2002, p. 15).

An interesting finding of Chen's 2002 study was that the exam seemed to have a greater influence on the teacher participants due to its high-stakes nature. The researcher offers some findings regarding teachers' perceptions of teaching and learning, reporting that some teachers expressed stress and confusion as to how best approach the new syllabus, while others said that they had focused their teaching efforts on the highest-achieving students, as these students and their parents tended to be more demanding regarding assessment outcomes (Chen, 2002). The final point here is particularly significant, as it shows the importance of including multiple stakeholders as participants when examining testing impact and washback. While teachers may be directly influenced by the nature and content of an exam, the pressures put on them by administrators, students, and parents can also have influential effects on what occurs in the classroom.

The empirical evidence reported in these studies supports the notion that washback does exist, but the degree to which it occurs will vary from test to test, and across teachers affected by the same test. Furthermore, it is clear that test creators consciously working for washback through examination design may not be able to effectively predict the consequences that an assessment may have on teaching and learning, due to variation among teachers, learners, and education contexts. Moreover, the aims and functions of a high-stakes language test may be in conflict with each other (Qi, 2005). Qi conducted a 2007 study in which she included stakeholder participants from test designer, teacher, and student groups. She found that the washback effects intended by the test creators were not manifesting in English writing classrooms, and that teachers abandoned communicative elements of writing in favor of styles that were more in line with the test and perceived scoring criteria (Qi, 2007). It is crucial that researchers take into account the experiences and views of stakeholders in high-stakes assessments when attempting to draw conclusions regarding washback, and these perceptions should be used in conjunction with class observations, materials analyses, or other sources of data in order to form effective recommendations for change.

Although teachers' perceptions and behaviors have often been at the heart of washback studies, some research has focused more on student experiences. Fox and Cheng (2016) compared experiences between groups of students who had varying degrees of success on the TOEFL iBT. This particular study was limited in that it excluded test scores when interpreting the questionnaire and interview data, and included a much smaller sample of participants who had previously taken the iBT (n=2) than who were characterized as "current test-takers" at the time of the study (n=375). An interesting finding of this study, though, was that beyond the predicted anxiety and stress associated with the actual items on high-stakes examinations, test-takers also shared that they had concerns regarding the computer skills and proficiency necessary to achieve success on the exam. Rodriguez and Allerano (2016) also looked at how students' past test success or failure could influence their exam preparation, and found that students who achieved relative success on high-stakes exams were much more likely to set higher education goals and be motivated to achieve success in their education endeavors.

Zhan and Andrews (2014) similarly focused on student experiences through interviews and diary entries from students participating in a case study as they prepared for the CET4, and exit examination required for bachelor's degrees in China. This study added to learning process inquiry the interesting aspect of students' language identities, and how those influenced their test preparation strategies. Finally, some studies have used only student questionnaire data (Cheng et al., 2014; Pan, 2014) to compare washback effects on learners in different contexts.

3.5.5 Washback Studies in South Korea

Washback research is dynamic and without many overlapping characteristics, but there are contextual elements that may be shared across some research studies, helping us to generalize findings and direct future inquiry. This section focuses on washback studies in South Korea, as contextual elements presented in these papers likely existed within the context of this thesis inquiry, too (see Chapter 2).

Choi (2008) conducted an extensive examination of high-stakes EFL examinations in Korea, and found that most teachers viewed these tests unfavorably, as they were seen to have negative effects on teaching and learning. He writes that the university admissions exam in South Korea has greater washback effects on the education system than any other EFL examination currently available, and that "the English section of the (university admissions test) has such a significant influence on EFL education that virtually all the exams developed at middle and high schools employ almost identical test methods" (p. 42). Prior to middle and high school, however, the controversial inclusion of limited English education opportunities in elementary public schools has contributed to the growth of a private industry of supplemental education that continues to dig into the financial well-being of many, if not most, Korean families.

In this scenario, parents, and particularly Korean mothers, are the major stakeholders in language testing education and preparation, and go to great lengths to ensure that their children are learning test-taking strategies that will help them succeed (Choi, 2008; Park & Abelmann, 2004).The current study proceeded with a keen cognizance of the reality described in the aforementioned Korean studies. While the classrooms at the research site were composed of students that of fairly similar English proficiency, their language education experiences varied as much as their heights, and influenced their attitudes towards English learning, and particularly towards English language testing. A further component of these experiential differences includes exposure to performance-based testing, which is extremely rare in the South Korean education system, but often used in North American schools (particularly in language learning environments). It may have been that students without these testing experiences were more adverse to performance-based testing practices, as these learners would have been unaccustomed to the format and expectations of such assessments.

Beyond the expansive impact that language testing has on South Korean society in general, teaching and learning practices in public schools across the country have also been vastly

affected by major English exams. Kim and O (2002), however, investigated these effects, and wrote that the intentions of the test designers and policy makers may not necessarily materialize in the teaching and learning environment of the classroom. Jung (2008) examined the listening portion of the university entrance examination in South Korea, and found that teachers perceived this test as having both negative and positive washback. The participants in his study reported that they believed the exam effectively provided guidelines for instruction, but that it encouraged teaching to the test. This notion of "teaching to the test," and whether it is exclusively negative, will be an idea that is tested during the current study, and instructors' perceptions of how the test affects their teaching practices will undoubtedly yield important data through which their voices can be heard.

Finally, Hwang (2003) conducted an analysis of textbooks used in third-year English classrooms in Korean high schools, and found that the admissions exam had strong washback effects on the materials used by students and teachers to prepare for the Korean university admissions exam, taken in November of students' third (final) year of high school. This study was unique in that it focused less on the stakeholders of the examination, but on the materials that were developed as a result of test and test policy creation, implementation, and subsequent reformation.

3.5.6 Studies Finding Limited Washback Effects

As washback effects are dynamic in nature, and can present themselves differently based on myriad contextual factors, there are studies that show that washback effects do not exist for some exams, students, and teachers. It is interesting to note that some of the research to follow was conducted by researchers who had either previously found washback effects to exist in different contexts, or observed washback occurring and published those findings in later papers. Cheng (1998) found that on a certain examination in Hong Kong, there was limited washback on students' studying and learning strategies, as well as negligible influence on their motivation to learn English. Cheng (1998) supposed that these effects may have been limited due to the fact that the examination had only been implemented for two years,

and thus the influence of the test perhaps hadn't had enough time to materialize in classrooms, "as is the case with any curriculum innovation" (p. 297) In another case of student motivation having an effect on observable washback effects, Shohamy et al (1996) wrote that the stakes of examines can have a profound impact on the intensity of washback, and that the low-stakes Arabic had differing levels of washback on students, perhaps due to weaknesses in the test design, or student motivation factors.

While some studies have shown great variation in the levels of washback on student processes, many more have reported minimal changes on teaching behaviors and practices, perhaps due to educators being more set in or comfortable with their approaches to teaching. Wall and Alderson (1993) found in their study that teachers showed no changes in their methodologies despite the introduction of an exit exam, although these same teachers reported that they perceived the test to have influenced their teaching processes. In another study, Cheng (1999) saw that teaching content was changed much faster than teaching methods, and that classes that were more teacher-centered before the introduction of the exam remained as such, despite the introduction of more communicative tasks and activities. Qi (2005) wrote of similar observations, providing the explanation that there was incongruence between the intended aims of the test, perhaps due to poor communication between teachers and test designers. Finally, Watanabe (2000) found that teachers were reluctant to make changes to the curriculum, even after test reform had occurred. The researcher proposed that this may have occurred due to cultural or school pressures, or teacher's perceptions of what methods worked best for students.

3.5.7 Methodological Considerations

A number of washback studies have focused on the use of questionnaires to collect large amounts of data from many participants simultaneously. Cheng (1998) administered over 1000 questionnaires to students twice during a 1 year period. While this study lacked strength in the form of mixed methods triangulation, it did use the within-method to improve the validity of the findings, and the large number of respondents helped to decrease possible

error in the questionnaire responses, although this particular research focused only on one group of stakeholders. In this same vein, Herman and Golan (1993) gathered responses from matched pairs of teacher participants in their washback study of standardized testing, but again this data came from only one group of stakeholders. Stecher, Chun, and Barron (2004) similarly used only questionnaires in their washback research, but gathered responses not from student stakeholders, but from both teachers and principals, while Cheng (2004) used within-method triangulation and focused on teachers' perceptions and behaviors through the use of questionnaires. It becomes clear from these and other studies that questionnaires can be a valuable part of washback research, but questionnaire data alone is severely lacking in how effective it is to describe washback effects (Bailey, 1999).

In order to collect information-rich data that most accurately represents the unique cases of all participants, interviews are often used in washback studies either alone (Qi, 2004) or in conjunction with classroom observations (Watanabe, 2004; Wall, 2005; Alderson & Wall, 1993). Bailey (1999) and Alderson and Wall (1993) advocate for the inclusion of mixed methods washback research, and specifically mention the importance of classroom observations and interviews. Alderson and Wall write "that it is increasingly obvious that we need to look closely at classroom events in particular, in order to see whether what teachers and learners say they do is reflected in their behavior" (1993, p.127). Bailey (1999, p.36) refers to the combined use of interviews and observations in washback studies as "watching and asking," where researchers should strike a balance between watching and asking about teaching and learning behaviors. Interestingly, although Alderson and Wall (1993) push for the use of observations in washback studies, they also acknowledge that observations alone simply allow the researcher to see and hear *what* is occurring, but not *why* it is occurring. Thus, it becomes clearer that not only multiple methods of a quantitative or qualitative nature should be used, but that a mixed methods approach is critical to forming a complete understanding of washback effects.

3.6 Summary

This chapter presented a framework to guide my investigation into the washback effects of the GMATE. First, this study was critical in nature, and sought to improve the specific testing situation in which the GMATE was being administered. This meant that I embraced my own judgements regarding the test, and remained cognizant of them while problematizing the testing policy and suspected washback effects. Washback effects themselves are defined in several ways and described by multiple models, due to the fact that they are dynamic in nature, and can vary between and within individuals depending on myriad factors including but not limited to the educational setting, assessment, background of the participants, and stakes of the exam investigated.

However, exit examinations and policies requiring their administration in secondary and tertiary education have been examined thoroughly throughout the literature, with particular attention paid to those graduation tests related to English language proficiency, and the resulting test preparation behaviors of students and teachers, collectively defined as "washback effects" in several papers, although the manners in which these behaviors manifest are as diverse as the contexts in which they occur. Washback studies, therefore, have appropriately been undertaken in regions around the world, the findings from which have added to the greater understanding of these complex phenomena. Through the lens of past studies, I strived to find commonalities in the theoretical and methodological underpinnings of the available washback research in order to make effective and appropriate decision regarding my own methods and study design, outlined in the next chapter.

Chapter Four: Methodology

In this section, I begin by introducing the paradigmatic framework under which I conducted this research, after which I describe the case study, and give the reader evidence for why this study's research aims required such an approach. I then present my main research questions and sub-questions, along with explanations of each, and describe the qualitative and quantitative research methods used in this thesis, while providing justification for these choices. Finally, I outline the research procedures, planned and undertaken, along with acknowledgements of methodological limitations and ethical considerations.

4.1 Theoretical Framework and Methodology

Ernest (1994, p. 8) regards educational research as a "systematic inquiry with the aim of producing knowledge," while Anderson & Arsenault (1998) refer to research in education as "a disciplined attempt to address questions or solve problems through the collection and analysis of primary data and the purpose of description, explanation, generalization, and prediction." Furthermore, Burns (1997) repeats the notion of research as being *systematic investigation,* while Mertens (2014) echoes the aforementioned views of research as being efforts to "understand, describe, predict, or control."

This focus on *systematic* and *disciplined* efforts in these definitions is a testament to how crucial it is that researchers be clear about from which philosophical framework they are approaching their study, and that they be faithful to those ideas that have informed their work. These philosophical belief systems, or paradigms, are theoretical in nature (Bogdan & Biklen, 1998), provide the "motivation for undertaking a study," (Cohen, Manion, & Morrison, 2007) and influence the way that knowledge is created and understood. Indeed, it is the researcher's choice of paradigm that "sets down the intent, motivation, and expectations for the research, (for) without nominating a paradigm as the first step, there is no basis for subsequent choices regarding methodology, methods, or research design" (Mackenzie & Knipe, 2006).

This thesis research is conducted under the critical, or transformative paradigm, allowing for a fairly broad range of both qualitative and quantitative research methods. However, as noted above (Mackenzie & Knipe, 2006), the selection of research methods in any given inquiry is not always marked by such freedom. The research methods employed in qualitative and quantitative research are often regarded as incompatible and "belonging to," or falling under certain theoretical frameworks (Lincoln & Guba, 1985). Generally, quantitative methods concern the strict quantification of data and the careful control of empirical variables, while qualitative methods refer to a broader class of empirical procedures "designed to describe and interpret the experiences of research participants in a context-specific setting" (Denzin & Lincoln, 2000).

O'Leary (2004) presents qualitative and quantitative as

> "adjectives for types of data and their corresponding modes of analysis, i.e. qualitative data – data represented through words, pictures, or icons analyzed using thematic exploration; and quantitative data – data that is represented through numbers and analyzed using statistics" (p. 99).

This focus on the collected data and the tools used to obtain that data shows that these terms don't deserve assignment to one paradigm or another. Instead, we see research methods as traditionally "qualitative," such as interviews, case studies, and observations, and others as "quantitative," for example, experiments, tests, and scales (Mackenzie & Knipe, 2006). While these data collection tools are viewed as stand-alone research methods themselves, they are most effectively regarded as integral pieces of the strategies that employ them (Creswell, 2009).

4.2 The Case Study

When looking at the research aims of this project, it became clear that a case study design was most appropriate. Case studies allow researchers to focus on bounded systems and

specific instances, which in this thesis translate to the English department within this specific university context. Cohen, Manion, and Morrison (2007) provide a brief summary of several "hallmarks" of case studies (p.253):

> -They are concerned with a rich and vivid description of events relevant to the case
> -They blend a description of events with the analysis of them
> -They focus on individual actors or groups of actors, and seek to understand their perceptions of events
> -They highlight specific events that are relevant to the case

It takes but a brief glance at the aforementioned characteristics of case studies to quickly see that these notions match up very well with my research questions. Describing teachers' and students' perceptions, behaviors, and feelings as they relate to the specific context of the research is a repeated theme across all of the research questions, and is depicted quite clearly as an important defining element of case study research. Although case studies deal with focused, geographically bounded, and well-defined contexts (Hitchcock and Hughes, 1995), generalization is possible as long as these ideas are strongly supported by the evidence and collected data. A possible generalization from this study, for example, might include Korean universities that employ language exit exams, or more focused still, that use English language exit exams.

Interviews and observations are closely tied with case study research, but in this particular study, it was useful to gain further insight through questionnaires. While there are some that have proposed that questionnaire and case study research methods are mutually exclusive, Gable (1994) provides compelling evidence that they are in fact complementary, and that the inclusion of both methods can effectively strengthen a study overall. Some strengths of case studies are that the results can be more easily understood by non-academics, they provide insights into other similar situations and cases, and the research can be done by a single researcher (Cohen,

Manion, and Morrison, 2007). These strengths were very important to me, as I aimed to share my research with policy makers that may not have had a deep understanding of language testing and washback effects, and I was not able to hire a team of assistants or researchers to help with my study. Moreover, the questionnaire allowed for a higher degree of control, repetition, deduction, and generalization, which Gable (1994) explains might be less so for case studies.

Because of the aforementioned complexities that contribute to washback effects, it is crucial to collect different types of data through varying methods before drawing conclusions. In order to most accurately describe phenomena and tell the complete stories of research participants, a mixed- or multi-method approach is often used. This results in *triangulation*, and is an important element in qualitative research used to provide evidence of validity (Cohen, Manion, & Morrison, 2007). Triangulation is highly recommended in washback studies (Cheng, Yoshinori, & Watanabe, 2004; Wall, 2005; Wall & Anderson, 1993), and can exist in several forms. Bailey (1999) lists four types of triangulation (p.38):

> "The first is data triangulation, in which data from more than one source are brought to bear in answering a research question (e.g., the data from teachers, language learners, and inspectors in the study by Shohamy et al., 1996). Second, investigator (or researcher) triangulation refers to using more than one person to collect and/or analyze the data. In theory triangulation, more than one theory is used to generate the research questions and/or interpret the findings. Finally, in methodological (or technique) triangulation more than one procedure is used for eliciting data- for instance, Wall and Alderson's (1993) use of interviews and classroom observations."

Furthermore, methodological triangulation can be referred to as *within-method,* in which the same research method is used at different times, or *between-method*, which uses different methods (sometimes called *mixed methods*) (Brannen, 1992).

Examples of this mixed methods approach to research are the positivist researcher who codes and quantifies transcripts of a series of interviews in order to analyze them using statistics software, or the critical theorist who interprets the numerical responses of a questionnaire from a position seeking justice and emancipation for those in a problematized situation of power inequality. “While some paradigms may appear to lead a researcher to favor qualitative or quantitative approaches, in effect no one paradigm actually prescribes or prohibits the use of either methodological approach” (Mackenzie & Knipe, 2006). This thesis uses questionnaires, interviews, and classroom observations to satisfy the need for triangulation and high levels of concurrent validity. Following sections will discuss rationale for the choice of these specific research methods.

4.3 Research Questions

1. What are the perceived washback effects of the GMATE on teachers’ teaching?

The aim of this research question was to both understand what was occurring from teachers’ perspectives, but also to help make teachers more aware of their teaching practices through self-reflection. This was accomplished through a combination of interviews and questionnaires, and I hoped that teachers not only became more cognizant of the effects of the GMATE on their instruction, but also that they became more empowered through participation in the study. Freedom to teach how and what teachers’ felt was necessary for students to succeed, ideas of professionalism, and the value teachers placed on the exam and test preparation were investigated here.

1.1 To what extent do teachers feel differently about the GMATE, and how does their teaching and test preparation vary depending on the English proficiencies of their students?

1.2 To what extent do teachers feel differently about the GMATE, and how does their teaching and test preparation vary when their classes include students for whom the GMATE is a high-stakes exam?

1.3 What are the observed washback effects of the GMATE on the classroom environment?

2. What are the perceived washback effects of the GMATE on students' studying?

Language testing is awash in complex moral and power issues (Johnston, 2003; Shohamy, 2001), yet students are often not included in much of the decision making process when tests are implemented. As the first research questions strive to empower teachers and make them aware of their own teaching while also providing me with data regarding washback effects on teaching, this second set of research sub-questions give students a voice concerning English language testing in the research context through questionnaires. These questions were used to try and identify differences in motivation, stress, and attitudes towards English courses and exams as expressed by the learner participants.

2.1 To what extent do students of different English proficiencies feel and act differently towards English courses and the GMATE?

2.2 To what extent do students for whom the GMATE is a high-stakes exam feel differently towards English courses and the GMATE than those students for whom the GMATE is low-stakes?

2.3 To what extent does the length of term influence students' approach to English language learning and test preparation?

4.4 Data Collection Methods and Instrumentation

In the current study, important value-driven notions of empowerment and identity (Shohamy, 2001) helped to critically analyze the content, procedure, and use of the GMATE assessment. Beyond the exam itself, however, were the stakeholders in the test, and it was their voices that provided the greatest insight concerning the effects, whether positive or negative, which occurred as a direct or indirect result of the language test used in this particular Korean university. However, there does not exist an accepted standard set of instruments used in washback studies (Bailey, 1999). Questionnaires vary greatly from study to study due to the numerous contextual differences between schools, classrooms, and exams, and additionally these items are often administered in the native language of research participants, making accurate translation and adoption a challenge (Bailey, 1999).

Furthermore, accepted observation schemes, interview formats, and interview questions have yet to emerge, creating the need for researchers to adapt existing tools to their unique environments, while also creating and testing new items (Cheng, 2008). This study followed a mixed-methods design, as it combined analysis of both quantitative and qualitative research methods. The following subsections outline in detail each research method, the research-based evidence supporting the use of said method, and the instrumentation that was used to assist in the collection and management of data.

4.4.1 Questionnaires

To build a jumping-off point for this research, I chose to use separate questionnaires for teachers and students, in the hope that their responses would help steer the development of subsequent data collection tools (interview guides in the case of teachers, and different forms of questionnaires for the students). Questionnaires offer distinct benefits for researchers, allowing for the collection of large amounts of data, while also yielding structured data that can be analyzed easily to create generalizations and draw conclusions (Cohen, Manion, & Morrison, 2007). Furthermore, questionnaires allowed me to be absent from classrooms

while students took the questionnaires, making it much easier to collect data from a large number of separate classes spread over campus locations and class times. Also, questionnaires can be easily administered in participants' native language, which aids in avoiding participant confusion and difficulties researchers may have when it comes to interpreting responses without follow-up questioning (Wall, 2005, p.28).

While there are advantages to using questionnaires in washback studies and education research, there of course exist numerous drawbacks, as well. First, while questionnaires provide an easy method for collecting a comprehensive sample of data that could represent the larger population, this data can be somewhat limited in nature, due to the use of predetermined questions and response choices (Brown, 2001; Cohen, Manion, & Morrison, 2007). Moreover, as there is currently no accepted standard for questionnaires in washback studies (Bailey, 1999), I would have had to take the time to adapt existing questionnaire items or develop a completely new instrument, which would have then required additional time to be piloted and analyzed. Taking into account these disadvantages, along with time and human resource limitations, I decided that questionnaires could be used as a tool to inform other data collection methods, as well as to provide triangulation.

There are many factors that must be taken into account when designing research questionnaires. The type of questions, scales, and nature of question items must all be considered in order to collect quality, accurate data while still adhering to ethical standards by avoiding question that may distress or upset participants (Cohen, Manion, & Morrison, 2007, p.333). Furthermore, while questionnaire instruments are sometimes adopted directly for subsequent research studies, contextual differences and even the most minute of item changes can drastically affect the validity and reliability of the questionnaire (Creswell, 2009, p.150). With these caveats in mind, I decided to adapt a questionnaire used in previous washback studies (Chen, 2002) for teachers, as this questionnaire was to be more fluid, and changed depending on questionnaire and interview responses, as well as on what I observed in the classrooms. Although Chen (2002) extensively analyzed her questionnaire for validity and reliability, I was unable to use her questionnaire in its entirety, due to contextual

differences and the inclusion of some irrelevant items for my study. Therefore, the questionnaire was piloted and checked to ensure it was valid and reliable.

For the students, however, I chose to design a new questionnaire tool, as the student participants were fairly homogenous themselves (see section 4.5.2), but exhibited very different characteristics from student participants in past washback studies. However, although there were contextual differences between my context and those examined in the literature, there were ideas fundamental in washback studies that carried over and found themselves in the student questionnaire instrument. Because students were not interviewed in this research, it was crucial that the questionnaire instrument developed for this research was effective in gathering a large amount of accurate data to help draw conclusions and make generalizations, where appropriate.

In order to avoid collecting an overwhelming amount of text data, which would have subsequently required time-consuming coding and analysis, I decided to use closed rather than open questions. Although open questions would have provided me with an enormous wealth of data unique to my participants and context (Cohen, Manion, & Morrison, 2007, p.321), the questionnaires in this study were used more to inform the observation and interview components, which yielded significant qualitative data themselves. Furthermore, the student questionnaires were administered in Korean, and responses to open questions would have required much more translation work, which would have added strain to my research resources. Furthermore, based on past washback studies and my own research experiences, I decided to use a Likert scale for the student questionnaire, while the teacher questionnaire included Likert scale items in addition to some open questions. The teacher questionnaire was administered entirely in English, and with a much smaller sample of participants, reduced the need for time-intensive coding and translation work.

After deciding to use a Likert scale questionnaire for the student and teacher participants, I then had to make formatting choices regarding the responses and number of items. The literature shows that it may be desirable to remove the mid-point option from a five- or seven-

point scale when the item may require participants to give answers that are not socially acceptable (Garland, 1991). For example, students in the current study may have been wary of expressing negative feelings towards the exam with the university and teacher in mind, or conversely may have been averse to supporting the GMATE if they were aware of their classmates' conflicting views. Additionally, the literature (Lee & Green, 1991; Cohen, Manion, & Morrison, 2007) shows that Koreans, as well as many East Asian research participants, show a central tendency when responding to questionnaires while avoiding extreme responses, and one widely-cited study reports that fewer response options may in fact be optimal in situations that are time sensitive, or in which participants may become frustrated with the time-consuming nature of the questionnaire instrument due to a multitude of choices (Preston & Colman, 2000). Although the teachers in the study were of Western origin, both the number of response choices and questionnaire length can influence participant response patterns (Burchell & Marsh, 1992), thus I chose to remove several items from the Chen questionnaire that were not only deemed irrelevant to my research context, but also seemed somewhat redundant when looking at the questionnaire as a whole.

As useful as questionnaires can be in education research, they are not without their shortcomings. Cohen, Manion, and Morrison (2007, p.327) mention several drawbacks of questionnaires, including a lack of equal intervals between responses, no practical method to check that participants are being truthful, and a lack of opportunity for the respondents to qualify or add comments to their responses. Additionally, there is a great number of resources that provide direction for questionnaire item creation, along with do's and do not's. For example, because of the unidimensionality of the Likert scale, items must be carefully worded so that they do not measure more than one thing at a time, and the semantics involved in actually wording the items can further confound the results of questionnaire research (Cohen, Manion, & Morrison, 2007). Brown (1997) additionally provides some guidelines for item creation, writing that items should be succinct, clear, and free of bias.

4.4.2 Interviews

To better describe the unique aspects, attitudes, and behaviors of the participants in this research, I chose to interview teachers at several stages of Phases I and II (conducted during the regular fall and winter intensive terms, respectively). Interviews are very useful in case studies and washback research, as they not only provide a means of collecting information-rich data, but they also provide opportunities for the researcher to respond to participants in real time, allowing for verification, clarification, and follow-up questions. In this regard, interviewers and interviewees are working together to co-construct the interview itself, as the information exchanged flows in both directions (Walford, 2001). Furthermore, interviews help to identify human participants as the generators of knowledge, and that the social and cultural context within which the participants live and work influences interview data (Kvale, 1996).

Patton (1987) outlines the characteristics, strengths, and weaknesses of three basic types of qualitative interviews. *Informal conversational interviews* are more natural interviews that are not necessarily conducted, but that occur in immediate context. The less formal nature of these interviews means that interviews are well matched to unique participants and circumstances, but these interviews are less systematic, and researchers may lose out on opportunities to ask important questions if these ideas don't present themselves in the natural course of conversation. The *interview guide approach* is marked by topics that are specified and outlined in advance, which helps improve the comprehensiveness of the data, although these interviews are still of a fairly conversational nature. Unfortunately, the interview questions are often sequenced or worded in such an unstructured way that participant responses can be difficult to compare and contrast. *Standardized open-ended interviews* require a more precisely worded and sequenced set of interview questions, in contrast with the interview guide approach. Standardized open-ended interviews provide strengths in that the responses can much more easily and reliably be compared, as interviewer effects are effectively reduced. However, because of the rigidity of this format, some questions may be posed unnaturally, and may be irrelevant for certain participants or circumstances.

After weighing the benefits and drawbacks of each interview type, I followed in the footsteps of many education and washback researchers and chose to use the interview guide approach. This type of interview was most appropriate for my study, especially because it best responded to the uniqueness of each participant, and the dynamic nature of washback effects on different classes. Indeed, as questionnaire responses varied from one teacher to the next, so did the interview guides slightly differ from one another each interview period, depending on participant responses and what had been observed in classrooms (see Appendix C). Cohen, Manion, and Morrison (2007) write, "The more one wishes to acquire unique, non-standardized, personalized information about how individuals view the world, the more one veers towards qualitative, open-ended, unstructured interviewing" (p.354). As the questionnaire answers helped me drive my interview and observation methods, the interview responses allowed me to connect the 'why' to the 'what' in classroom observations. Without the valuable data provided through periodic interviews, I would have been left with only an uncaptioned snapshot of the language learning environment in this particular context.

4.4.3 Classroom Observations

Questionnaire and interview data can be combined to form a fairly clear notion of what is occurring in a washback study, but they continue to rely on the accuracy of participants' accounts, and can sometimes fall out of line with what is actually presented in the classroom (Robson, 2002). Therefore, Alderson and Wall (1993) suggest that it is important to include classroom observations and interviews both when undertaking washback research. Observations provide data that is contextually sensitive, and allow researchers to notice things they may have missed, as well as things that may not have come to light during questionnaires or interviews (Cohen, Manion, & Morrison, 2007). Because of the varying nature of washback effects, it is crucial that these nuances are included in washback research. However, because there is no accepted standard observational instrument for washback studies, it can be difficult to create, pilot, and employ novel observational schedules and instruments for each context.

Some washback studies have used or adapted the COLT (Communicative Orientation of Language Teaching) scheme (Burrows, 2004; Green, 2007). In these studies, though, the COLT scheme was molded in some ways in order to better fit the context and characteristics of washback effects. In this thesis, however, I was interested in both differences and the washback effects the GMATE had on teachers and students, but the COLT scheme still did not provide a perfect observation tool even in its entirety. Green (2007) explains that the COLT may be more appropriate for observing classes in teacher-centered, rather than student-centered classrooms. In my research context, however, I anticipated teacher and classroom differences, and therefore some elements of the COLT were incorporated into the observation scheme. Furthermore, the GMATE is composed of four distinct items (introductions, picture description, process explanation, and providing advantages and disadvantages of a subject; see Appendix F), so these were included as separate activities on the observation sheet. However, because of the rich mix of unique contextual factors influencing classroom behaviors, I decided to develop and adapt the observation tool based on data collected from previous washback studies along with my own questionnaires and prior observational data (see Appendix D).

Observations were conducted in specific weeks (Phase I) or on certain days (Phase II) outlined in Table 4.3. 8 classes lasting approximately 70 minutes each were observed during each observation week during Phase I. Because of limited resources and time, only some observations were conducted in a direct manner; I was able to observe these classes in person while the class was simultaneously being recorded for later confirmation. Other classes were observed indirectly, due to time conflicts, thus these were recorded and viewed as soon as was possible after having taken place.

I had hoped to rotate which classes were observed only indirectly in order to decrease classroom and teacher behavioral differences that may have occurred due to my presence (a phenomenon referred to as reactivity), as well as to attempt to minimize observer bias as it relates to attention, construct validity, and data or

memory selection (Cohen, Manion, & Morrison, 2007, p.410). However, due to students' heightened anxiety regarding the presence of video cameras, as a result of a string of scandals throughout South Korea at the time related to hidden cameras, many students and a few teachers declined to be videotaped, which decreased the number of observable classes. This is discussed in more detail when addressing research limitations in 4.9.1.

4.5 Participants and Selection

The current study aimed to examine the washback effects of the GMATE test, but it should also be noted that these effects were likely to vary from teacher to teacher, depending on the educators' unique personalities and identities. The educators participating in this study were all of different ages, educational backgrounds, and experience levels, and of course carried with them different teaching method preferences and perceptions of testing and teaching. All of these factors worked together to form their identities as teachers, and this most certainly influenced the findings of the current study. In this same vein, the current study did not only investigate the washback effects of the GMATE on students in the university, but also compared differences that may have occurred regarding these effects between students that were subjected to the GMATE as a high-stakes exam, and those who were fortunate enough to be preparing for the test as a low-stakes exam, due to policy changes enacted in 2015.

In the current study, it may have been that the students for whom the GMATE was a low-stakes exam followed a pattern of behavior previously found by Shohamy (1993), and could have been expected to earn lower scores than their older classmates of similar ability subjected to the same exam in its high-stakes format. Because washback effects are highly dependent on contextual and participant factors (Chen, 2002; Cheng, 1997; Cheng, 1998; Cheng, 1999; Tsagari, 2009; Watanabe, 2000), it is crucial to select a large enough participant sample to capture these individual variations and properly represent the larger population of stakeholders (Patton, 1987). Simultaneously, due to logistical constraints, the participant sample must not be so large that the data is overwhelming, thus making it difficult

to interpret and analyze the important nuances that exist, particularly within the qualitative data yielded through interviews and classroom observations (Perry, 2005).

While research conducted using quantitative methods often refers to individuals as subjects, qualitative research generally views these people as participants in the research project, reflecting the more participatory worldview held by those researchers operating under paradigmatic frameworks that support constructivist, participatory knowledge claims (Creswell, 2009). Furthermore, quantitative research often employs sample selection that prioritizes the representativeness of the target or general population. That is, sampling is done in a way that is appropriately random, and allows the researchers to better and more accurately generalize their findings, referred to as external validity (Perry, 2005). Conversely, qualitative research is interested in collecting a wealth of high quality data, even if this comes from a small number of cases or participants (Gall, Borg, & Gall, 2006). These somewhat competing ideas are collectively referred to as "sampling paradigms," with strategies used to maximize information labeled the "information-rich paradigm," and the complementary sampling framework focused more on generalizability called "the representative sampling paradigm" (Perry, 2005, p.59).

The current study, being mostly qualitative in nature, strived to find a balance between these two paradigms. The use of mixed methods aided in this pursuit, as questionnaires allowed for a larger number of student participants, and thus more accurate generalizations were drawn, while classroom observations and interviews provided the opportunity to collect information-rich data that is so crucial to conducting good qualitative research (Alderson & Wall, 1993; Bailey, 1999). Because the research methods in this study were employed with the distinct aim of collecting a wealth of quality data, it was thus necessary to purposefully select participants that were able to provide diverse perspectives and experiences, despite not being representative of "typical" cases, keeping in mind that in washback research, the complex nature of the phenomena allows for a wide range of successful research methods to be used (Watanabe, 2004). The selection of participants can, and should when necessary,

be changed as the research progresses, assisting the researcher in broadening or focusing the scope of information being collected.

To best achieve both objectives of representativeness and information-rich data, a combination of sampling strategies was employed in the current study. In order to allow for the generalizing of data interpretations, analysis, and generalizations to the greater population of university students in the university English program, proportional stratified sampling was used in part in the current study (Perry, 2005). The English program is composed of beginner (level 1), intermediate (level 2), and advanced (level 3) students, as placed by an abbreviated form of the GMATE, called the General English Language Test (GELT), which uses the same rating criteria as the GMATE, but is administered in a much lower-security fashion, and does not include the use of the third-rater system (see section 2.8). This greatly reduces the accuracy and reliability of the placement test, but according to instructors at the university, has been a marked improvement over the previous class groupings by major. English courses were made up of 12 to 18 students in total, and the program on the whole was composed of 32 classes of level 1 students, 55 classes of level 2 students, and 23 classes of level 3 students. To get a proportional, stratified sample of students in this study, 5, 9, and 4 classes were selected from levels 1, 2, and 3, respectively, from all teacher volunteers (purposeful convenience sampling used here). During the intensive term, there were far fewer courses offered, so participant numbers were adjusted accordingly.

4.5.1 Teacher Participants

Phase I Participants

24 English instructors responded to my request for questionnaires, and from those 24, 18 were purposefully chosen for interviews and observation research participation, with 17 completing the study. Teacher participants were originally from a variety of countries, including Canada, the U.S., England, New Zealand, South Korea, and Australia, and had

been teaching in Korean universities for an average of between 8 and 9 years, with an average of close to 5 of those years in their current position. They ranged in age from 33 to 64 years old, with an average of 39 years old. All but one of the participants was a native speaker of English. While I had hoped to recruit more Korean teachers of English in order to compare any differences that might exist between those two groups of participants, there were very few to contact, and most were part-time adjunct faculty, which meant they generally traveled around the greater Seoul area teaching at multiple universities, making interviews more difficult. All but one of the participants held graduate degrees, with 12 participants holding master's degrees related to TESOL, and another 3 participants holding certificates in TESOL (such as the CELTA). 15 of them were teaching only one level of classes that semester, while the other 3 taught adjacent levels (levels 1 and 2, or 2 and 3).

Phase II Participants

With only 7 total speaking courses being offered during the intensive period, I contacted all 7 teachers to participate in this phase of my research, and all but one agreed. There were 3 level 1 classes, 2 level 2 classes, and 1 class of level 3 students. Because level 1 students more often fail the GMATE than higher level students, more level 1 classes were offered during the intensive period. The 6 instructor participants here were from Canada, the U.S., and England, and ranged in age from 33 to 40 years old, with an average of just over 36 years old. They had been teaching in Korean universities for an average of 7 years, with just under 5 of those spent in their current position. All 6 of the participants held master's degrees related to TESOL, and one of them also held a certificate in TESOL.

4.5.2 Student Participants

Phase I Participants

As the general English courses are often referred to as "freshman English" by teachers and students alike, it wasn't surprising that a great majority of the student participants in the

regular term were in their first year of university study. 268 (77%) of the students were freshmen (first year), 35 (10%) were sophomores (second year), 13 (4%) were juniors (third year), 27 (8%) were seniors (fourth year), and 4 (1%) were in their fifth year or later. I was informed some students put off taking their required English classes until the last possible moment, providing a possible explanation for the larger number of seniors than juniors in this sample.

Following this trend, participants' western ages are generally younger, as is shown in Table 4.1. Korean age is calculated in a different way than is western age, but Koreans are well aware of both, so I was not initially worried about inaccuracies here. However, the data shows a larger number of 19-year-old students than one would expect in Korea, as students generally don't delay their university admission after graduating from high school, so age was not used as a reliable factor when investigating trends and relationships in the data.

Table 4.1 Phase I Student Participant Western Age

Age	18	19	20	21	22	23	24	25
N	30	155	92	30	13	14	9	4

Although the research site is considered an upper-tier university in Korea, and thus attracts mostly students who have achieved higher scores on the national entrance exam, the English levels of these students, as measured by the university placement exam, varied greatly. There were 59 (17%) level 1 students, 212 (61%) level 2 students, and 76 (22%) students from level 3.

Phase II Participants

According to the instructors, students taking classes during the intensive term generally fall under one of several categories. Some are taking English for the first time after having postponed it for various reasons, such as to focus on their major-related courses, or because of fear of failure, others are retaking the course after

having failed it the first time, or in hopes of replacing the first score with one higher (grades are often included on job applications), and still others take the course during the intensive period because instructors “have a reputation of grading easier, giving less homework, or helping students pass the GMATE” (Alex, Int1) during this time. Of the 122 students taking English during the winter intensive period, only 2 (2%) were freshmen, 14 (12%) were sophomores, 36 (30%) were juniors, 58 (48%) were seniors, and 12 (10%) were in their fifth year of university study. Thus, for 98% of the students in this phase of the study, the GMATE was a higher-stakes test that required them to achieve certain minimum scores in order to pass the course and graduate from the university.

Appropriately, student participants taking courses during the winter intensive period were older, with no students aged 18 or 19 years. Participants’ western ages are shown in Table 4.2 below.

Table 4.2 Phase II Student Participant Western Age

Age	20	21	22	23	24	25
N	8	22	34	22	16	20

Contrary to the strict class size limits used during the regular term, intensive classes have a higher, more flexible class cap. There were 72 (59%) level 1 students in 3 classes, 30 (25%) level 2 students spread across 2 classes, and 20 (16%) level 3 students in a single class.

It should also be noted that in this study, students and teachers were of very different backgrounds, and thus likely held more varying views of education, testing language learning, etc. based on their unique experiences. Some of these possible differences were outlined in the second chapter of this thesis, and will be addressed in more detail in later chapters, as well.

4.6 Data Collection Procedures

This research was conducted in two phases. Before beginning these main phases, though, I piloted the teacher and student questionnaires and teacher interview questions during the second week of the term, when little to no teaching was taking place. This allowed me to analyze the questionnaire for validity and reliability, clarity, and usability, and make changes as needed.

The student questionnaire was piloted to 3 classes of 18 students each, of levels 1, 2, and 3. The questionnaire was administered in students' L1 of Korean in order to avoid code-switching and difficulties understanding items (especially for level 1 students). Questionnaires were done on students' smartphones, as the questionnaire software was optimized for such administration, and according to data provided by the software, the questionnaire took an average of just under 4 minutes. Afterwards, students were asked if there were any items they felt were difficult to understand, unclear, ambiguous, or irrelevant, but no items in particular were pointed out by the student participants, and thus the questionnaire used in the pilot was used also in Phases I and II.

In the piloting stage of the study, 6 volunteer teachers took the questionnaire and met with me to discuss the individual items that needed clarification or explanation. Some of the items were found to be slightly ambiguous to several of the participants, due to the wording used. For example, item 21 on the teacher questionnaire stated, "I usually use the textbooks because they cover the topics on the GMATE." Several participants mentioned that this was confusing, as they indeed do usually use the textbooks, but for varying reasons. One teacher expressed, "I use the textbook because the students complain about it if we don't. There's some good stuff in there, but I would prefer my own materials if I could."

Two other items on the questionnaire, items 49 and 50, used words to describe emotions, "embarrassed" and "guilty," respectively, as responses to students performing poorly on the GMATE. Pilot stage participants asked questions about this item, with one clarifying, "I don't

feel *guilty*. I just feel bad, because I know that they've (the students) worked hard all term, and I know they can do better." These items were amended based on participants' feedback, and pilot response data was not included in the findings and discussion chapters that follow. The questionnaire software showed that the questionnaire took an average of 18 minutes to complete, and although this was quite a bit longer than optimal, the participants were very motivated participants, and thus I wasn't worried that this would affect dropout rates.

I piloted the interview questions with 3 pilot study participants, in large part based on their responses to the questionnaire. As I interviewed participants, I stopped periodically to ask them if questions were clear and concise, and also took my own notes regarding participants' answers and questions to me in order to better create questions that best captured participants' experiences, attitudes, and feelings. The final version of these preliminary questions can be found in Appendix C. It should be noted that the interview question guides diverged quickly between participants, depending on their questionnaire and interview responses, and whether or not they were participating in the observation component of the research.

The first data collection phase (Phase I) was approximately 14 weeks in duration, and used questionnaires, interviews, and classroom observations to gather data regarding what washback effects may have been occurring during regular semester courses, which are generally taken by first- and second-year students who have not previously taken the course or GMATE exam. This schedule is shown in Table 4.3.

The second data collection phase (Phase II) was 20 days in length, and used the same research instruments to identify, compare, and contrast any washback effects that might have occurred during "intensive" English courses, which are only offered during the summer and winter vacation periods, and tend to be composed of older students (third and fourth year students) and those students who have previously failed the GMATE. The intensive English courses are designed to follow a similar syllabus as the courses offered during the academic

term, with the major difference being that a week's worth of material must be covered per day, as the 15-week term is taught over a period of 3 weeks (15 class days).

Other than the different length of term and intensity of study, the teachers, levels, and materials were generally shared between the regular and intensive courses (fewer teachers were involved in the intensive courses, as fewer courses were offered). Student questionnaires and teacher questionnaires were administered electronically using the online questionnaire host, QuestionPro, an online questionnaire tool, as this allowed me to create and administer questionnaires to a large number of students in electronic form, due to its ease of use on mobile devices such as smart phones and tablet computers. Furthermore, QuestionPro provided a simple and secure method for organizing and analyzing data, as well as exporting this data to SPSS for more detailed analysis (see section 4.4.1 for more info regarding the questionnaire instrument).

I chose the aforementioned schedules of data collection to allow me to make slight changes to questionnaire items and interview questions based on what I heard from respondents and saw during observations. Alderson and Wall write, "Without observations, researchers are unlikely to know what questions they should be asking and may not understand (or be sufficiently critical of) the answers they are given (1993, p.65). By scheduling brief intervals of time between questionnaire and interview dates, I hoped to be able to reflect on what I saw and heard during observations to mold the questions asked, while also changing the specific classroom behaviors I more closely focused upon based on the answers given during interview and questionnaire collection periods.

However, due to logistical difficulties that arose just weeks before the data collection period commenced, I was forced to adjust my observation aspirations, and focused only on four instructors who fulfilled the requisites of teaching at non-conflicting times during the day, provided consent, and whose students also provided consent. Of these four teachers, one allowed me to record her classes on video for later referral and confirmation of my field notes.

Lastly, educational research can be affected by seasonality. Freeman (1996) writes of seasonality as the natural shifts in teaching and learning behaviors that occur throughout a school year, due in part to scheduled exams, vacation periods, etc. Watanabe (1996) also identifies seasonality as a possible influence in washback studies, suggesting that two teachers interviewed at different times of the academic term or year might give very different answers or exhibit contrasting behaviors based in part on the inherent contextual differences that occur throughout an academic period. One study showed that teachers reported that their teaching "intensified" as testing periods approached (Shohamy, Donitsa-Schmidt, & Ferman, 1996), while another (Shohamy, 1993) reported that washback effects vary greatly over time. For this reason, the final three weeks of the term involved more data collection, as the literature suggested that there may have been more observable washback effects closer to the GMATE testing period (week/day 15). By spreading out observations, questionnaires, and interviews over the academic terms, I hoped to identify some of the effects that seasonality may have had on washback effects, and to discuss those when answering research sub-question 1.3.

4.7 Data Analysis

The questionnaire data collected through electronic questionnaires was analyzed using the applications built into QuestionPro, in addition to SPSS Version 18. Descriptive statistics were used within samples (Phase I or II), and inferential statistics were used to identify significant patterns of similarities or differences between the participant groups in phases I and II. For the teacher's questionnaire, categories of questions were separated, and relationships between categories were investigated, but this data was used only to build an informed and unique interview guide for each teacher participant, and is not discretely reported in the findings. These analyses allowed me to better focus both my interview questions and observation scheme, which were coded and examined for patterns and trends between teachers, classes, levels, and terms of study. These findings and analyses are presented in the next chapter with specific justification for data analysis techniques described below.

Cohen, Manion, and Morrison (2007) present five distinct ways of organizing and presenting data analysis, with support for each. In this study, it made sense to me to incorporate elements of several of these presentation methods, as I organized my findings by research question, yet within each question (representing a distinct issue related to washback) and question group (representing in cases a unique participant group – either teachers or learners), I chose to further separate the analysis by instrument. Finally, I compared data collected through the questionnaires and observations with interview data, allowing me to match and contrast participant accounts and perceptions of what was occurring within the research context. This was of particular importance, as I found some major differences between teacher participants' beliefs and behaviors, as well as gaps between the attitudes regarding the testing situation held by instructors and students (see Chapter 5).

There were cases, for example, in which the interview guide I used for a particular participant produced responses that differed from their questionnaire answers, on which the guide itself was based. In others, I recapped observations during interviews as a form of member checking, to make sure that what I was perceiving was in line with what was actually occurring. In general, teachers' interview data matched up with their questionnaire responses and that which I observed in classrooms, however, I did confirm all of the quoted passages and my interpretations of the same with respective participants.

4.7.1 Quantitative Analysis

The responses to the student questionnaires composed the quantitative data in this study. After collecting the responses through the questionnaire software, and removing partial and incomplete responses, I began the quantitative analysis process through exploratory data analysis. This allows the researcher to respond to the data itself, as it is the data that tells the story of what the respondents experience and believe (Cohen, Manion, and Morrison, 2003).

In addition to frequency data and cross-tabulations of the quantitative data, I also calculated Spearman rank order correlation values to investigate patterns of behaviors and perceptions

that may have differed by age, grade, level, and length of term. Magnitudes and directions of correlations are presented and discussed in the following chapters.

4.7.2 Qualitative Analysis

Richards presents different levels representing the collection, analysis, and presentation of data, as "discovery," "analysis," and "interpretation," respectively (2003, p. 263). At the analysis level, he explains that qualitative researchers should keep in mind that data analysis is not a stage of the research, but that it occurs throughout the entirety of the research process. Cohen, Manion, and Morrison (2007) write that it is of upmost importance that researchers first establish a clear purpose for the data before choosing the appropriate analysis to be performed. In the case of this study, I used the Washback Hypotheses to help drive research questions and analysis, but allowed important ideas to emerge from the data throughout the coding and categorization process, much in the tradition of grounded theory. Classic grounded theory, however, requires the researcher to minimize preconceptions and research questions before analysis, and therefore the qualitative analysis I used in this thesis bridged and combined deductive and inductive approaches. Auerbach and Silverstein (2003) refer to the starting point in this analysis as "research concerns," which help direct the researcher with slightly more focus when identifying relevant passages in interview data, and thus the Washback Hypotheses served as my research concerns (or a priori theoretical framework) as I began the coding process.

In this research, I was interested in how my data might tell the stories of the participants as they related to pre-existing theory – in this case, the Washback Hypotheses. I transcribed the interviews, and according to recommendations in the literature, removed interjections and other information not relevant to the participants' message (Dey, 2003; Seidman, 2006), being extremely careful not to misconstrue their expressions, and also noting non-verbal communication found in my notes or shown through laughter, hesitation, etc. In order to facilitate this process, I used Dragon Naturally Speaking transcription software. For some of the participants, I was able to run their interviews through the software and with only minimal

editing for punctuation and to denote participant and researcher. Other participants spoke with accents too far removed from my own, to which the software was calibrated. Rather than try and teach the software each participant's unique accent and voice, I instead respoke those interviews into the dictation software while listening to the original interviews through headphones. This was somewhat inefficient in the end timewise, but allowed me to engage with and think about the interviews before marking them using my coding scheme.

After transcription was complete, I began coding and categorizing the data, revisiting the data on multiple occasions to confirm relevancy and accuracy in my identification of themes and patterns. Richards writes, "In a sense, categorization has already begun when the aims of the research are formulated because implicit in their selection is an element of conceptual identification" (2003, p. 273). Thus, although I was beginning the coding process in a sense, the actual categories and themes I expected to emerge from the data had in many ways been somewhat predetermined by my problematization of the testing situation and context, and my choice of research questions based on the Washback Hypotheses.

Throughout the selective coding "stage" of the research (remembering that in actuality, coding had been going on throughout the study), I employed the method of constant comparison (Boeije, 2002; Glaser, 1965). This allowed me to more accurately categorize my qualitative data within my coding scheme, and helped me to compare interview quotes that I assigned to specific categories both within and across participants. After my initial coding guided by notions presented in the Washback Hypotheses, I revisited the data, and compared participants' comments and observation incidents to the categories that had further emerged from the first round of coding (Cohen, Manion, and Morrison, 2007). Finally, I looked at the efficiency with which my categories included the data, and modified them to better incorporate the ideas that had been expressed during the interviews. As a result of the constant comparison process, I was able to capture and describe the data in themes discussed in later chapters.

4.8 Ethical Considerations

Whether research is being conducted in a manner regarded as qualitative, quantitative, or mixed methods, ethical responsibilities must be considered at all stages of the inquiry (Creswell, 2009, p. 88). Ethical interests arise in such issues as personal disclosure, authenticity and credibility, the role of researchers in cross-cultural contexts, and issues of personal privacy through forms of Internet data collection (Israel & Hay, 2006). Kvale (In this thesis, several steps were taken to adhere to ethical guidelines. First, all participants were given a very clear representation of the research, its aims, and the expectations I had of their participation through individual information sheets (see Appendix E). Students were provided with these information sheets in Korean, explaining their more limited participation through electronic questionnaires and classroom observations, while teachers were given English versions explaining that in addition to questionnaires and observations, they would be asked to occasionally participate in small group and individual interview sessions. Furthermore, a bilingual consent form (see Appendix E) was provided to all participants that outlined in detail what steps were taken to protect their identities and personal information, as well as how the data produced through this research would be used.

Personal data that could be used to identify participants was collected only when deemed absolutely necessary, and in such cases stored securely on my personal, password-protected computer. As voice recordings and observation video data could be used to identify participants, this data was deleted as soon as interviews were transcribed, and observations had been coded and analyzed. The students in this study did not provide any information that could identify them, as I did not request their names, and the large number of participants meant that other personal information (year in school, English course level, etc.) would not have been sufficient to separate them individually.

The teachers in the study, however, were given pseudonyms assigned alphabetically and randomly, as the teachers participated in interviews, in addition to filling out questionnaires. This allowed me to easily identify them to the reader when providing or referencing interview

excerpts, and is recommended even when participants would not be at high risk of being identified by their responses (Seidman, 2006).

Kvale (1996) identifies not only informed consent and confidentiality as ethical issues in interviewing, but also the consequences of the interviews. It was therefore important for me to look for signs of distress in participants, and to always make sure they were cognizant of their right to end interviews and withdraw from the research at any time, even though this information had been conveyed in the consent form and information sheet. Finally, Smith (1992) notes researchers must strive for equity when choosing times, places, and locations of interviews, and thus interviews were arranged according to the wishes of participants.

4.9 Methodological Limitations and Obstacles

This study was not without its shortcomings, and like anything in life, even after many months of planning, there were still certain aspects of the study that did not progress as I had planned, sometimes due to factors beyond my control or ability to foresee. In this section, I address some of these issues, and explain how I was able to shift my approach when necessary in order to best maintain reliability throughout the data collection and analysis phases of the research.

4.9.1 Observation Difficulties

Influential studies in the field of washback research recommend the inclusion of observation data along with questionnaires and/or interviews in order to provide triangulation as well as a clearer picture of what is occurring in the learning environment (Alderson and Wall, 1993; Bailey, 1996). For this reason, I chose to add an observational component to the current study. However, in the weeks leading up to the beginning of the data collection period, a scandal captivated South Korea, as hidden video footage of women in changing rooms, and on subways and escalators was found to have been recorded and uploaded online. This issue was cited by several classes and teachers as reason for them to not agree to my

recording their classrooms for research purposes, and thus my time was stretched quite thin as it concerned achieving a balanced sample of teacher and classroom observations during the data collection periods.

As a workaround, I asked teachers to keep notes on their activities and classroom tasks, and used these notes along with lesson plans and interview questions recapping their teaching activities in order to help fill in the gaps regarding what was taking place in the learning environment. While true observations across all participants would have been ideal, in some cases they were not a feasible option, yet I believe that the choices I made to collect information regarding actual classes from teachers provided a fairly accurate view of the goings on in the classrooms themselves.

4.9.2 Researcher Bias

The analysis of qualitative data begins during the collection of data and continues throughout the study, and therefore it is essential that any biases that exist be addressed in order to best help the reader understand why data is being collected or explained in the ways chosen (Perry, 2005, p.152). Having worked in the university and department in which this study was conducted, I knew several of the teacher participants, and had fairly extensive experience with the test itself. I had thus formed my own opinions about the test's benefits and shortcomings prior to engaging in the research.

Rather than avoiding some of these feelings about the test and context, I embraced them, and they helped me to form my research questions and hypotheses. It should be noted that I held a fairly balanced mix of positive and negative views of the GMATE assessment, and on many occasions argued for policy changes against its use in certain ways, while also strongly advocating for its continued inclusion in the program in others. In order to minimize the effects of researcher bias on the study and collected data, I made sure I remained cognizant of my biases, and made sincere efforts to reflect on times during the project in which these views could influence my interview questions or conclusions.

4.9.3 Interview Bias

In addition to the personal views I held regarding the GMATE assessment and its value to the university and English program, there were also possible issues of different bias occurring during the interviews themselves. Cohen, Manion, and Morrison (2007) write of various sources of bias in interviews, and warn against the negative effects these biases can have on validity and reliability. Misunderstandings, misconceptions, confirmation bias, and even the demographics of interviewers and interviews all influence the data, and therefore, the researcher and reader must be aware of these factors. Furthermore, there have been published numerous articles that contradict one another concerning best interview practices, such as Morrison's case against the use of leading questions in research (1993), and Kvale's (1996) stance that leading questions may assist researchers in attaining real, true answers from hesitant participants.

When all of the possible combinations of interview variables are considered, it can be quite difficult to determine the strength of the data. However, some of these issues can be avoided, if the researcher takes precautions while preparing for and conducting interviews. For example, although the interviews themselves followed a semi-structured approach, in order to allow me to explore different paths based on the interviewees' responses, the questions and prompts that started each string of thought, as well as their wording, remained unchanged from one participant to the next, which allowed me to better compare and contrast responses across participants. While conducting the interviews, I was keenly aware of my own biases in order to best avoid leading interviewees towards answers that might confirm my own expectations regarding the GMATE assessment, and asked leading questions only to confirm or clarify what had already been stated by the participants.

4.9.4 Power Issues

Having worked as a teacher and coordinator in the research context, I had valid concerns about power relationships, past and present, affecting participants' responses and behaviors

during the study. Furthermore, as the findings of the study were shaped by participants' experiences and perceptions, and were also to be used to help improve the program and testing policies if possible, I worried that teacher participants might be less open with some of their more critical opinions. Cohen, Manion, and Morrison write, "the notion of power is significant in the interview situation, for the interview is not simply a data collection situation, but a social and frequently a political situation" (2007, p. 151). Further to this, Kvale (1996) explains that power differences can define the way that the interview is carried out, along with what is said by the respondents, and the situations in which the interview and interview find themselves.

Beyond the power issues that may have affected the teacher participants' behaviors and responses, I was concerned that the student participants would be hesitant to critique their university or professors, in particular if the students were in their first year at the school. As explained in chapter two, Korean students are incredibly deferential towards their instructors, and in my own experience, this respect peaks during their university studies. If students felt in any way that they might be disadvantaged as a result of their questionnaire responses or classroom behaviors, they may have answered or acted in ways that did not represent the truth of the learning situation. In addressing both the teachers' and students' possible hesitation to be open and speak with me during the project, I repeatedly drew their attention towards the ethical guidelines directing my study as the researcher, as well as the steps I was taking to ensure their anonymity. There were instances, for example, in which teacher participants confirmed with me that their answers would not be tied to them in any way before sharing frustrations or negative experiences.

4.9.5 Student Participant Voice

As has been discussed in prior chapters, washback effects are not limited to only teachers in any given learning situation, and thus it is important to collect data that represents learners' voices in addition to those of educators. In the current study, I hoped to achieve balance between student and teacher experiences and perspectives, but it became quickly apparent

that it would be very difficult to recruit a representative sample of students who would be willing to share their time on a consistent basis for the sake of interviews and follow-up clarification. I believe that the questionnaires provide a glimpse into the student experience in the research context, yet because I was unable to delve deeper into their attitudes, behaviors, and perceptions, I chose a somewhat asymmetrical approach to the data analysis and presentation, focusing more on the teacher responses and behaviors throughout the study.

4.10 Summary

Previous washback studies have used a variety of instruments and methods in order to describe and explain washback effects in numerous contexts. However, there are clear trends in the methodological choices that have been made by washback researchers. Some researchers have advocated for a mixed-methods approach, with particular emphasis on including observations (Alderson and Wall, 1993; Bailey, 1996; Bailey, 2000; Watanabe, 2004). However, as is shown in Table 3.1, washback studies have been conducted using diverse methods, and with a variety of participants and stakeholders, depending on the research questions posed. A common methodological thread tying together the majority of these investigations, though, has been the use of a mixed methods approach, the inclusion of interviews with stakeholders, and when feasible, classroom observations.

This thesis investigated a well-defined and bounded system in the context of a single Korean university, showed temporal characteristics as the study progressed throughout two terms of varying lengths, focused on individual groups of actors and their perceptions of events when looking at both students and teachers, and was particularly concerned with rich, chronological description of the events taking place within the system. All of these elements provide strong justification for the use of case study research (Hitchcock and Hughes, 1995), and this project proceeded as such.

In order to answer my research questions, it was critical that I chose the appropriate research methods for data collection. It would have been ineffective, for example, to use only

questionnaires to fully understand teachers' feeling and perceptions related to the testing environment within which they work. The table shown on the next page (Table 4.3) provides a brief summary of each research question, and how it was answered through the use of specific research methods.

Table 4.3 Research Methods Summary

Research Question	Research Methods	Justification	Data Collection Period
To what extent do teachers feel differently about the GMATE, and how does their teaching and test preparation vary depending on the English proficiencies of their students? To what extent do teachers feel differently about the GMATE, and how does their teaching and test preparation vary when their classes include students for whom the GMATE is a high-stakes exam?	Interviews	Interviews provided richer, more accurate descriptions of the events within the context, while allowing me to focus on specific actors in the bounded system of the university. These interviews were guided based in part on information received from interviewees' questionnaire and prior interview responses at different times throughout the data collection period.	Phase I Weeks 5, 11, 14, and 15 Phase II Days 5, 11, 14, and 15
What are the observed washback effects of the GMATE on the classroom environment?	Classroom Observations	Observations were a crucial part of developing a clear picture of not only *why* events occurred (as provided by interview and questionnaire responses), but exactly *what* was occurring. As teacher accounts of classroom activities could be subjective, these observations helped to provide stronger evidence to corroborate instructors' interview responses.	Phase I Weeks 5, 7, 9, 11, 13, 14, and 15 Phase II Days 5, 7, 9, 11, 13, 14, and 15

To what extent do students of different English proficiencies feel and act differently towards English courses and the GMATE? To what extent do students for whom the GMATE is a high-stakes exam feel differently towards English courses and the GMATE than those students for whom the GMATE is low-stakes? To what extent does the length of term influence students' approach to English language learning and test preparation?	Questionnaires	In order to collect responses from a large number of students of varying English proficiencies and years in school, I used questionnaires written in students' native language of Korean. This yielded a large amount of data that allowed me to gain a better insight as to what the student experiences, perceptions, and behaviors were. Teacher questionnaires were administered at the beginning of both research phases, but results were used only to inform and direct interview questions during the first meeting with each teacher participant.	Phase I Weeks 3, 9, and 15 Phase II Days 3, 9, and 15

Chapter Five: Findings

The findings of the study are presented in this chapter. The data is presented in this section to help the reader clearly see how research questions have been answered, and it is also organized chronologically to help tell the story of the study as it unfolded. Research questions are addressed, with findings from each phase presented separately and then compared, when appropriate and relevant to research sub-questions. As the first research questions focus on washback related to teaching, data from teacher interviews and observations are presented first, while the second major research question looks at student washback, and thus student questionnaire responses and trends are found thereafter. This chapter focuses only on the data and analysis of such, while a more in-depth discussion of these findings accompanied by appropriate comparisons with the literature is made in chapter six.

5.1 GMATE Washback on Teachers

This section deals specifically with research questions related to washback effects on teaching, and is organized into different subsections based on washback effects related to the level of the students, the stakes of the exam, and washback effects that became apparent during classroom observations. The findings related to the first two research questions in this section are answered through the analysis of teacher participants' interview responses, while the third involves the summarization and presentation of classroom observations for four of the teacher participants.

5.1.1 Washback on Teaching Based on Level

This subsection deals with research question 1.1: To what extent do teachers feel differently about the GMATE, and how does their teaching and test preparation vary depending on the English proficiencies of their students? As mentioned prior, each teacher participant's interview varied slightly based on his or her responses to the questionnaire. Subsequent

interview questions were developed based on observations, when appropriate, and participants' ideas and experiences expressed in prior interviews, and thus interview guides differed even more as the term progressed. Teachers' responses are organized in this section into themes based on classroom tasks and activities, content and teaching methods, materials, and attitudes and perceptions regarding the GMATE.

Tasks and Activity Variation

12 of 17 teachers interviewed in this study responded that they depend on the textbook for most of their in-class activities, and that the majority of their classroom tasks and activities vary little from year to year.

> Gemma: The books change every year, but I'm comfortable sticking to what helps my students, and what I've prepared and done before. If something is really different, I might make something, but the book is pretty good.

Another teacher talked about recycling or reusing materials and activities.

> Alex: I think I've been doing a lot of the same things for years. The students seem happy, and it works. I make little changes. Not a lot, but if something works in a class on Monday, I might do that again in a Tuesday class.

Several of the teachers who had been teaching at this university for more than a few years repeated these ideas of recycling activities from class to class and year to year, even as the textbook and curriculum changed.

> Ryan: We can just click our old syllabus for the new term, so I guess you can make changes…I do things different each year, but I don't update my syllabus that often. Most things are pretty much the same.

When asked about differences in their activities across class levels, though, all teacher participants said that they use very different activities for different level classes. Carla said this was somewhat natural, as the textbooks varied, and teachers were "strongly asked to use the books when possible. Otherwise students complain." Another teacher was more specific in his approach.

> Frank: I really only makes major changes for level 1 students. Level 2 and 3 students can pretty much do the same things, but level 3 students can do them alone for longer. Level 1 students need more direction and help. I can just leave level 3 alone and they'll do the activity for fifteen or twenty minutes.

Several teachers commented on the autonomous learning and ability to stay on track for upper-level students, while mentioning that level 1 students required "far more (teacher-student) interaction" to create an effective learning environment. This was certainly the case when I examined teacher-student interaction trends during the observations, as well.

I inquired specifically about tasks and activities related to the GMATE, and there was surprisingly less variation across teachers here. 15 of 17 teachers interviewed said that they rely on the same GMATE preparation tasks and activities in their classes, regardless of students' proficiency level. One teacher focused on the use of mock tests.

> Monica: Students need to get to know the test, the timing, the wording. Doing practice tests is the best way for that, and students have a lot of practice items that I give them, so they can practice with partners in class without my help.

This practice was supported by another teacher, too, mentioning shared needs across levels.

> Alex: It doesn't matter if they're level 1 or 3. They want to do well on the GMATE. It's part of their grade. A big part. Practice tests are the easiest way for them to get familiar with the test. They are used to practice tests from high school, so freshmen pick that up really quick.

Of the two teachers saying that they preferred to do different activities for different levels, one said that she did far more indirect activities in class.

> Donna: I don't want to waste class time on practice tests. I give them practice questions, and they do those on their own, or for part of their homework. I've made some games and things that get students practicing the skills, like describing pictures and processes, but I think they like those more than just practicing the GMATE every class.

Finally, several teachers talked about not being able to do activities they thought would benefit the students' learning because of time limitations related to addressing GMATE preparation.

> Jamie: Near the end of the term, I have to cut things I planned a week or so before, because students need more time to practice for the GMATE. Sometimes they are actually ready but just want more practice time, so I give it to them. You can't do everything you plan to do.

A common thread between teachers regarding the GMATE, reflected in both the questionnaire responses and interview questions, was that they felt limited in the range of tasks they would like to have done with their students, but were not able to do so due to the GMATE requirements placed upon them and their students.

Content and Methods Choice

Focusing a bit more on the content and methods employed by teachers allows for another look at washback effects stemming from the GMATE. This theme from the interviews was

related to *what* and *how* the teachers the taught. Level 1 teachers echoed notions that they were a much more integral part of the classroom environment, while level 3 teachers responded that they felt more comfortable giving their students higher degrees of freedom and responsibility for content during class times. Level 2 teachers, though, were fairly split in their approaches.

> Les: I get a good sense of what level 2 students can really do. Some classes have such a wide range of abilities that I have to be careful about who is paired up. If two low-level 2's are in the same group, they'll just sit there…other classes, I just give an activity and walk around the room keeping them doing the discussion.

There seemed to be much more variation in how teachers grouped students in different level classes, as well. Patrick, for example, said that he preferred larger groups regardless of ability, because it translated to "more ideas, so less awkward quiet," while Donna chose to divide her students into smaller groups in lower levels, because "they need more talking time." Teachers generally said that they focused on discussion activities during the bulk of their classes, with these discussions occurring in groupings of two to six students. With class sizes averaging eighteen students per class, only a handful of the teacher participants stated that they organized whole-class discussions.

> Patrick: Whole-class discussions aren't really needed. If students get mixed up from week to week, they'll hear a lot of different ideas and meet new students. Doing a discussion as a whole class means what, that every student can speak for one or two minutes?

It is interesting that here his ideas regarding grouping students in larger groups to increase the number of ideas being spoken is replaced by Donna's preference for increased talking time.

Virtually all of the teachers talked about a division between content related to discussion and presentation, and course content related to the GMATE. However, when asked how they saw their course content connecting with the requirements of the GMATE, 12 of 17 said that the textbook made that connection for them, as it was "designed with the GMATE in mind." Only one of the teachers chose to focus on grammar and vocabulary in his class, despite those elements not being a major part of the textbooks. Neal said that he asks students to do grammar worksheets and memorize vocabulary words, because "if someone can't use good grammar, they can't communicate, and no one will understand them." However, he stated that this was less due to the GMATE, and more part of his own teaching preferences.

This division in content focus was more pronounced near the end of the term, when many of the teacher participants talked about dedicating certain days or weeks solely to GMATE preparation. Using a "GMATE review day" just before the exam was a very common choice among participants, with 16 of 17 saying that they included it in their syllabi, regardless of level. The lone dissident of this group, Edward, stated that he chose instead to do GMATE practice for the first ten minutes of each class throughout the term, and then just before the exam week gave his students a review sheet with practice items and advice to achieve a higher score.

> Edward: I know GMATE review day is pretty common, but a whole GMATE only takes like five minutes. Students don't usually do more than a couple [practice tests] before getting on their phones, in my experience. It's better for them to do a little each day, especially for level 1 students.

Materials

While almost all of the teachers said they created their own supplements to the textbook, some relied upon these materials more than others. One teacher was quite frank.

> Brynn: I don't like the textbook one bit…I would prefer to just do what I want with my students. That's what being a teacher is, after all. Being able to create your own content for your students' needs.

However, she mentioned the same experience as others in the study, in that she made sure to refer to a variety of pages in the textbook, as the students are "expected to buy it, and can get pretty loud (complaining) if we don't at least make an effort to use it." I thought it was very interesting that there existed so many opinions regarding the textbook, as it is created and edited by the teachers themselves, and thus they have almost complete control over what is included. One teacher was more supportive of the provided materials.

> Edward: I like it (the textbook). I've worked on it, so I'm probably biased, but it's not too bad. We're not book makers. We're teachers. So it's pretty good, considering.

Supplements to the textbook were incredibly varied between teachers. Upon request, Ingrid brought some of her lesson plans and supplements to one interview, and proudly displayed what she called "years of experience refining these to where I finally think they're really good." She said that she took a very deliberate approach to organizing her classes and creating content, and depended on notes she writes on her supplements during classes in order to improve them for the future. Ryan, on the other hand, said that his only "supplements" consisted of pages from prior textbooks that had been omitted from the current edition.

> Ryan: I don't do the textbook committee, so I don't have a lot of say about what goes and stays. There were some things I really liked and always do, so I just photocopy those and give them to my students.

There was an incredible range of supplements that teachers created, photocopied, and downloaded, but attitudes about sharing their supplements tended towards feelings of ownership and distrust of colleagues' intentions.

When asked whether or not she shared her supplements and teaching ideas, one participant was more protective of her work.

> Brynn: I worked really hard to create good activities for my students. I don't want some teachers just coming along and using them without giving me credit for that (hard work).

Asked about perhaps trading supplements or activities, she was somewhat skeptical.

> Brynn: I suppose that might work, but really, every teacher is so different. What works for one person won't necessarily be right for another person. I think it's better for us to stick to our own activities. We know our students best.

This was a surprisingly common idea mentioned among the teacher participants. Two teachers spoke of this idea specifically.

> Neal: They can use them (his supplements) if they want, but I don't think anyone does. No one asks, anyway.

> Frank: Everyone is so focused on their own classes. We don't do a lot of collaboration, except maybe during PDC (professional development) meetings. Everyone is pretty busy during the term. And different schedules.

Attitudes Regarding the GMATE

Many teacher participants seemed quite eager to talk about the GMATE itself. Some teachers thought the test itself was "good," but was not good for students or teachers.

> Kyle: The GMATE is good, in theory. I know it's a good test. I mean, it's not like the TOEIC or TOEFL, where students just memorize everything and can do well. You have to be able to speak English to get a high score (on the GMATE). But it's definitely distracting for students and teachers.

Within this quote we see a very good example of teachers' perceptions of the GMATE as having strong construct validity. Teachers seemed to be very aware of how the GMATE was originally designed, and for what purpose, despite expressing feelings that the exam caused negative washback. In a group interview, several teachers agreed that the GMATE forces teachers to "teach to the test." When asked if "teaching to the test is always bad," those same teachers backpedaled a bit.

> Donna: It doesn't *have* to be bad, but I think here in Korea tests are always thought of as the bad guy. Students are so stressed about tests all the time. Having another test that they have to pass just makes learning English the same as it was for them when they went to *hagwons* (private English academies) every day.

"Teaching to the test," which in reality can be described as washback effects, was quoted by more than a few participants. There seemed to be a very common shared idea that washback effects exist primarily in the negative form, and that testing, especially standardized testing, is inherently bad. There were more than two dozen instances during the data collection period of teacher participants using such phrasing, such as Carla, who said that she "wants to avoid teaching to the test, but it's pretty much unavoidable, the way the book is set up." Edward stated that he felt "like there's no option. The test decides the syllabus for most of us." Ingrid mentioned that "for most students, there's this feeling that the test is more important than other stuff we do. So I spend a lot of time talking about it and practicing for it. It shows in my evals." Finally, Alex mentioned teachers' cognizance of the importance of the GMATE.

> Alex: We all do it (teach to the test), even if we don't talk about it. Actually, we *do* talk about it. I remember people complained about it when we asked to change the pass or fail part of the test, but admin said 'no'.

On a related note, there was a recurrent theme of teachers stating that they felt that the GMATE and related policies undermined their ability to teach effectively, and eroded their sense of professionalism. As Brynn was quoted in the previous section regarding materials and washback, she thinks that requiring teachers to use the textbook, which is in large part based around the GMATE skills, doesn't allow for teachers to rely upon their professional abilities to address students' individual needs. Other teachers were not necessarily softer in their criticism of the test as it relates to their professional identity.

> Frank: I think we need to think seriously about what kind of talent we have in our department. We all have master's degrees. I mean…c'mon. We don't hire people without master's degrees and experience. We shouldn't need the handholding.

A couple of the participants were somewhat defiant.

> Neal: I've got a pretty good idea of what I'm doing. I don't need to be told what and when and how to teach, like some *hagwon* teacher.

> Jamie: It's ridiculous that we should really need that. In my last university, there was much more freedom (to teach what we wanted). I think most people want that. If not, they're not real teachers.

5.1.2 Washback on Teaching Based on Exam Stakes

Here I present findings directly related to research question 1.2: To what extent do teachers feel differently about the GMATE, and how does their teaching and test preparation vary

when their classes include students for whom the GMATE is a high-stakes exam? The majority of students taking general English courses during the regular term are freshmen and sophomores, for whom the GMATE is not a pass/fail exam. However, any student who started at the university prior to 2015 is required to attain a certain score on the GMATE, as determined by her major, or she automatically fails not only the exam, but the course, as well. These older students make up the majority of classes during the intensive period of study. While it is possible for freshmen and sophomore students to fail the exam, if they do not take the exam at all, for example, this does not necessarily result in a failing grade in the course, and thus the stakes for different students in a single course can vary wildly. This section deals with washback effects on teaching related to participants' interview responses that dealt with the stakes of the exam, and is divided into similar sections as Section 5.1.1.

Tasks and Activity Variation

Teachers were very open about their willingness to tailor their classroom tasks to their students. The teachers in the intensive term, for example, candidly said that they prioritized the GMATE above all else, and communicated that to their students on the first day of class.

> Patrick: I like the intensives. They're pretty laid back, and I don't have to worry about a lot of the things I usually do during the fall (term). There's not time, but also the students just really want to pass the GMATE…it's totally different. Totally. You know, and the students know, if they don't pass the test, the class is a waste, and they have to re-take it. It's my job to prepare them, so we do.

Another teacher also said she spends much more time on direct GMATE practice for older students, but that it varies by level.

> Ingrid: We can take a look at students' student numbers and know how many pass-or-fail students we have. In level three, I don't really worry

> about them, except for maybe the English majors. In level one, I spend a huge chunk of time on practice every day. We only have three weeks.

Participants during the regular term, though, generally assigned tasks and activities to their classes based on the assumption that their students were subjected to the lower stakes version of the GMATE.

> Neal: You know which students are seniors. You know they have to get a certain score. But I can't spend a lot of time in class only on the GMATE for them, because there are like fifteen other students that want to do other things.

> Les: In in the past, it was easy. Everyone had to pass. Now you might have one or two seniors or juniors, and you feel a little bad, because you know you can't just focus on the GMATE like you did in the past. But I think we still do enough practice. I don't think many students fail.

This is interesting, because the student questionnaire responses showed that students generally viewed the GMATE as fairly high stakes regardless of the fact that only certain students were required to attain a certain level on the exam, while teachers viewed the exam stakes slightly more objectively, accordingly placed more or less importance on preparation activities and tasks. Lastly, the intensive teachers talked about assigning and grading fewer assignments during the intensive term. This seemed to be more related to the length of the term, rather than the stakes of the test, but it was interesting nonetheless.

Content and Methods Choice

Although intensive teachers have the same number of contact hours with their students as they do during the regular term (45 hours in total), efficient use of time, and choice of content focus was mentioned in some fashion or another by almost all of those teacher participants.

> Les: You gotta decide what you're gonna focus on. You can't do the same things you do during the term, because you just can't use the time the same way. It's just not realistic to expect students to focus for three hours every day of the week during their vacation. We (teachers) can't either.

Another participant said that it was this perceived limitation on time that helped him make the choice to avoid volunteering to teach intensives.

> Oliver: It's tough enough to get through everything I want to during the term. The idea of trying to do all that in a shorter amount of time is just not appealing at all.

It wasn't only Oliver that expressed discomfort with the prospect of cramming in fifteen weeks of content into a period of three weeks.

> Gemma: I used to do them (intensive classes). I tried using my same syllabus, because we're told they're supposed to be the same as the regular term. There's no way you can do that, though. Students are burned out, and each day is like a week's worth of materials. It takes a lot of cutting and reworking. It's like teaching a totally different class, really.

There was a marked difference in the teaching methods employed by intensive teachers, too. Five of the intensive teachers stated in some form that they spent more time lecturing to their intensive students, and that this was mostly related to GMATE content or testing tips.

> Carla: I don't want to take the chance that my students forget something important related to the GMATE, and then have to take the course again. I'm sure they get tired of it, but I'd rather hammer them with review than see that a few students failed. It means more talk time for me, but it's a tradeoff.

> Patrick: I don't normally like the whole lecture thing, but you know, these students are used to that for their whole lives. They're really good at taking in information from their teachers' lectures. If that's what it takes, sometimes I think that it's fine. Even better, maybe.

There is generally somewhat of a demonization of teacher-centered instruction in language learning environments here in Korea, particularly when native English speakers discuss Korean public school English teachers, so it was interesting to hear that some of these teachers espoused that teaching style when approaching their intensive courses.

However, there was a major difference between the intensive teachers when looking at their levels. While all of these participants said that they cut the course content from the regular term courses in some ways, level 1 intensive teachers explained that they cut far more.

> Alex: It's basically all about the GMATE…if I'm walking into a level one classroom during the summer or winter term, I know that those students have one thing on their minds. That's the GMATE. I include some parts of the book, like the process and description chapters. But that stuff is on the GMATE, so it helps them prepare. Everything I do is basically to help them get ready in the short time we have.

This view was in stark contrast to the opinion of Ingrid, who taught a level 3 intensive course.

> Ingrid: Level three students aren't generally retaking the course, unless it's to get a better grade. It's more stressful for me to teach them during the winter, I think, because it's not so straightforward. I have to choose which chapters I can teach in the short time, whereas level one teachers can just focus on the GMATE more.

Materials

While activities and content varied with stakes of the exam, there were surprisingly fewer differences in the materials that teachers used, and these slight variations generally followed participants' responses in previous themes.

> Ingrid: I don't really do a lot of homework during the intensives. There's just not time for it, and we're getting daily contact and practice time, so there's really no need to give them (the students) speaking journals, or things like that. They'll do them, but I don't like giving homework just to give homework…I'll give them the (speaking journal) sheet. I just don't require it.

Another intensive teacher, Donna, said that she provides her supplements to her intensive students electronically.

> Donna: I give them all the regular handouts and assignments, just in case they want them. I just put them up and tell them that they're there. Students who want them are happy, and I'm happy it doesn't waste paper on things I might not have time for.

There seemed to be a strong connection between materials and content choice during the regular or intensive terms, yet this was not necessarily based on the stakes of the GMATE for the students, but more likely because these teachers felt they were not able to cover the same amount of content.

Attitudes Regarding the GMATE

I expected teachers to have fairly consistent feelings about the GMATE in general, regardless of the stakes for their students, but some teachers were very forward about how they felt about the GMATE as a higher stakes exam. Brynn, who didn't actually teach an intensive

course during the research period, but said she had on numerous occasions in the past, was not supportive of the current policy.

> Brynn: I think it's absolute rubbish, really. That we have students in one classroom held to two standards, and one can pass the course with a moderate low while another fails. I'd like to see them (administrators) treat all students equally and fairly, but good luck with that.

Another teacher who didn't take part in the intensive term agreed with Brynn's comments.

> Hector: You feel awful when you see that an older student in your class failed the GMATE, and you know she's going to be taking it again. You know how hard she worked all term, and then this? One test that lasts what, four minutes? Five? I've had A-plus students fail the GMATE, and the computer won't let me give them anything but an F.

Somewhat surprisingly, there were some sentiments explained by intensive teachers that conflicted with the aforementioned comments. I had expected intensive teachers to express themselves more in opposition to the GMATE, as their students would be taking the exam as a higher-stakes test, but this wasn't always the case.

> Patrick: I don't know that it's completely *fair*, but I wouldn't say that it's completely *unfair*, either. It's not like these students don't know the requirements. They're pretty straightforward about what the students need to achieve. It's not like they're required to do the TOEIC or something.

It was interesting that Patrick seemed to qualify the GMATE as a more just exam than the TOEIC.

> Patrick: At least the GMATE is something they can use. They have to write or speak in English. I don't know how many people I read about that get perfect TOEIC or TOEFL scores and can barely speak a lick of English.

This was one of a handful of comments that I heard in defense of the GMATE during the research period. These notions were almost always qualified by comparisons to more well-known standardized English language exams, extolling the benefits of an exam score that might more accurately correspond with a test taker's English proficiency.

5.1.3 Washback Effects Observed in Classes

Finally, in this subsection, I organize and present relevant findings from the classroom observations in order to answer research question 1.3: What are the observed washback effects of the GMATE on the classroom environment? Observations provide an important window into the classroom, and allow researchers to connect participants' accounts of class activities with what actually happened, as these can sometimes differ. I was able to observe four teachers during Phase I of the research period, and the findings of those observations are presented in this section, divided by teacher in order to better show similarities and differences between the participants.

Carla

Carla taught level 2 students during the regular term, and conducted her classes in a very organized, predictable manner. When asked about this, she said that it helped students to stay on task, knowing that they would do certain tasks for specific amounts of time each class period. After organizing students into groups of four or five students by students' choice, she spent the beginning 10 to 12 minutes on GMATE practice tests, providing students with practice questions shown at the front of the classroom using the overhead projector.

Students were left to work on these alone, and Carla generally stayed at the lectern located at the front of the classroom, with a few exceptions during which she walked around the classroom, but without interacting with students. This was interesting, because she said during interviews that she tended to interact with students and talk much more during the intensive terms.

The middle and longest portion of all but one of Carla's classes was dedicated to discussion activities. She followed the textbook in most cases, starting the discussion with a minimum of 5 to a maximum of 12 minutes of what was listed as a "warm-up," consisting of five to seven fairly straightforward questions related to the day's discussion topic. Students navigated these questions on their own in their groups, with minimal interaction with their teacher, Carla. She worked more as a facilitator during her classes, calling out pages and directing students' attention to certain areas of those pages, while offering simple explanation as to her expectations.

After the warm-up section of the discussions, she gave her students the option of choosing activities from several provided in the textbook, with the exception of asking her students to skip sections on grammar. She justified this decision in a later interview.

> Carla: They can do these on their own, but they're really not worth doing in class. I'd rather use that time for speaking as much as possible.

Students spent between approximately 20 to 30 minutes on this section, and Carla added that she let students engage in longer discussions if they were fairly active as a class, and would stop them earlier if the discussion slowed. In most classes, Carla chose a video clip listed in the textbook, and showed it using the projector. She then engaged the class in the whole-class discussion for the remaining time in class, posing questions to students she called out by name.

In one class, Carla asked students to work on presentations explaining how to do something, called “process presentations.” This was directly related to item 3 on the GMATE, and was included in the textbook, along with description presentations (item 2), and opinion or persuasive presentations (item 4). She talked in one interview about balancing her teaching goals with students’ learning preferences.

> Carla: The students hate presentations. They *hate* them. But they’re part of the course, and good practice for them. I would prefer more discussion, and I think they would, too, but the class is called ‘discussion and presentation’, so we do a presentation every few weeks or so.

Besides the time spent practicing for the GMATE, and the time practicing for the process presentations, I didn’t find many connections between Carla’s class activities and the test itself. One could make the argument that students were sharing their opinions during discussions, and thus this was indirect preparation for item 4 on the GMATE, but this item is a bit more focused on advantages and disadvantages of something, which wasn’t necessarily a part of discussion activities.

Donna

As a level 3 teacher, Donna conducted her classes in a very different way from Carla, while still relying heavily on the textbook. She did not spend any class time during the observed classes on explicit GMATE practice.

> Donna: Level three students really don’t need it. If they keep up with the speaking journals, they’ll be fine. Really, they should already be at MH (moderate high, the highest score attainable on the GMATE) level based on their GELT (placement test score).

Unlike Carla, Donna divided her students into groups of her choosing, rather than allowing the students to separate themselves into groups on their own. She said that this saved a few

minutes, as “students can be really shy about getting into groups, even when they can speak like level threes.” Like level 2 students in Carla’s class, Donna asked students to do what she called “warm-up questions,” although these were not marked as such in the level 3 textbook, as they were in the level 2 text. Students discussed these questions for up to 15 minutes, after which time Donna asked students to choose one of four sections in the day’s chapter to discuss. The level 3 textbook was very different from the other levels, offering little direction or variation in activities, and instead providing students with four specific sections of questions relating in different ways to the overall chapter topic. Students engaged in these longer discussions for close to 30 minutes in each class, after which time Donna asked students to prepare a short presentation based on directions in the textbook, and present those ideas to their group members.

During both the discussions and short presentations, Donna floated around the room and participated in discussions with students, adding her own ideas and posing questions to student group members, while offering encouragement and praise. A very interesting difference in these classes was that there was almost no direct preparation for the GMATE whatsoever, but at the end of the term, students responded on the final questionnaire that they felt Donna had adequately prepared them for the exam. I asked Donna about this point, and she said they did do a single session related to GMATE questions near the exam day, and she assigned the very common speaking journals as homework, but felt her students didn’t need much more preparation beyond that.

Kyle

Kyle taught level 1 students, and somewhat predictably spent more time in class talking about or teaching the GMATE, as several teachers during their interviews reported that they spend more time talking when they teach level 1 courses. Kyle spent a full 15 minutes or more at the beginning of each of the observed classes on explicit GMATE preparation in one form or another. On one occasion, he talked about specific strategies for the picture description task (item 2), and then allowed students to practice with pictures from their smartphone galleries

or the internet. In another class, he asked students to focus on item 3 (process), and provided mock test items for the students to use in pairs. Towards the end of the term, he used both the beginning and ending periods of classes for GMATE practice.

Between these GMATE practice periods, Kyle followed the textbook, but spent more time explaining activities, and allowed students much less time to work on activities between explanations. While the other three teachers who I observed spent only a few minutes talking to the class between tasks, Kyle spent almost 1 minute talking to students for every 2 minutes the students were engaged in activities or discussions, on average. The activities in the textbook were much more similar to those in level 2's book, but with more focus on drills and practicing specific language patterns such as agreeing and disagreeing, describing using prepositions of place, and the use of modals.

As I observed Kyle's classes, there seemed to be far more connections between the activities and the GMATE than in other classes, even after excluding explicit practice time. In each chapter there was a focus on building vocabulary used to describe things, such as adjectives for food, movies, or sports. These words could be used by students when attempting any of the items on the GMATE, but would be especially useful on the picture description task (item 2). Furthermore, many chapters' discussion question sections contained examples that directly related to GMATE item 4, dealing with advantages and disadvantages, asking students to provide pros and cons of travel, movies, exercise, plastic surgery, or learning an instrument, for example.

Overall, the activities in Kyle's classes were far more related to the GMATE both directly and indirectly, than in any of the other teachers' classes I observed. Additionally, Kyle spent more time explaining tasks and activities to his students than did the other teachers, which was somewhat understandable considering the students' lack of English proficiency, and Kyle's perceived need to "keep them on track." The textbook seemed to draw upon the GMATE items for activities and tasks, as well, but it is interesting nonetheless that Kyle in particular

found the GMATE to be a better test in some regards, but still “distracting” for teachers in his university department.

Patrick

Patrick was another level 2 teacher, but unlike his colleague Carla, chose a different approach when preparing students for the GMATE throughout the term. He organized his students into groups in the same way as Carla, grouping his students into slightly larger groups of fives and sixes, but spent no time at the beginning of classes on GMATE item practice. He gave a similar reasoning for this as Donna did.

> Patrick: The homework assignments I give them are enough, I think, when you consider we also do GMATE review days during the week before finals.

This was of note because while Donna taught level 3, Patrick was teaching level 2 students, and thus I would have expected him to spend a bit more time directly addressing the exam during class time.

With this exception, he conducted his classes in a very similar fashion to Carla, following the textbook for the bulk of his activities and tasks, although in two classes he provided supplementary activities to the students, which he said he copied from previous versions of the textbook, as they had been edited out of the current edition.

> Patrick: I think everyone has their favorite activities, and they might be there the next year, or they might disappear. Fortunately we can just copy something, as we get all the levels’ books each year.

Unlike any of the other teachers, though, Patrick seemed to make a fairly strong effort to follow the textbook precisely, going through each activity in a unit as time allowed, and often

simply skipping activities near the end of chapters as classes came to a close. He justified this.

> Patrick: I figure the textbook committee chose each activity in a certain order for a reason. They kinda build on each other…it's nice to break up a discussion with some reading or something like that sometimes.

Despite going through the chapters more methodically than did Carla, who chose certain activities while excluding others, I didn't find the activities in Patrick's classes to be extremely well-connected with the GMATE items. Again, there were indirect connections that could be made, as students would describe things or offer their opinions on certain topics, as did Carla's students, but there wasn't the same degree of test preparation that was observed in Kyle's level 1 classes.

5.2 GMATE Washback on Students

Despite the very short time (just under 4 minutes, as reported by the questionnaire software) that the questionnaire took to complete, there were some students who dropped the questionnaire before finishing it. This incomplete data was excluded from the study. After randomly removing 5 responses from the first questionnaire and 3 from the second to achieve even numbers across all three, there were 347 student participants completing all three questionnaires in Phase I, and 122 in Phase II. Furthermore, there were some questions that were not particularly relevant in the second week of the term (questionnaire items regarding whether or not teachers had done an effective job preparing students for the exam, for example). Thus, student responses to the third and final questionnaire are described in the following section, followed by specific comparisons of response trends by proficiency level (research question 2.1: To what extent do students of different English proficiencies feel and act differently towards English courses and the GMATE?), year in school (year in school determines the stakes of the exam for these students, and is related to research question 2.2: To what extent do students for whom the GMATE is a high-stakes exam feel differently

towards English courses and the GMATE than those students for whom the GMATE is low-stakes?), and seasonality (research question 2.3: To what extent does the length of term influence students' approach to English language learning and test preparation?).

5.2.1 General Perceptions of English Exams

Given that the national university entrance exam is notoriously stressful for the majority of Korean high school students, and that most of the participants in Phase I would have been only a year or two removed from this testing experience, their responses to items related to English testing in general were quite surprising. When presented with the statement, "I have an overall positive experience taking English tests," 287 (82.7%) answered that they agreed. This response trend may have been influenced by their experiences taking English exams during their first term at university, but it was interesting nonetheless.

Although students expressed generally positive views regarding their experiences taking English exams, their faith in those exams' ability to accurately portray students' abilities waned. Participants responded to the statement, "English test scores are accurate representations of my ability," and while a large number, 199 (57.3%) agreed, 130 (37.5%) of those respondents agreed only somewhat, while 75 (21.6%) disagreed somewhat, and 59 (17%) disagreed. Here we see very mixed views in Phase I participants regarding the degree to which they trust in the construct validity of English exams, regardless of the quality of their experiences.

Continuing along the lines of inquiry into students' past experiences with and behaviors relating to English tests in general, participants in this phase were given the statement, "I prepare for English tests using materials separate from my classwork." The vast majority of students in this phase disagree with this statement (216; 62.2%), while 146 of those respondents either disagreed (103) or strongly disagreed (43). Of the remaining 131 (37.7%) participants who agreed, 89 only somewhat agreed. These responses reinforce previous notions that students strongly rely on their teachers and classwork to prepare for exams.

5.2.2 Student Knowledge and Perceptions of the GMATE

Students showed a fairly keen awareness of what the GMATE entailed, how it was organized, and they expressed fairly clear judgments about the test itself, shown by their responses to items summarized in Table 5.1 below.

Table 5.1 Student Responses to GMATE Knowledge/Perception Items

Item	Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Strongly Agree
I am aware of the format and items that appear on the GMATE.	4	13	20	109	146	55
I understand why I must take the GMATE.	45	63	96	88	48	7
The GMATE is more important to me than other parts of my English class.	22	65	85	106	61	8
It is important for me to be able to speak English.	4	1	2	32	135	173
The GMATE will help me with my future career.	82	77	72	85	29	2
The GMATE is a fair test of my English ability.	101	116	81	41	7	1
All students at my university should be required to pass the GMATE.	44	69	70	71	58	35

As is shown in the preceding table, 310 (89.33%) of the participants in this phase agreed that they were "aware of the items and the format of the GMATE," which came as little surprise, as most teachers reported spending increasing amounts of time reviewing and preparing for the exam throughout the semester, while others addressed the GMATE on a consistent basis from the beginning of the term. Because of this fact, it was interesting that such a relatively large number of students felt that they were unfamiliar with the types of test items and the way those items are organized.

Beyond the *what* of the GMATE, I was interested in whether students understood the *why* of the test. As tests are commonplace in Korean classrooms, I wondered whether or not students were being informed as to the reasons beyond the GMATE and its related policies,

or if learners were simply accepting this assessment as yet another moment in life of a Korean test-taker. When students were given the statements, “I understand why I must take the GMATE,” participant responses formed a fairly normal distribution around slightly disagreeing and slightly agreeing, indicating uncertainty in their understanding of why this language exam was being administered. In detail, 96 (27.7%) somewhat disagreed with the statement, 63 (18.1%) disagreed, and 45 (12.9%) strongly disagreed, while just 88 (25.4%) somewhat agreed, and 48 (13.8%) agreed. This is evidence that instructors may be preparing their students for the exam itself, but are not necessarily helping them to become cognizant of why the test is in place.

Somewhat surprisingly, students in Phase I didn’t value the GMATE as highly as I would have expected, having likely grown up giving their test scores high regard, particularly in the case of English language exams. Furthermore, the GMATE is a higher-stakes exam for at least 23% of these participants, as their ability to pass the course and graduate hinges on their test scores. Additionally, despite slightly lower stakes, twenty-percent of freshmen students’ course grades depend on their GMATE scores, which led me to expect more importance placed on this exam. However, student participants who agreed and disagreed were split down the middle when presented with the statement, “The GMATE is more important to me than other parts of my English class.” While almost exactly half of students both agreed and disagreed with this, a larger proportion of the agreeing students (106 out of 175) only somewhat agreed, while 85 of 172 somewhat disagreed, 65 of 172 disagreed, and 22 strongly disagreed. Here we see that these students have very differing valuations of the GMATE, perhaps because of the lower stakes (a possible correlation covered later in this section).

Even though students placed little importance on the GMATE itself, attitudes towards the communicative English skills are generally very positive across South Korea, with the great majority of Korean citizens and companies believing that English ability is fairly indispensable for one’s success in school or the workplace (indeed, as discussed in earlier chapters, their study and spending habits support this). Therefore it was of little surprise to me that a full 98%

of all participants agreed to some degree that “it is important for (them) to be able to speak English,” with 49.8% of all students strongly agreeing. This is important to note, as it is clear here that students’ feelings towards the importance of English language abilities are not likely influencing their perceptions of the value of the GMATE. Truly, one would expect students to place slightly higher trust in a performance test designed to measure their real communicative abilities.

While practical English ability is a highly-valued asset when applying for jobs and promotions in South Korea, test scores are still fairly standard, required documentation for these endeavors. Many university students in Korea, therefore, spend their vacation periods attending cram schools to prepare for standardized English tests, hoping to pad their resumes and get a leg up in an ultra-competitive society. Because of this, and the fact that a small number of large employers accept the GMATE and MATE, I expected students to value their GMATE scores as they concern future employment opportunities. To the contrary, only a third of participants agreed to any degree that “The GMATE will help (them) with (their) future careers.” A full 23.6% (82) strongly disagreed with this statement, while another 77 (22.2%) disagreed, and 72 (20.7%) somewhat disagreed. Only 2 students strongly agreed, while another 29 agreed (8.3%). Not only did student participants in this phase not place the expected importance on this exam when compared to other material in their English classes, but they also felt that this test was of little value to them after graduation.

As seen earlier, the student participants in this phase of the study held fairly strong beliefs that English exam scores were accurate representations of their true abilities. However, their trust in the GMATE as a fair test of their English skills was very different. Students overwhelmingly disagreed with the statement, “The GMATE is a fair test of my English ability,” with 298 (85.9%) disagreeing, and 41 (11.8% of all participants) of the remaining participants only somewhat agreeing. These responses are clear indicators of students’ feelings towards the exam as a language assessment tool, and are perhaps related to students’ not understanding why they must take the GMATE.

In this section, we have seen that students held quite negative beliefs regarding the fairness, importance, and usefulness of the GMATE. It would follow, then, that these students might think that students shouldn't be subjected to such an exam. However, and to my surprise, almost half (47.3%) of all participants agreed that, "All students at the university should be required to pass the GMATE." Sentiments here were fairly evenly spread across all responses, with 20% of respondents somewhat disagreeing, 20% disagreeing, and just under 13% strongly disagreeing. It came as a bit of shock to see the previous trends unfold, only to be followed by feelings that despite their belief that the GMATE was a "bad" test (in that it is unfair, useless, and unimportant), participants thought that it should remain as a university requirement.

5.2.3 Student Test Preparation for the GMATE

Student participants in this phase not only put their faith in their teachers and classwork for English test preparation in the past, but also believed that the skills they were learning in class were directly applicable to the GMATE assessment, shown by students' responses in Table 5.2 below.

Table 5.2 Student Responses to GMATE Preparation Items

Item	Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Strongly Agree
The English skills I have learned are related to the GMATE.	5	14	16	100	166	46
I have taken practice GMATE tests.	85	113	21	43	66	19
I do activities in class that are directly related to the GMATE.	8	13	37	126	130	33
My instructor prepared me adequately for the GMATE.	1	4	18	77	159	88
My instructor spends too much time preparing students for the GMATE.	40	175	79	40	11	2
I study directly for the items on the GMATE outside of the classroom.	121	122	50	38	14	2

An overwhelming majority, 312 (89.9%) agreed to some degree that, "The English skills I have learned in class are related to the GMATE." It must be noted, however, that the vast majority of these students, as freshmen, would not have actually taken the true GMATE exam before participating in this research, and thus would have been drawing this conclusion based perhaps on what their instructors or older classmates had told them.

Focusing further on the classwork, activities, and tasks with which students were presented, students were presented with statements regarding what types of in-class preparation they do. Interestingly, despite instructors' access to previous test items, and the fact that teachers reported using these in class to help students prepare for the GMATE, 219 (63.1%) reported that they hadn't yet taken practice or mock GMATE exams, despite the final questionnaire being administered in the days just prior to the exam itself. This conflicted with teachers' accounts of how they helped students prepare for the GMATE, as all but one of them explained that they provide students with past test items and class time to prepare for the GMATE as closely as is possible to the actual exam. On this note, though, students may have felt that their practice for the GMATE was not related to actual "practice exams," as the GMATE itself is taken via computer, and none of the teachers had access to more than one computer in their classrooms.

Providing a possible explanation for the contradiction between instructors' and students' accounts of the use of practice test items in class, a very large majority of student participants agreed to some degree that they had done "activities that directly related to the GMATE during class time." 289 (83.3%) agreed with this statement to some degree, showing that they either considered test practice items to be directly related to the test, although not necessarily practice tests on the whole, or perhaps students were able to draw clear connections between what they were doing in class and what they knew of the GMATE content and requirements.

In addition to students' perception that the class content was sufficiently related to the GMATE, participants also expressed strong beliefs that their instructors had properly

prepared them for the GMATE. 325 of the 347 participants (93.7%) agreed with the statement that instructors had "adequately prepared them for the GMATE," with 88 of those students strongly agreeing. In a similar vein, when presented with the statement, "My instructor spends too much time preparing students for the GMATE," 294 (84.7%) disagreed, which could be a reflection of the importance of exam preparation to them, or that their instructors are acutely aware of their students' test preparation needs.

Student behaviors and study habits within the walls of the classroom can be strongly influenced by the tasks and activities presented by their instructors, but what they do on their own time is directed as such to a lesser degree (specific homework, for example, is an exception). 293 (84.4%) participants responded that they "did not do activities or tasks related to the GMATE outside of class time," even though some instructors related in interviews that they assigned such activities as homework. Here, students' responses provide more support for the idea that students depend on their teachers' directions and in-class work to prepare for exams.

5.2.4 Student Emotions Related to the GMATE

Despite students' views that the GMATE was of very little importance to them, when compared to the other materials they were learning, students had some interesting responses to statements regarding emotional influences of the exam. Student responses to items related to emotion are shown below in Table 5.3.

Table 5.3 Student Responses to GMATE Emotion Items

Item	Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	Strongly Agree
The GMATE causes me stress.	25	45	40	109	60	68
I am afraid of failing the GMATE.	28	42	26	86	67	98
I would prefer not to take the GMATE.	4	26	29	97	71	120
I change the way I study English because of the GMATE.	9	37	56	134	91	20
The GMATE motivates me to study English harder.	56	95	79	86	26	5

Students were given the statement, “The GMATE causes me stress,” and an incredibly large number of students agreed with this. 237 (68.3%) of participants here agreed to some degree with this statement, while 60 of those respondents agreed, and 68 strongly agreed. Only 25 (7.2%) students strongly disagreed, showing that although there may not be much attention paid to the GMATE for its importance, it still remains in the thoughts of students from an emotional standpoint.

While tests in general are often stressful events for many students, I was curious as to the source of the stress for these participants. As most of these students were freshmen, and thus couldn’t actually “fail” the exam, due to the recent change in policy, it was quite a surprise to see that a great number of them actually harbored fears of failing. 251 (72.3%) of participants agreed to some degree that they were “afraid of failing the GMATE,” with 98 (28% of all respondents) answering that they strongly agreed. Only 28 (8%) strongly disagreed. This is a bit odd, as level 3 students would have already been identified at proficiency that would allow them to achieve the highest score possible on the GMATE, even without specific instruction.

It follows, then, that students would likely avoid the GMATE, if possible, despite their belief that all students should be required to take the exam. This was indeed the case, as 288 (83%) responded that they “would prefer not to take the GMATE,” with only 4 (1.1%) strongly disagreeing with that statement. Although this response trend is in opposition to students’

earlier beliefs that the GMATE exam policy should require all students to take and pass the test, it is clear here, as outlined by their emotional responses to the GMATE, that it is an experience that they would certainly rather sidestep.

Whether out of fear of failure, or for some other reason, I wondered whether students actually shifted their study strategies regarding English as a result of the GMATE. 245 (70.6%) agreed to some degree that they "change the way (they) study English because of the GMATE." In detail, though, only 20 (5.76%) of all participants strongly agreed that they had changed their English study methods due to the exam. Although there seemed to have been changes made, it would be necessary to collect data on these in much more detail in order to determine to what degree these washback effects had occurred.

Lastly, I was interested in any positive washback effects that might be attached to a motivational or emotional component. Students were presented with the statement, "The GMATE motivates me to study English harder." I was careful to make it clear here that the item indicated a motivation to study *English* harder, rather than simply to study harder in general, which would follow in a system well known for its test preparation. 230 (66.3%) of respondents disagreed with this statement, and of the remaining participants that agreed to some degree, 26 (7.5%) agreed, and only 5 (1.4%) strongly agreed. From students' responses in this section, we see that although the test may be of little overall importance to students, it nonetheless causes them stress, perhaps due to fear of failure, and doesn't seem to have overall positive washback on their motivation to study English. However, there were many students who responded that the GMATE did in fact affect the ways that they studied English, so it may be interesting to delve into the qualitative nature of this change in future inquiries.

5.2.5 Student Differences Based on Level

Student responses varied across items, and I wanted to investigate whether or not their level, as dictated by their English placement exam scores, would correlate with their questionnaire

responses. Prior to calculating Spearman's correlation coefficient values for the Phase I student questionnaire data, I chose to set the significance level at .05 for non-directional alpha (a = .05). Based on Ramsey's (1989) updated table of critical Spearman's correlation coefficient values at this level, I accepted values greater than +/-.197 as significantly correlated for n=100 (the largest n available in this table of critical values).

In Phase I, two items showed a slight positive correlation with student level, while three others were moderately, negatively correlated with level. Item 1, related to having positive past experiences with English exams, showed a correlation of .22 with level. This made sense, as students in higher levels would likely have enjoyed greater success on English tests, thus perhaps influencing their recollection of the examination experiences themselves. In a similar vein, item 23, concerning the importance of being able to speak English, was correlated with students' level at a value of .20. Here there may be some correlation with motivation due to the perceived importance of English, and English scores and level. However, the data in this study showed only a slight positive correlation, regardless of the cause.

Item 6, which stated, "The GMATE causes me stress," was negatively correlated with student level at -.36. As students in level 3 would have minimally begun the term at the lower border of the highest score bracket (Moderate High), it follows that these students would be the least stressed about achieving a good, and in some cases, passing, score. Even for those students in level 2 for which the GMATE was higher-stakes, their beginning ability, as determined by the placement exam, would have allowed them to likely earn a score of Moderate Mid, which perhaps resulted in less stress. However, level 1 students, who are categorized by the placement exam as starting their studies at Moderate Low level, would have likely felt the most stress regarding the GMATE, regardless of the stakes. This was further illustrated in students' responses to item 15, regarding students' fear of failing the GMATE. Level was negatively correlated with this item at -.45, showing that even those students for whom the GMATE was not a pass/fail exam, due to their falling under new policies, the prospect of failure was a very real concern in the lower levels.

Following this trend, student wishes regarding not taking the GMATE (item 21) were negatively correlated with level at -.32. That is, students in lower levels responded more that they preferred not to take the GMATE, while upper-level students responded in this way less so. Continuing aforementioned ideas regarding stress and failure, it is understandable that lower level students would respond in this fashion.

In Phase II, though, there were far more, and stronger correlations between items and students' English level. Nine items were shown to be positively correlated with student level during this phase of the research, while only one was negatively correlated. The only item showing significant negative correlation at the .05 cutoff for critical values was concerning fear of failing the GMATE (item 15). This makes sense to some degree, as student participants taking courses during the intensive term were more likely to have failed the GMATE in the past. Also, because there were more students for whom the GMATE was a higher-stakes exam studying during this period, the negative correlation for these students was slightly weaker, at only -.26, as compared with -.45 during Phase II.

Only two other items were significantly correlated with level both in Phase I and II. Item 1, about positive past experiences with English exams, showed a very strong positive correlation of .41. While this item was also positively correlated with level in Phase II, it was so much more moderately (.22). Furthermore, item 23, related to the perceived importance of English speaking skills, was slightly more strongly correlated with level in Phase II, at .26, but this was not as strong as I would have expected due to the greater number of older student participants during this phase, who were more likely to be preparing to enter the job market. As stated in Chapter 2, English speaking skills are often prized by employers in Korea, and are often an area of the job application in which potential employees believe they may be able to gain an advantage over the competition.

Item 5, which states, "I study directly for the items on the GMATE outside of the classroom," was positively correlated with level in Phase II at .24. Level 1 teachers often mentioned providing their students with additional outside GMATE preparation, but they also often said

that they spent a large amount of time in class focused on GMATE item practice. Conversely, although teachers of higher levels during the intensive term also mentioned that they made the GMATE a priority, they used less class time for test prep, and therefore may have assigned activities and practice items as homework more often. Indeed, this notion is supported by student responses to item 18 ("My instructor provides us with extra materials for the GMATE"), which was positively correlated with level at .24. Furthermore, level 3 students may have been more capable of successfully attempting and completing outside GMATE preparation on their own, due to their higher level of proficiency allowing them to better understand directions and assignments.

Item 9, regarding student awareness of the format and items of the GMATE, was positively correlated with level at .35. I thought that this may have been due to an increased ability to understand teachers' explanations of the GMATE more than the teachers' descriptions themselves, as all of the instructors during this phase of the study said that they addressed the GMATE with great consistency. Similarly, item 10, which concerns how well the instructor prepared students for the GMATE, was correlated with level at .37. This may have been due to differences in confidence among students, as those students in higher levels were perhaps more likely to succeed relative to those students of lower abilities. Finally, students' perceived understanding of why the GMATE was necessary (item 11) was also positively correlated with their level at .32. Again, this may be related to students in higher levels being able to understand their teachers' explanations of the exam, or perhaps that they are more able to direct this type of question to their instructors.

A much weaker positive correlation of only .21 was shown between level and students' belief that the GMATE would help them with their future careers (item 13). While a stronger correlation here in Phase II than in Phase II, maybe due to the fact these students are older and thus closer to entering their future careers, it is nonetheless a fairly weak correlation, and is not very surprising, as very few Korean corporations recognize or accept GMATE scores on job applications. Lastly, item 22, stating that "the GMATE motivates me to study English

harder," was positively correlated with level at .3. This is a moderate correlation, but may indicate some positive washback on students' general English study habits in this context.

5.2.6 Student Differences Based on Year

While still using Ramsey's (1989) .05 cutoff for critical correlation coefficient values for n=100 (.197), there were only three items that positively correlated with year in Phase I, and none that showed a negative correlation. However, in Phase II, eight items were positively correlated with year in school, and one showed a negative correlation. In Phase II, item 6, related to the GMATE causing stress, was correlated with student year in school at .25. Although the majority of students taking English courses during the regular term were freshmen, older students do in fact take the course, and those students for whom the exam was higher stakes may have felt more stress. Furthermore, students who were in later years of their university education careers were more likely to agree that they were "afraid of failing the GMATE," as item 15 was positively correlated with year in school at .28. This is not terribly surprising, of course, as only freshmen and sophomore students are subjected to the updated testing policy under which the GMATE is a lower stakes exam. Lastly, and following in the same vein as the previous item correlations, item 25, related to students' preference not to take the GMATE, showed a correlation with year in school of .23. These three correlations begin to paint the picture that the higher stakes exam may be more stressful for those students who must pass it in order to pass the course and graduate from the university, although as was discussed in the summary of Phase I questionnaire responses, the GMATE seems to be a fairly stressful event in the lives of student participants regardless of the stakes.

Turning to Phase II, it was particularly interesting that year in school was positively correlated (.34) with the statement, "I have an overall positive experience taking English tests" (item 1). This was a bit surprising, as students taking courses during the intensive term are very often re-taking the course after having failed the GMATE in a past term or terms. Similarly, students in later years of university during this term seemed to put more faith in the validity of English tests, as item 2 ("English tests are an accurate representation of my English ability")

was correlated with year in university at .35. As a minor surprise, item 13, dealing with students' knowledge of the GMATE format, was positively correlated with year in school, but at only .22. I had somewhat expected this to be a much stronger correlation, again, as I assumed older students would have had previous GMATE experiences, and therefore would be more familiar with the exam itself.

While students may have not necessarily been crystal clear about what appeared on the GMATE, and in what format the exam was administered, older students were more likely respond that they "understood why they must take the GMATE" (item 11), as this item was correlated with year in school at .34. Whether or not this understanding was due to teachers' explanations or perhaps a tendency of older students to seek out such information is not evidenced here. When looking at what students were actually doing in the classroom, and comparing that with teachers' lesson plans and recollections of what they did, the fact that item 12 was correlated with year in school at .3 is very interesting. This item states, "I do activities in class that are directly related to the GMATE," and because level is not significantly correlated with year in school (.02), it is a bit odd that year in school would be correlated with students' agreement with this statement. Intensive term teachers, regardless of level, generally responded during interviews that they placed a major priority on early and consistent in-class GMATE practice, so I would expect there to be very little correlation related to this item, as all levels and years would respond similarly.

Student participants during Phase I that were in later years of their university studies were more likely to respond that the textbook helps to prepare students for the GMATE (.23). This was surprising, as the vast majority of teachers believed that the textbook was directly connected to the GMATE, and indeed, the text contains several pages tied to GMATE preparation through practice items and exercises. Knowing this, I would have thought that there would exist no significant correlation with this item. In a small example of possible positive washback, item 22, related to the GMATE motivating students to study English harder, was correlated with year in school at .27. It may have been that the GMATE

positively affected students' drive to become more proficient in English, but this would require follow-up inquiries with the student participants in order to accurately make such a claim.

However, older students in this phase of the study were more likely to agree that "it is important to be able to speak English" (item 23), as this item was correlated with year at .26. Placing importance on the ability to communicate in English may have been tied to students' overall motivation to learn, therefore influencing their responses to the aforementioned item 22, and vice versa. Finally, and perhaps most shocking, was the negative correlation of student year in school with item 21 ("I would prefer not to take the GMATE") at -.21. While not a very strong negative correlation, this finding is in direct opposition to that shown in the Phase I data, explained earlier. At present, I have no explanation for why older student participants studying during the intensive term would seemingly prefer to take the higher stakes GMATE, while those during the regular term (Phase I) would prefer to avoid it.

5.2.7 Student Differences Based on Seasonality

Looking at trends in student questionnaire responses throughout the term, it was difficult to find any meaningful differences that couldn't be possibly explained by the aforementioned factors of age, level, or year. In one class, for example, responses directly related to the GMATE spiked towards the end of the term, particularly concerning those items regarding emotion and test preparation. This is shown in Table 5.4 on the next page.

Table 5.4 Student Questionnaire Responses in Agreement (Strongly Agree)

Item	Beginning of Term	Middle of Term	End of Term
I have taken GMATE practice tests	0 (0)	3 (2)	17 (15)
I study directly for the items on the GMATE outside of the classroom	0 (0)	0 (0)	7 (4)
My instructor prepared me adequately for the GMATE	0 (0)	0 (0)	12 (5)
I do activities in class that are directly related to the GMATE	0 (0)	2 (1)	14 (9)
I am afraid of failing the GMATE	14 (10)	15 (10)	14 (8)
The GMATE causes me stress	19 (12)	19 (14)	19 (15)
My instructor provides us with extra materials	0 (0)	4 (0)	16 (10)

It is interesting to note in the table that responses related to student stress were fairly constant for students in this class from the beginning to the end of the term, while there were marked differences in responses related to classroom activities and preparation throughout the term. While this first appeared to be evidence of seasonality, it seemed to be a direct result of the teacher's increase in talking about and teaching to the GMATE exam towards the end of the term, although interviews with this teacher did not support this, despite my direct line of questioning regarding this issue. In another class, for example, the teacher reported addressing the GMATE and employing preparation activities fairly consistently from the beginning to the end of the term, and in this class there were no real changes in overall student responses related to the GMATE. One reason for this may be that students receive the vast bulk of their information regarding the exam format, structure, items, and importance

from their instructors, and thus their perceptions of the GMATE might be more dependent on their teacher's behavior. Although teachers reported creating student-centered learning environments, these students, and in particular freshmen students, will have been much more accustomed to relying upon their instructors for this sort of knowledge, while taking a more passive role in test preparation.

I believe it is likely that instructors' actions and behaviors influenced students' questionnaire responses to some degree; however, it is difficult to precisely correlate these, as data regarding the former consisted of mostly instructors' recollections of classes combined with lesson plans, due to the limited number of observations that I was able to conduct. Without complete observational data, it is difficult to ascertain whether or not class activities and lectures influenced students' responses. For example, in one class, a teacher was observed talking about the GMATE, but when we reviewed his lesson plans and notes during a subsequent interview, he didn't recall directly addressing the GMATE in any way on that date. This sort of situation could have presented itself in multiple cases, therefore confounding some conclusions I would have drawn regarding seasonality.

For example, in a contradicting example to the aforementioned class, another class of students exhibited response patterns that were fairly consistent from the beginning of the term until the end. This is shown in Table 5.5 on the following page.

Table 5.5 Student Questionnaire Responses in Agreement (Strongly Agree)

Item	Beginning of Term	Middle of Term	End of Term
I have taken GMATE practice tests	19 (15)	19 (16)	19 (16)
I study directly for the items on the GMATE outside of the classroom	0 (0)	16 (12)	18 (12)
My instructor prepared me adequately for the GMATE	0 (0)	8 (4)	14 (9)
I do activities in class that are directly related to the GMATE	0 (0)	15 (8)	17 (9)
I am afraid of failing the GMATE	16 (10)	14 (7)	12 (9)
The GMATE causes me stress	19 (12)	17 (11)	19 (10)
My instructor provides us with extra materials for the GMATE	0 (0)	16 (12)	17 (12)

Here we see that responses lined up fairly well with the teacher's interview responses and lesson plans, as she started the term by explaining and outlining the GMATE, and even gave the students a practice exam during the first week as an introduction to the exam. She also drew connections with class and textbook activities to the GMATE, which supports the student response patterns above. One bit of interesting data in the above table is that student fears of failing declined only slightly despite the instructors' consistent GMATE tasks and practice, while student stress due to the GMATE dipped during the middle of the term, but rose again as the exam drew nearer.

Looking at seasonality in Phase II (the intensive term), there was almost no significant difference whatsoever from the beginning of the term to the end, in any of the participating classes. One reason for this could have been that there was only a week between questionnaire periods. Looking at the age and year in school data, though, reveals that the majority of students during the intensive period were retaking the GMATE or taking it for the first time in the slightly higher stakes pass/fail situation, which may have influenced their response regarding emotional influences of the GMATE. Furthermore, all of the teachers that participated in this phase of the research explained that they place a stronger focus on the test than during the regular term, as is shown in section 5.1 of this thesis.

To summarize, intensive teachers generally spent a greater amount of time talking about and preparing students for the GMATE regardless of level, which was a stark contrast to the trends of regular term teachers. This pattern of teaching behavior was shown even in participating teachers who taught both terms. Instructors were keenly aware of their students' priorities during the intensive terms, and it could be said that teachers in most cases adopted these priorities. That is, intensive term teachers sacrificed time and effort spent towards general English speaking and presentation skills in favor of GMATE preparation, but this didn't show any major fluctuations throughout the three-week intensive term.

5.2.8 Differences Based on Type of Course

While Sections 5.2.1 through 5.2.7 focused on student response from Phase I (the regular term), this section aims to examine student data from the final questionnaire in Phase II (the intensive term), and compare that with patterns summarized in previous sections of this chapter. Looking first at intensive term students' perceptions of English exams, and following the same structure as the summary of Phase I data, Phase II students generally agreed that they had had positive experiences with English exams, as 82 (67.2%) agreed to some degree, and 20 of those (16.4% of all Phase II respondents) strongly agreed. This is slightly lower than the 82.7% of Phase I students who agreed with this idea, and may have been due to the likelihood that Phase II participants, being older, would have taken more standardized English

exams, and therefore would have had more opportunities for disappointing or stressful testing experiences.

As they may have taken more English exams, these experiences could have reinforced their faith in English exams' ability to represent students' true ability. When presented with the statement, "English test scores are accurate representations of my English ability," 68.3% of participants agreed to some degree, which was slightly higher than the 57.3% of Phase I students who agreed. However, like the Phase I participants, the student responses from the intensive term showed a central tendency, with the bulk of those students in agreement agreeing only slightly (48 of the 82), while the largest number of disagreeing participants also disagreed only slightly.

In a shift away from Phase I response patterns, Phase II students were fairly evenly divided when it came to their use of materials separate from their assigned classwork to prepare for English tests. 54% of the intensive term students agreed that they study for English tests using outside materials, while 62% of Phase I disagreed, and many of those strongly disagreed. This may again be due to the fact that intensive term students tended to be older, and thus may have developed more independent study habits, whereas those students taking English courses during the regular term likely depended heavily on their teachers throughout public school, and perhaps had yet to move away from this style of teacher-centered exam preparation.

Moving to students' perceptions and understanding of the GMATE, students very much agreed with the statement, "I am aware of the format and items that appear on the GMATE," with 100 of 122 (82%) agreeing, and 30 of those 100 strongly agreeing. This comes at a bit of a surprise, though, as a great many of the intensive term students were retaking the course after having failed the GMATE in previous attempts, so they would have had more awareness of the format and items due to their own experiences. Furthermore, with the priority placed on the GMATE by teachers during the intensive term, it is quite odd that Phase II students

would report less understanding of the testing format and items than those in Phase II, of which 89% agreed with this statement.

While approximately 57% of Phase I students disagreed that they understood why they had to take the GMATE, 63% of Phase II students disagreed, with almost 20% strongly disagreeing. It may have been here that Phase II students felt defiant against having to pass yet another university-administered exam in order to fulfill graduation requirements, or that their teachers simply did not spend much time discussing such issues, but it was somewhat interesting, regardless. However, while Phase II students reported that they didn't know why they were being asked to take the exam, they seemed to place very high value on it compared to their Phase I counterparts. Students in the regular term were split down the middle when presented with the statement, "The GMATE is more important to me than other parts of my English class," while nearly 70% of intensive term students agreed. This makes sense when considering the difference in stakes for most of these students.

While Phase II students placed higher importance on the GMATE than did those students in Phase II, they also valued the ability to speak English at a slightly higher level, as well. Although 98% of Phase I students agreed that being able to speak English is very important for them, and Phase II only 95%, the Phase II students who agreed to some degree tended to agree or strongly agree more (40 and 52%, respectively) than did those respondents in Phase II. I thought that perhaps the Phase II students placed higher value on English speaking abilities due to the fact that they were approaching their entrance to the working world, but those feelings were certainly not extended to the GMATE. Although they valued the GMATE above their other English coursework, Phase II students generally disagreed (71%) that the GMATE would help them with their careers. It is interesting to note these last three response patterns, as these students seemed to value English ability on the whole, place importance on the GMATE as a priority in their course, but dismissed the exam as having any further significance or use beyond graduation.

While I mentioned earlier that students were generally trusting of English exams' validity, this was not the case for students' perceptions of the GMATE. Just as in Phase II, a full 87% of students in the intensive term disagreed that the GMATE is a fair test of their English ability. That this exam was researched and rigorously designed and piloted over many years is irrelevant here, as I was simply interested in students' feelings regarding the exam, and those sentiments seem to be very clear in this case. Lastly, and just as surprising as in Phase II, almost 44% of Phase II students agreed that all students should be required to take the GMATE, despite their previously reported suspicions regarding its validity or usefulness beyond university.

Moving on to perceptions of classwork and preparation related to the GMATE, Phase II students agreed only slightly more than their regular term counterparts with the statements, "The English skills I have learned are related to the GMATE." 91.8% of intensive term students agreed with this, while 89.76% of Phase I students agreed to some degree. However, in a very surprising response pattern, only 35% of Phase II students agreed that they had taken GMATE practice tests during the term, which is quite a bit lower than the 46.9% of regular term participants who agreed. While the Phase I responses were very surprising, as teachers had often mentioned during interviews that they gave practice exams at different times during the term, that the intensive term students would agree to such a low degree is even more astonishing, given that interview and observation data are at direct odds with this notion. Perhaps it was that students did not realize that they were taking practice exams during the regular term, or that the intensive term students, who would have likely had real GMATE experiences, did not associate mock exams with the real test because they were not done on the computer, as the actual exam is administered.

Interestingly 87% of Phase II students agreed that they had done activities that directly related to the GMATE, which was very similar to the responses given by Phase I students (82.4%). Here, too, I am inclined to wonder why students more often disagreed that they did mock and practice exams, when this was in opposition to observations and teacher interviews, but here overwhelmingly agree that they did related activities. Although students may have

considered practice tests and activities to be separately related to the GMATE, but this still would not explain students' not connecting their in-class practice tests with the GMATE, particularly when intensive term students would have had more accurate knowledge of the exam format and items themselves.

When presented with the statements, "My instructor adequately prepared me for the GMATE," intensive term students agreed in a similar fashion (91.8%) as regular term students (93.6%). This is positive news for instructors, as they geared their courses towards the GMATE during the intensive term, and the 8.2% of students who disagreed in Phase II only slightly did so. However, it was interesting that Phase II students were much more divided concerning the amount of class time that their teachers allotted for GMATE preparation. 84.7% of Phase I students disagreed that their instructors spent "too much time preparing students for the GMATE," while only 57.1% of intensive term students disagreed, and many of those (21.3% of 57.1%) only slightly disagreed. This is peculiar, because intensive term students certainly place high value on the GMATE, are stressed by the exam, and have a real fear of failure, but it may be that the priority teachers place on the GMATE during the intensive term is such that many students feel that other tasks could receive more attention.

Lastly, there was a discrepancy between the student responses to the statement, "I study directly for the items on the GMATE outside of the classroom." 84.4% of Phase I students disagreed with this, which as mentioned prior, is strange because teachers during the regular term often assigned GMATE practice items as homework. However, Phase II students also disagreed (63.1%), but intensive term teachers were more split about how much GMATE-related homework they assigned, so it is interesting that the responses in both terms seem to be in opposition to the perceptions and recollections of instructors.

Looking at items related to student emotion in Phase II, 87% of students agreed that the GMATE causes them stress, with 42.6% strongly agreeing. This was quite a bit higher than the 68.3% of Phase I students who agreed, perhaps because students studying during the intensive term were taking the exam as a higher-stakes test, or because many intensive term

students would have had experiences failing the GMATE in past terms. It follows that Phase II students would harbor more fear of failing the GMATE, and this was the case. 88.3% of intensive term students agreed that they were afraid of failing the GMATE, while only 72.6% of Phase I student participants agreed. The combination of more often being caused stress due to the exam coupled with the fear of failure would lead one to think these students would be more likely to wish to avoid taking the GMATE. Here again, it was clear that these feelings were related, as only 83% of regular term students agreed that they would prefer not to take the GMATE, while a full 92.8% of intensive term students agreed. These response patterns may lend some support to the notion that higher-stakes English language exams cause students more stress, and deserve further inquiry.

Turning my attention to GMATE washback effects regarding students' study habits, a slightly lower number of students in Phase II agreed that the GMATE influenced the ways they choose to study English in general (60.7%), while 70.6% of regular term students agreed. I have to wonder if this was because older students had developed their own study methods or practices throughout their university careers, while the regular term students, who were more likely to be in their first year of study, may have been more affected by exams. Finally, the GMATE seemed to be a demotivating factor in regular students' English studies, as 66.3% of those learners disagreed that the GMATE motivates them to study English harder. Conversely, 53.1% of Phase II students agreed with this notion. Again, I am curious as to whether or not this difference in motivation may have stemmed from the difference in stakes between the two terms, or from something else entirely.

Chapter 6: Discussion

In this chapter, I focus on several issues from the findings, and connect them to the literature. After briefly presenting a summary of the key findings from chapter five, I organize and discuss important issues in three categories: exam stakes and perceptions, teaching and learning, and contextual factors.

6.1 Summary of Key Findings

The findings of this study present some interesting issues occurring in this particular research context. Teachers in the study tended to use the textbooks and materials created by the department faculty members for all classes regardless of focus on test-related items or otherwise, and paid little regard to the stakes of the exam for their students during the regular term. This slight indifference was shown by teachers' consistency in activity and task selection across class levels. Instructors tended to spend more time talking and lecturing in lower-level classes, however, and direct preparation for exam items decreased as level increased. Regular term teaching washback effects were shown in a somewhat different way than those occurring during the intensive term. Teachers participating during the three-week intensive session reported omitting much more content from their original syllabi in order to focus more on the GMATE, and held stronger negative perceptions of the test and policy during this shorter teaching period. These teachers also assigned fewer assignments, choosing instead to practice for the exam during much more of their available class time.

The students in this study provided me with an interesting look into their experiences through their questionnaire responses. Washback effects on their learning and study habits seemed to have been occurring, as they reported to have changed the ways they studied English based on the GMATE, although they responded that the study materials provided by their teachers were most important to their success. Furthermore, these students knew that their test scores would be significant in their lives only in the very short term, but still believe that the test overall was very important, and felt stressed or worried about GMATE failure even

though many of the students in this study were incredibly unlikely to receive such an outcome due to their having already achieved a certain corresponding score on the placement exam prior to the start of the term. Finally, these concerns and emotions persisted throughout the term, and did not necessarily change in intensity as the exam date approached, despite many teachers reporting that they focused class activities on test preparation closer to the end of the term.

6.2 Exam Stakes and Perceptions

Exam stakes are mentioned throughout the literature in different ways, but by far the main focus of most studies has been on high-stakes exams, and the washback effects that may occur dependent on those tests. Indeed, even the Washback Hypotheses state that "Tests that have important consequences will have washback," and, "Tests that do not have important consequences will have no washback." (Alderson & Wall, 1993, p.120-121). However, the stakes of the exam in this study varied across students and teachers in both perceived and real ways. As noted in chapter six, there were some students for whom the exam was very much a high-stakes one, acting as a gatekeeper to their ability to pass the English course with which it was connected, and ultimately graduate from the university, too. For others, though, the exam was simply a portion of their class grade, and not weighted particularly heavily relative to other assignments or tests. While it would have been ideal for the GMATE to have fit into the model proposed by the Washback Hypotheses, it was not nearly so neat and tidy.

Shohamy, Donitsa-Schmidt, and Ferman (1996) investigated tests of varying washback in the same context, and reported findings that supported the Washback Hypotheses. In this study, teachers made little to no change in their teaching approaches regarding the lower-stakes Arabic language test, while teachers of English were very strongly affected by the higher-stakes English language test. In my study, though, the teacher participants were dealing with the same test of different stakes for their students, which were often in the same classroom together. This created a complex teaching and learning environment with unique implications

for washback effects. Discussed in sections 5.1.1 and 5.1.2, there were conflicting ideas shared regarding how to best adapt the curriculum, materials, and teaching methods to the students' needs based on the stakes of the exam.

Beyond the *actual* stakes of the exam, though, are the *perceived* stakes of said test (Green, 2007). Although the stakes of the GMATE for students could be easily defined by whether or not they had to achieve a certain score in order to pass the course and graduate from the university, the stakes of the test in teachers' views were not so clearly laid out. This is in line with recent findings presented by Cheng et al. in that perceived value and stakes of an exam can be very much dependent on social contexts (2014), and that the value that is placed upon the test itself determines its stakes (Xie, 2015). Teacher participants who viewed their students' GMATE results as a reflection of their teaching ability, for example, held the stakes of the GMATE as much higher than for those teachers who did not feel this way.

Accountability to one's students is also reported by Tsagari (2011), as she writes that "accountability towards parents and employers made teachers feel stressed and anxious as success in the exam was the yardstick their professional status was judged by." Here we see that teachers' perceptions of the exam being of certain stakes might impact the teaching methodologies that they employ while preparing their students to take the test. Feelings of added pressure as a result of testing policy change are a common finding in the washback literature (Solorzano, 2008), but these reports of stress or fear are usually connected only with high-stakes exams. It was particularly interesting in this study that the stakes of the GMATE were perceived very differently between teacher participants, and even within individual teachers, depending on when during the term they shared their views.

Although perceptions of the stakes of the GMATE varied much during the regular term, there was more consistency between teacher participants during the intensive term. These findings support Shohamy's (1993) conclusions that there would exist some narrowing of the curriculum. However, in her study, Shohamy reported that teachers did not vary their instruction much, citing the materials as a possible explanation for this phenomenon. In the

current study, conversely, teachers reported making wild departures from their usual methodological choices when the stakes of the exam, both actual and perceived, were clearly higher. This “subjective reality” is described by Fullan (1993) as often conflicting with objective reality, mandating that educators navigate testing and education policy in their own ways.

A clear difference between the current study and the aforementioned research studies was the intent and aims of the stakeholders. In several papers (Qi, 2004, 2005, 2007; Shohamy, 2001, 2007) high-stakes test implementation was done with very real purpose. That is, high-stakes exams were put in place in order to effect specific, intended changes within an educational context. In the current study, however, there was no explicit explanation of the GMATE as such a vehicle for change. Teacher participants made no mention whatsoever of testing policy having been created in order to change the English program or curriculum, and even those teachers who had been teaching in the program since before the creation of the GMATE expressed no such ideas relating to the possible intent of university administrators. The student participants in the current study were subjected to objectively different stakes regarding the consequences of their GMATE scores, yet their perceptions of the GMATE as stressful, valid, or important did not differ to the same degree.

The literature supports the notion that students for whom an exam is low-stakes will exhibit lower washback (Cheng et al, 2014; Shohamy, Donitsa-Schmidt, and Ferman, 1996; Xie, 2015; Xie & Andrews, 2013), and indeed, this is such an important tenet in the field of CLT that it is found as one of the Washback Hypotheses (Alderson & Wall, 1993, p.121). However, in the current study, there were few real differences shown between the groups of students taking the low-stakes version of the GMATE, and those required to attain a minimum score in order to pass the course and graduate from the university. In fact, the area that showed correlations between year in school (which translated to exam stakes, as students having started school earlier were subjected to the higher-stakes version of the test) was related to stress and emotion, while other factors as measured by the questionnaire were relatively similar between students of different years in school.

This characteristic, though, has been mentioned in previous studies, as Bachman and Palmer (1996) proposed that taking and preparing for a test can affect students, which would make sense in consideration of many older students in the current study having had taken the exam prior to the research period. Furthermore, Shohamy et al. wrote that “even minor changes in the test cause strong washback in terms of anxiety, fear, the production of new material, and in the teaching and learning behaviours of those involved” (1996, p.314). I would argue, however, that changing the stakes of an exam do not necessarily constitute *minor* changes, in this context, and that there are perhaps other factors important to explaining this key difference from the literature.

Students’ motivation in connection with exam stakes is another point worth mentioning here. Watanabe (2004, p.129) writes that a large number of washback studies concern themselves with student motivation as a key washback effect, and that some proponents of exam-driven curriculum suggest that without high-stakes tests, students would not study sincerely. Previous papers have shown that students will put forth much less effort in preparing for and taking an exam when the stakes are lower (Wise & DeMars, 2005; Xie & Andrews, 2013; Xie, 2015). In the current study, there was some very weak evidence to support this idea, noted in section 5.2.6. As shown previously in chapter six, it may have been that older students had their eyes on their future careers, and were keenly aware that English skills would help them with their goals in securing employment. Thus, the higher valuation of English skills may have worked as a motivator for those students, and affected their responses regarding motivation in connection with the GMATE. However, Rodriguez and Arellano found that failure on an exam can demotivate students and lead to negative outcomes on subsequent scores (2016).

Finally, there were some very interesting correlations that occurred with regard to students’ perceptions of English exams, and of the GMATE in particular, when exam stakes were considered. The results of the questionnaire data analysis show that students for whom the GMATE was higher stakes were more likely to place higher value on the test, be aware of the items and format of the test, and believe that the test was not a valid indicator of their English

levels, but were not more likely to respond that the GMATE caused them stress, as virtually all students responded that the GMATE was stressful, regardless of level or year in school. These findings support conclusions presented in the literature (Xie & Andrews, 2013; Xie, 2015).

Conflicting results were reported in the 1996 study by Shohamy et al., as they wrote that 90% of their student participants reported *not* knowing what the test in that study covered, and 64% believed the test was a poor indicator of their language ability. 82% of Phase II students reported being aware of GMATE format and items, and a full 89% of Phase I students do so, too, while 87% of students in both phases of the current study thought the GMATE was not a good indicator of their English language abilities. However, students in the Shohamy et al. study believed the Arabic language test in that case was of "no importance," while students in the current study tended to place greater importance on the GMATE above their coursework, but generally responded in the questionnaire that the GMATE was of little importance to them beyond graduation from university.

6.3 Teaching and Learning Processes

As described in Bailey's (1993) model of washback, teaching process are any actions taken by teacher participants that relate to the process of learning. Creating and using specific materials or activities, organizing students into certain groupings, and employing particular teaching methodologies are all behaviors which could be considered as teaching processes in this model. In the current study, teachers mentioned and exhibited a great variety of teaching processes that in many ways supported the literature, with some exceptions. Teachers cited the GMATE as a reason to conduct teacher-centered lectures in class, with a major focus on test items and activities. Some similar findings were reported by Shohamy et al. (1996, p.301):

> "Teachers stopped teaching new material and turned to reviewing material. Teachers replaced class textbooks with worksheets that were identical to

> previous years' tests. The activities were all 'testlike'. Review sessions were added to regular class hours."

In the current study, too, I was able to not only hear from teachers during interviews that these sorts of things were occurring, but I was able to see for myself during observations that there existed a lot of overlap with the aforementioned findings and my own research. As noted in the findings chapter, some teachers focused the curriculum in response to the exam in order to review test materials, and they dedicated whole days or even weeks only to GMATE preparation, resulting in the need to omit certain sections of the textbook from their syllabi. Furthermore, these activities were generally reported and observed as simply being mock tests using previous years' test items, precisely in line with what Shohamy et al. observed in their 1996 study.

Au (2007) identified three key themes when reviewing a large number of washback studies that presented themselves to varying degrees in the current study. He wrote that the washback effects noted in a great majority of the papers he examined were taking place in the form of content control, in which the actual content was contracting or expanding due to the test, formal control, in which teachers taught in smaller chunks directly related to the test, and finally, pedagogic control, that support a pedagogic shift towards more teacher-centered instruction in reaction to testing changes (p.262-263). Interestingly, all of these themes could be seen in the current study, in the previously mentioned narrowing of the curriculum, the focus on individual test-related skills such as picture or process description tasks, and the tendency for teachers to build test preparation lectures around themselves in a less participatory format.

Two influential washback studies found that washback effects influence content much more than methodology (Cheng, 1999; Wall & Alderson, 1993), but this was supported only in part by the findings of the current study. Across the 17 participating teachers, for example, 16 of them stated that they omitted sections of the textbook in direct connection with the GMATE, while 12 said that they taught in very different ways because of the exam. While slightly

fewer of the teachers in the current study showed or mentioned methodological influences from the GMATE, there were still enough of these teachers that in this context, the exam seems to be exerting fairly even influence over content and teaching methodology. In explaining this contrast, it is necessary to look at some of the more detailed findings from past studies. Wall and Alderson (1993) wrote that "many teachers are unable, or feel unable, to implement the recommended methodology. They either lack the skills or feel factors in their teaching situation prevent them from teaching the way they understood they should." In the current study, though, there were few if any mentions of an inability to properly prepare students for the GMATE using appropriate methodology. In fact, there was a widespread confidence among the participants in their general ability to do just that. A recent paper supports the notion that teachers' learning experiences as students influence later methods choices when teaching, particularly as those choices relate to what *not* to do (Moodie, 2016), and therefore this may have led teachers to be more sure of their classroom approaches.

As previous studies have correctly highlighted, however, there are individual factors unique to teaching participants that must be considered when drawing conclusions regarding washback effects. In Watanabe's 1996 study, for example, two teachers were observed to be fairly consistent in their choice of teaching methodology independent of the course. He presented three possible explanations for these findings (p.330-331):

1) the teachers' educational background and/or experiences
2) differences in teachers' beliefs about effective teaching methods
3) the timing of the researcher's observations

In the current study, there was variation in the age of teacher participants, but much less so when looking at educational background and experience teaching in higher education, particularly in programs similar to their current positions. There were, however, some major differences in what teachers considered to be best teaching practices, particularly concerning preparing their students for the GMATE. Also, in line with Watanabe's suggestions above, I was only able to observe a small group of the 17 teacher participants, which limited my ability

to more carefully compare what was said was done during interviews with what was actually done during observations. Watanabe reported that when he spoke with two teachers about their teaching methodology practices, both said that they used the same methods despite the course, which ran counter to the observational findings (1996, p.331). These differences manifested themselves not only in what teachers said and did during classes, but also how they incorporated the textbook into the their teaching. As shown in section 5.1 of this thesis, there were varying opinions and practices regarding the use of the textbook in classes, from teachers who valued it as an important part of their curriculum, to others who mentioned that they would rather not use a common textbook to teach their courses. Cheng (1997) wrote, conversely, that in her study teachers in Hong Kong were more reliant on the textbook, which had been similarly designed as a central pivot point around which classes could revolve.

Finally, teacher participants in the current study expressed feelings supporting findings published by Cheng (2004). In that study, teachers stated that they were increasingly worried about students' English ability levels as time progressed. Participants teaching level 1 in the current study, too, mentioned on several occasions that they "didn't think it was fair" to require level 1 students to attain the same minimum scores as higher levels. Comments of this sort were more common near the end of the term, suggesting that teacher participants were more worried about the issue at that time.

The ways in which students specifically behave regarding learning in response to exams are termed 'learning processes' in Hughes' model of washback (Bailey, 1999). The vast majority of influential washback studies have focused on teachers' perceptions and behaviors, and have even proposed that teachers are the key factor regarding washback effects (Baily, 1999, p.17-19). Bailey writes that "although language learners are the key participants whose lives are most directly influenced by language testing washback, there is relatively little research that documents their point of view or their washback-related behavior before and after tests" (1999, p.14). Other research, too, point to the importance of including multiple stakeholders when conducting washback research (Cheng, Andrews, & Yu, 2011; Fox & Cheng, 2016). I therefore thought it was an important facet of the current study to include student participants

in order to give them a representative voice when drawing conclusions regarding what is occurring in the research context.

Before discussing comparisons between the current study and the literature with regard to student learning processes, however, it is crucial that we remember that a disproportionate number of the major language exam washback studies to date have focused on standardized tests of reading, writing, and listening comprehension, or some combination thereof. In the current study, though, the GMATE is a holistically-rated speaking performance test delivered via computer, and therefore may show slightly different washback effects related to students' examination preparation. For example, one list of common learning processes as washback effects includes behaviors such as enrolling in test-preparation courses, or skipping language classes to prepare for the test (Bailey, 1999, p.13-14). In the current study, those behaviors were either not possible, in the case of the former, as GMATE preparation courses are not offered at the university, or not practical, in the case of the latter, because most teacher participants in this study were either observed or reported providing students with unique test preparation activities or opportunities during class time.

Student participants in the current study did report that they engaged in specific learning processes influenced by the GMATE, though these were generally directly connected to assignments, supplements, and homework that their teachers assigned to them. This was likely due to contextual factors, briefly described in the next section, as well as due to the uniqueness of the GMATE. That is, because the GMATE is only administered in the English program at this particular university, it is unlikely that students would have been able to find outside materials or sources to aid them in their independent test preparation.

When looking at learning process-centered washback effects cited in the literature, it is clear that there is very little supporting evidence shown in the current study. While questionnaire responses related to stress, anxiety, fear of failure, test importance, and test validity were very similar in this study to those findings reported in previous research, the specific actions and behaviors taken by students to prepare for the other high-stakes language exams were

not reported by students in the current study. Again, this may have occurred due to contextual differences related to the students themselves, or a combination of stakeholder factors and characteristics of the test itself, as the GMATE is a holistic, performance assessment of English speaking skills, and thus requires different preparation from those high-stakes language tests that focus on listening and reading comprehension, vocabulary, and knowledge of grammar.

6.4 Contextual Factors

The importance of context is mentioned repeatedly in the literature, with good reason, and it is similarly appropriate that I discuss the influence of context on the washback effects in this study. Cheng et al. looked at teachers in three different contexts around the world, and found differences based on social context (2014). However, an important characteristic of the research context in this study is that the teacher participants were all from outside Korea, while in the aforementioned paper the teachers were teaching in their home countries. In a similar example, Cheng (2004) worked with teachers from several different high schools known for sending students to “prestigious” universities, but made no mention of from where these teachers originated, allowing one to safely assume all were of Chinese origin.

Another study focused on the Israeli education system, and the teaching of Arabic as a second language (Shohamy, Donitsa-Schmidt, and Ferman, 1996). Again, with no specific mention of the teacher participants’ backgrounds, one can be perhaps taken that these educators were Israeli, or at least hailed from the region. As was covered in chapter two, regional and cultural differences regarding language, testing, and education can vary wildly (Cheng et al., 2014), and thus it was a unique advantage of this study to be able to look at a language assessment and teaching situation through the lens of teachers not necessarily educated within the dominant belief system of the research context.

Alternatively, it may have been that some teachers had not taken more Korean attitudes regarding education and testing, but that they had instead become hyper aware of these

phenomena, in part due to their time spent in Korea. For example, teachers said that they avoided teaching to the test because of negative views of the common grammar-translation approach here, while another wanted to avoid specific test preparation methods because of his experience in the American school system. That a teacher would avoid employing certain teaching techniques based on his or experience as a learner in the past is supported in a recent study conducted in South Korean schools (Moodie, 2016). These attitudes are further evidence of notions presented in the literature that washback "takes place within complex dynamic educational systems and social contexts. The complicated interactions between the factors that make up the micro-level teaching/learning context and the macro-level factors beyond the classroom make it difficult to attribute impact to any one factor alone, such as the test" (Cheng, Sun, & Ma, 2015, p. 453). However, that washback has been investigated and found in so many different testing contexts around the world is a strong statement regarding its *existence* across context and culture, although in what *form* and *intensity* these effects present themselves is a more challenging question to answer.

Korean teachers of English whose responsibility it is to prepare students for high-stakes English exams have been shown to often rely on grammar-translation approaches (Choi, 2008; Jung, 2008; Kim and O, 2002). While a common test preparation teaching methodology for teachers sharing a non-English first language with their students, this was not truly the case for the majority of the participants in the current study. However, this seemed to be due mostly to teachers' inability to speak the Korean language fluently. For example, some teachers reported preparing vocabulary or grammar translations to use in class, despite not being able to actually converse or communicate in Korea. In essence, this is at times very much in line with what occurs in classrooms during which times Korean teachers are using grammar-translation teaching methodologies.

These teacher participants may have been relying upon a common Korean teaching practice because they had been exposed to it, showing a contextual connection with their host nation, or they may have simply taken this approach as many teachers around the world, in varying contexts have throughout history when preparing students for language tests (Au, 2007;

Cheng et al., 2014; Cheng, Sun, & Ma, 2015). In any case, it must be considered that context, both current and past, plays an integral role in how washback effects are observed and perceived. As Bailey writes:

> "In considering the varied research about washback and language teachers, we can see that teachers' classroom behavior can either support or override the intended positive washback effect of new or revised tests. There have also been differences observed between novice and experienced teachers with respect to washback. We have seen that in many contexts teachers change the content of their teaching but not their methods as a result of examination changes" (1999, p.24).

It therefore remains incredibly difficult to generalize washback findings from one context to another without hesitation, as the many complex factors contributing to testing washback in either context are unlikely to overlap to a great degree, and individual differences between the teachers themselves contribute further to this dynamic.

The Korean context, as it relates to testing, and language testing in particular, is a fascinating landscape ripe with opportunities to observe washback effects. The context of the current study focused one step further, as universities in the vicinity of Seoul tend to be populated with students who have enjoyed greater academic success throughout their academic careers, which translates to specific attitudes and approaches to exams on the whole (Abelmann & Park, 2005). This is supported by the student participants' responses regarding their positive experiences with exams, as the majority of students reported good experiences with important tests (see section 5.2.1). Furthermore, the way in which students are virtually assigned to universities of certain rank dependent on their entrance examination scores almost guarantees that the student body of a university will be fairly homogeneous in many ways, allowing for unique research opportunities in which some factors that normally wouldn't, are able to be controlled for.

The students in the current study, perhaps as a result of their having grown up educated in the Korean education system, put incredible faith in the validity of examinations in general, without which the test-laden system would likely be rejected by students and parents alike. Because Koreans have accepted tests as fair and just, though, there may be unique washback phenomena occurring in the context of the current study. After years of uniform test preparation, defined for almost all students as evenings and weekends with private tutors and expensive classes at cram schools, students in this context may be at a loss when it comes to outside test preparation for a performance speaking exam such as the GMATE. Rote memorization and repetition are the methods used by almost, if not all students when preparing for important English language tests, yet these techniques offer little in the way of paths to probable success when used for a speaking test. Student participants in the current study therefore provide a fascinating, and perhaps more verifiable look at how student factors are related to context, and how these factors might influence the ways in which washback effects manifest themselves and are reported or observed.

Current theory offers only very vague caveats regarding the occurrence of washback effects in certain populations of teachers and learners, despite the vast majority of studies including some mention of the complexity and dynamic characteristic of washback influence. In the Washback Hypotheses, for example, we are presented with two conflicting statements (Alderson & Wall, 1993): “Tests will have washback on *all* learners and teachers.” And, “Tests will have washback effects for *some* teachers and some learners, but *not* for others.” At first glance this seems to be a somewhat confusing contradiction within one of the most important theories in the field of CLT, but I believe it can be further clarified when viewed through the lens of context.

Cheng, Watanabe, and Curtis present a variety of studies having taken place in a variety of contexts (2004). This collection of papers shows that washback effects can indeed occur in a variety of educational situations, but also that there may exist some underlying connections between different contexts that can, to some degree, account for trends in observed and reported washback effects. An example of this can be seen in the findings of Watanabe

(2004) and Cheng (2004). In both of these studies, teachers were found to resort to grammar-translation teaching methodologies when faced with test preparation responsibilities – a common strategy employed by English language teachers in China, Japan, and South Korea. While Watanabe's study was done in Japan, and Cheng's in Hong Kong, there are some marked similarities between these distinct regions' education cultures, which has been supported in recent investigations, as well (Cheng et al., 2014).

As covered in chapter two, South Korea's education culture and history share many ties with its neighbors, China, Taiwan, and Japan, and therefore it is to be expected in some ways that one might observe similar washback effects, regardless of the type of exam investigated. The results of this study seem to confirm this idea, as some of the teacher participants, all of whom were from outside the region, reportedly relied on test preparation strategies common in South Korea. Furthermore, with regard to student washback effects, there were some congruencies between reported questionnaire responses related to anxiety and stress in connection with the exam, while these findings were not as consistent in regions outside of East Asia.

Chapter Seven: Conclusion

This chapter provides a summary of the key findings and related conclusions drawn from analyzing and discussing the data. I then present the important implications this research has for stakeholders in this and similar contexts, and provide some recommendations for future research projects. Finally, I acknowledge the limitations of this study.

7.1 Key Findings and Conclusions

There were several important findings from this research related to how the GMATE affected teaching and learning in the research context. Through observations and interviews, teachers showed and reported that they only very occasionally varied their class activities and tasks related to the GMATE, even though a larger number of the lower-level students they were teaching would have been subjected to stricter and higher-stakes scoring guidelines. Some teachers mentioned that although they were aware of these students in their classes, those students were outnumbered by their classmates, who fell under the new lower-stakes guidelines, and thus the educators didn't feel it was appropriate to adjust their classroom activities for such a small number of students. I also found that these teachers generally relied upon the textbook for both regular and test preparation classes, with a few individual exceptions that relied more heavily upon supplements they created themselves. This did not vary much across levels taught; however, when teaching methodologies were examined more closely, there were larger differences between the highest and lowest levels in the amount of teacher-centered time spent during class periods, as well as how much time was spent on direct preparation for the exam, although again there were variations across teachers.

Teachers during the intensive term organized their classes and chose materials slightly differently when they were teaching students for whom the GMATE was a higher-stakes test. There was much more narrowing of the curriculum, as teachers reported eliminating more material to focus on GMATE preparation, particularly for those students in lower levels that were also subjected to the higher-stakes version of the test. Teachers also expressed more negative ideas related to the GMATE during the intensive period, in a few cases specifically mentioning the difference in the stakes of the exam. Finally there seemed to be a larger shift

away from homework during the intensive classes, as teachers accepted more responsibility for their students' success and failure as it related to the exam outcome. Some of the ideas and behaviors that emerged in this study were particularly interesting in that they did not reflect trends acknowledged in the literature, and I proposed that this may have occurred because of the unique contextual situation of this study in that none of the teacher participants in the study shared their L1, education background, or culture with their students.

Other interesting findings came from student questionnaire response patterns. Overall, students' responses seemed to follow expectations one might form based on knowledge of the education culture in South Korea. For example, students responded that they depended on their teachers for materials to help them prepare for English exams, and that they placed high importance on their GMATE scores, while at the same time acknowledging that it was likely of little value beyond their immediate graduation or course needs. Furthermore, the students in this study overwhelmingly responded that the GMATE caused them stress, and that they were worried that they would fail the exam. This was a curious response pattern because the majority of the students in the study were not subjected to the higher-stakes version of the exam, thus they wouldn't have failed the exam no matter what their results were. However, in the testocracy of South Korea, every exam seems to carry significant weight, and from a young age these learners are repeatedly trained to focus and prepare for every test with the utmost effort. Regarding washback effects on specific learning processes, students responded that they *did* change the ways that they studied English based on the GMATE, but that they *did not* study English harder.

Student response patterns varied with certain factors. There were correlations, for example, between level and stress from the GMATE, as although many higher-level students responded that the test caused them stress, it was to a lesser degree than for lower-level students, particularly those students in level 1 who would have had to improve much more throughout the term in order to achieve passing scores. Students in lower levels also showed more aversion towards taking the GMATE at all, while students in higher levels responded in this way less so. I had expected to see some seasonality in students' responses, perhaps seeing that they would become more stressed by the exam as the testing date approached,

but while direct test preparation in and outside of class increased greatly as the test drew near, students' reported stress and aversion to the test was maintained at high levels at the beginning, middle, and end of the term. Lastly, although most students responded that the GMATE did not motivate them to study English harder, this was correlated with year in school, so it may have been that older students, who also placed higher value on English skills in general, were more likely to receive positive washback effects in this instance.

7.2 Implications for Stakeholders

The GMATE, although not a high-stakes exam for the majority of the student participants in this study, has very real washback effects in the way it causes stress to both students and teachers, and in how it steers textbook and supplemental material creation. Furthermore, this exam influences the methodologies employed by teachers in the context, which had a direct impact on how students studied English and prepare for the exam. The students in this context did not follow some of the washback effect trends shown in washback studies conducted in other regions, though, and thus it may be useful for teacher and policy makers to keep in mind the tendencies of their students when creating future English exam policies.

This exam is unique in that it is a holistic performance test of English speaking ability, a skill deemed very desirable by Korean companies and indeed, society on the whole. I believe that the examination itself may not have been responsible for many of the negative washback effects found in this study, but instead the blame rests on the policy, and thus it is the policy itself that requires revisiting for consideration of change. Although some researchers have proposed that lowering the stakes of the exam will reduce the washback effects of a test, both positive and negative, the Korean context shows some unique attributes that may provide some resilience to that part of the Washback Hypotheses.

Therefore, I believe it is crucial that teachers maintain an awareness of the effects that exams and policies have on their teaching practices, while also keeping in mind the fact that their Korean learners of English may view exams in different ways due to sociocultural or contextual factors, and thus washback effects may occur in different ways and with varying intensity that these teachers may expect based on past experiences. Teachers should try

and bridge the power divide between students and instructors, particularly in Korean learning environments, to include and empower learners to participate in the formation of learning objectives and expectations, which can then be translated into teaching and assessment if appropriate.

Administrators should understand that the policy and testing choices they make have far-reaching consequences within their respective education context, and thus even seemingly minor shifts in testing policy must be made after sincere thought, and ideally, formal research into the possible washback effects that could occur on students and teachers. Finally, the test creators themselves should take a strong look at what it is they are trying to accomplish through each test item type. It has been shown time and time again that tests often have washback effects on the learning environment; however, we also know that washback effects can exist in both positive and negative forms. If test and materials creators wish to effect positive washback on the classroom, their work will require much more diligence and planning than seems to be involved in the test creation and administration process in its current form. In order to better realize learning outcome objectives through test creation and administration, test creators, policy makers, and teachers should come together (Tan & Turner, 2015) while also involving students to ameliorate ethical and test validity concerns, although latter efforts have been shown in similar contexts to achieve varying degrees of success (Yu, 2007).

7.3 Contributions to the Field

This thesis provided some answers to unique scenarios not thoroughly investigated prior. First, this study was conducted in a unique context as the bulk of washback studies conducted in Korea have focused on either high school teachers and students preparing for the national college admissions exam, or university-aged and older learners studying for the various major standardized English exams such as the TOEIC, TOEFL, or IELTS. In addition to investigating washback effects in this research context, though, the exam itself presented some distinctive attributes that aided in providing these findings with their own place in the literature. The GMATE is both a holistic speaking exam composed of individual tasks, and administered via computer. These characteristics make it an interesting English language

test to investigate. Finally, it is not easy to find non-experimental research opportunities in which one can observe groups of students and teachers preparing for the same examination, but with varying stakes dependent on the test-takers' year in school or major of study. This allowed me to compare teachers' and students' approaches to the exam, and draw connections with their behaviors and perceptions and they related to the version of the exam for which they prepared.

7.4 Recommendations for Further Research

This study raised some interesting new concerns that I think warrant further investigation in the future. The literature has grown not only in depth but in breadth of focus, with studies having been done in regions and contexts around the globe, and this study reinforced the important notion that further research must be done in different contexts, and with different types of exams, to continue building on this strong foundation. Washback is such an incredibly dynamic and complex phenomenon that it is imperative that researchers not discount opportunities for investigation because they seem to overlap with those ideas already put forth in the literature. Additionally, I believe that context itself deserves a stronger look from washback researchers in the future. The power context has to shape stakeholders belief systems, and thus the manifestation and intensity of washback effects should be at the forethought of all researchers as they undertake important research into washback and testing.

Researchers have been responsible in adequately referring readers to the complexity and dynamic nature of washback with regard to stakeholders' individual characteristics, but the body of research has grown to such a degree that I believe it is appropriate to further focus on *why* washback effects will occur for some teachers and learners, but not for others. I believe it is now time that we place context more centrally in the spotlight, so that stakeholders can more effectively make test and testing policy changes that work for positive washback (Qi, 2004). With a firmer understanding of context's influence on these complex phenomena, educators will more effectively find ways to accomplish their goals in specific teaching

environments, with appropriate reflection given to the unique characteristics of a region's teachers, students, and the educational culture on the whole.

Washback effects tend to focus on perceptions and behaviors, or processes, directly related to the exam being investigated. However, in the current study, there was an interesting pattern of responses regarding professionalism and professional identity that emerged from teachers' interviews. Teacher participants mentioned that they worried about how their students' success or failure on the GMATE would reflect on their evaluations and subsequent contract opportunities, as contracts were given on an annual basis in this particular program and university. Beyond the more simply defined feelings of stress in connection with their students' scores, though, many teachers in this study mentioned on several occasions that the test and testing policy itself diminished in some ways their professional identities as teachers.

When coding the interview transcription data, I noticed a trend in the ways teachers talked about the curriculum and test sometimes stifling their creativity in how they approached their courses, from the limitations placed on them regarding their required use of a department textbook, to the assigned weighting of their grading schemes. As I spoke with teacher participants throughout the data collection period, there seemed to be a very clear and shared concept regarding teacher professionalism and professional identity, and that framework was distinctly mentioned as not including standardized testing.

This notion seemed to be strongest among both teachers who were older, and teachers who had been working as teachers, in any context, for the longest time, which is very much in accordance with findings in the literature (Valli & Buese, 2007). Teachers with a stronger, more concrete conceptualization of their professional identity showed a tendency to be more strongly influenced by washback effects, and in the current study, it may have been that those teachers felt the most identity conflict with regard to their professionalism and roles as educators.

The Washback Hypotheses specify that “a test will influence attitudes to content, method, etc. of teaching/learning” (Alderson & Wall, 1993). However, very little has been written focusing on the impact of tests and testing practices on teachers’ identities as educators and professionals. I believe this warrants concentrated study in the future, as the findings of such research could help to drive teacher education and support systems within schools if tests are shown to have real effects on the ways teachers perceive themselves and their roles. This is an area that I think could yield incredible understanding into how tests and testing policies might affect teachers at different stages of their careers, and how these effects could knock on to create separate issues in teacher education and professional development in the future.

7.5 Limitations of the Study

An important factor included in major washback models is learning outcomes. Being able to discern how an exam influences not only the ways in which students and teachers study and teach, respectively, but also how it positively or negatively impacts students’ learning outcomes is an important notion to mention, and quantify, when possible. Because the overarching objectives connected with the study related to telling the participants’ stories in hopes of effecting policy change in the research context, this additional group of ideas was beyond the scope of the current study. Additionally, because this thesis concentrated on a unique population of students along with a small group of instructors, the findings of this study should not be generalized too judiciously to other contexts, despite the outward similarities a learning environment might share with the context described in this research. While I believe there are many factors contributing to great homogeneity within the student participant population in this paper, and I outline in chapter two several ways in which the learner population of South Korea is more homogeneous than in contexts outside the region, it is still crucial to keep in mind that part of the complexity of washback comes from the fact that stakeholders are incredibly unique in and of themselves, and thus washback studies undertaken in different contexts should be conducted as novel projects on relatively blank slates.

It is important to remind the reader of some of the methodological limitations, as well. Ideally, I would have been able to observe a great number of classes than were made available to me, due to resource limitations in addition to difficulties unique to the context at the time (see chapter four). Furthermore, through multiple interviews with teacher participants I was able to accurately, and in great detail, provide them a voice with which to tell their stories, while again I was limited in my ability to do so for the student participants, instead electing to rely upon questionnaire responses from a greater number of participants rather than a more detailed case study approach with fewer students. Although this provided me with excellent data from which to draw clearer conclusions regarding the overall student experience in the program, it would have been a welcome luxury to have been able to follow up with students individually to expand on and clarify their questionnaire responses.

Finally, I entered this project with very clear biases regarding the nature of the exam, and its possible effects on the students and teachers in this particular university context. This was appropriate due to the critical nature of the research, and is in fact a tenet of conducting research under the critical paradigm, but my own values regarding the GMATE, and standardized English exams in general, could very well have had an effect on how I identified themes in interview data and while conducting observations. I made a sincere effort to maintain appropriate cognizance of these biases throughout the data collection and analysis period, as well as while writing up my findings and the discussion of the same, but it is important, I think, to again mention the limitations of holding such judgments when conducting qualitative research.

Appendix A1 Student Questionnaire Korean

학생용 설문조사 초안 (한국어)

만 나이 _______ 학년 _______ 영어 강의 레벨 _______

귀하가 느끼고 있는 감정을 가장 잘 나타내고 있는 것을 숫자를 선택해서 아래의 문항을 완성하여 주십시오.

1 = 전혀 그렇지 않다 **2** =그렇지 않다 **3** = 약간 그렇지 않다 **4** = 약간 그렇다 **5** = 그렇다 **6** = 매우 그렇다

1._____영어시험을 볼 때 대체적으로 좋은 경험이 됐다.

2._____영어 시험 점수는 나의 영어 실력을 정확하게 보여준다.

3._____나는 영어 시험을 대비 할 때 학교 수업 자료 이외에도 외부 자료도 참고하여 공부한다.

4._____나는 모의 GMATE 시험을 본 적이 있다.

5._____나는 수업 시간이 아닐 때에도 GMATE 기출 문제를 풀어본다.

6._____GMATE 시험 때문에 스트레스를 많이 받는다.

7._____교수님이 GMATE 시험을 준비하는데 너무 많은 수업 시간을 할애하신다.

8._____내가 수업시간에 배운 영어 전략들이 GMATE 와 관련되어 있다.

9._____나는 GMATE 에 출제되는 문제 유형과 시험 형태를 잘 알고 있다.

10.____ 교수님이 GMATE 를 꼼꼼하고 정확한 방법으로 준비할 수 있도록 지도해주셨다.

11.____ GMATE 시험을 보는 것이 왜 필요한지 잘 이해하고 있다.

12.____나는 수업 시간에 GMATE 와 직접적으로 관련된 활동을 한다.

13.____GMATE 점수는 훗날 내가 취업을 하는데 도움이 될 것이다.

14.____수업에서 GMATE 에 관련된 내용을 다루는 것은 다른 어떤 내용보다도 중요하다.

15.____GMATE 시험에 불합격 할까봐 두렵다.

16.____수업 교재가 GMATE 를 준비하는데 도움이 된다.

17.____내 영어 실력을 평가하는데 GMATE 만큼 공정한 시험이 없다고 생각한다.

18.____교수님이 GMATE 시험 관련 자료를 주신다.

19.____GMATE 시험을 준비할 때 평상시에 내가 영어를 학습하는 방법과 다르게 공부 한다.

20.____수업 교재는 GMATE 를 준비하는데 적절한 교재이다.

21.____GMATE 를 보지 않았으면 좋겠다.

22.____GMATE 로 인해 영어 공부를 더 열심히 한다.

23.____내가 영어를 말할 수 있다는 것은 중요하다.

24.____전교생이 GMATE 에 합격해야만 한다고 생각한다.

Appendix A2 Student Questionnaire English (translated)

Student Questionnaire (English)

Age (Western) ________ Year in school _________ Class Level _________

Please complete the following by selecting the number that best describes your thoughts/feelings:

1 = Strongly Disagree 2 = Disagree 3 = Somewhat Disagree 4 = Somewhat Agree 5 = Agree 6 = Strongly Agree

1._____I have an overall positive experience taking English tests
2._____English test scores are accurate representations of my English ability.
3._____I prepare for English tests using materials separate from my classwork.
4._____I have taken practice GMATE tests
5._____I study directly for the items on the GMATE outside of the classroom
6._____The GMATE causes me stress
7._____My instructor spends too much time preparing students for the GMATE
8._____The English skills I have learned are related to the GMATE
9._____I am aware of the format and items that appear on the GMATE
10.____My instructor prepared me adequately for the GMATE
11.____I understand why I must take the GMATE
12.____I do activities in class that are directly related to the GMATE
13.____The GMATE will help me with my future career.
14.____The GMATE is more important to me than other parts of my English class
15.____I am afraid of failing the GMATE
16.____The textbook helps prepare students for the GMATE
17.____The GMATE is a fair test of my English ability
18.____My instructor provides us with extra materials for the GMATE
19.____I change the way I study English because of the GMATE.
20.____The textbook is appropriate for the requirements of the GMATE.
21.____I would prefer not to take the GMATE.
22.____The GMATE motivates me to study English harder.
23.____It is important for me to be able to speak English.
24.____All students at my university should be required to pass the GMATE.

Appendix B Teacher Questionnaire

Age __________ Years Teaching EFL at University Level _________

Years Teaching at Current University __________

Highest Degree Obtained _____________________________ (BA/BS, MEd/MA, EdD/PhD)

Current GELT Level Courses ______________ (1, 2, or 3 – write as many as apply)

Are you a native English speaker? __________________ (YES/NO)

Please complete the following by selecting the number that best describes your thoughts/feelings:

1 = Strongly Disagree 2 = Disagree 3 = Somewhat Disagree
4 = Somewhat Agree 5 = Agree 6 = Strongly Agree

1. ____ The GMATE motivate me to implement activities to promote my students' test-taking skills.
2. ____ My time allotment in class would be different if the GMATE were cancelled.
3. ____ I arrange my classroom activities to meet the requirements for the GMATE.
4. ____ I spend more time instructing grammar other than communication skills because I think grammar is more likely to be tested on the GMATE.
5. ____ I rarely use specific teaching activities to promote my students' language skills just for the GMATE.
6. ____ I teach test-taking strategies, especially as the GMATE testing date gets closer.
7. ____ I arrange my classroom activities based upon different factors but not just based upon the objectives of the GMATE.
8. ____ I change my teaching methods to help students to succeed on the GMATE.
9. ____ I would select teaching methods to help develop my students' skills that are more likely to be tested on the GMATE.
10. ____ I neglect some teaching methods that are not able to prepare my students for the GMATE.
11. ____ The GMATE has little impact on how I teach.
12. ____ I rarely change my teaching methods to help my students succeed on the GMATE.
13. ____ I usually use the textbooks because they cover the topics on the GMATE.
14. ____ I use materials not the textbooks if they will help my students succeed on the GMATE.

Appendix B (cont.)

15. ____ I give students materials to review content expected to be on the GMATE.
16. ____ I have my students do mock tests to familiarize students with the GMATE.
17. ____ The GMATE influences which supplementary materials I use.
18. ____ The GMATE affects my syllabus, including practicing the kind of items that are to be tested.
19. ____ I pay little attention to the GMATE while constructing my teaching syllabus.
20. ____ The GMATE influences my decision about which language skill is more important to be taught.
21. ____ I emphasize the skills which are more likely to be tested on the GMATE while planning for my syllabus.
22. ____ I skip over certain sections in the textbook because they are less likely to be tested on the GMATE.
23. ____ I adjust the sequence of my teaching objectives based on the GMATE.
24. ____ I focus on certain sections in the textbook because they are more likely to be tested on the GMATE.
25. ____ I include some relevant content to help my students perform well on the GMATE.
26. ____ My course content is established to reflect the objectives of the GMATE.
27. ____ The GMATE has little impact on what I teach.
28. ____ I cover every section in the textbook although some sections are unlikely to be tested on the GMATE.
29. ____ I include different technique to evaluate my students.
30. ____ I evaluate my students' works by using the criteria used by raters when rating the GMATE.
31. ____ I adapt test items from the mock tests in my classroom quizzes.
32. ____ I include listening tests in my classroom quizzes.
33. ____ I include speaking tests in my classroom quizzes.
34. ____ My assessment has been changed for the GMATE.
35. ____ I would teach whatever I think is important to teach no matter whether my students like it or not.
36. ____ I spend less time on certain sections of the textbooks because my students are less interested in them.

Appendix B (cont.)

37. ____ I often teach what will be tested because my students expect me to do so.
38. ____ I expect my students to perform well on the GMATE.
39. ____ Students' learning attitudes influence my teaching.
40. ____ I will feel embarrassed if my students perform less well on the GMATE than other students taught by my colleagues.
41. ____ I will feel guilty if my students do not succeed on the GMATE.
42. ____ Students' test results influence how people judge me as a good teacher.
43. ____ The GMATE gives me important feedback about how I teach.
44. ____ Improving students' test scores is stressful to me because my school administrators often compare my students' test score results with others.
45. ____ I feel pressure from my administration to improve my students' test scores.
46. ____ I feel pressure from other teachers to improve my students' test scores.
47. ____ I feel pressure from students to improve their test scores.
48. ____ I am aware of the changes of the GMATE.
49. ____ I teach to the test especially when I am aware of the test format on the GMATE.
50. ____ I have opportunities gaining information about the GMATE objectives.
51. ____ I regard the GMATE as a test, which will influence students' future career.
52. ____ Students' GMATE score result will have significant sanctions to most teachers.
53. ____ The GMATE is a fair test to students.
54. ____ The GMATE is able to test my students' language ability.
55. ____ Students' GMATE score result will have significant rewards to most teachers.
56. ____ The GMATE will influence students' admission to the secondary schools.
57. ____ The GMATE provides little feedback about students' learning.
58. ____ The GMATE impacts curriculum innovation.
59. ____ The GMATE has changed my school's language teaching policy.
60. ____ My curricular planning and instruction are influenced by teaching time.
61. ____ My school's policies influences my curricular planning and instruction.
62. ____ My personality influences my curricular planning and instruction.

Appendix C1 Interview Guide First Meeting Example

General Profile

How long have you been teaching in the program?
Have your responsibilities changed during that time? How?
Has your approach to the needs of the students changed during that time? How?
How would you describe yourself as a teacher?

Teaching

Do you have any preferences for materials or activities in your speaking classes?
How much control do you feel you have on what and how you teach?
How much time do you estimate you spend on preparing for the GMATE each week?
How do you decide what and how to teach each week? Does this change? How?

GMATE Perceptions

You answered that you believe the GMATE is/isn't a fairly important part of the curriculum.
Could you please explain in what ways you believe so?
What are your overall thoughts on the GMATE?
Do you think those feelings are shared by your colleagues?
How would you like to change the GMATE policy?

Content and Test Prep

What are you planning to do for the next couple weeks related to GMATE prep?
How would you like to change the textbooks?
What do you believe is the most effective way for students to improve their English speaking ability?
What do you think is the best way for students to prepare for the GMATE?
How do you approach the GMATE for different ages or levels of students?

Is there anything else you'd like to add that we haven't covered today?

Appendix C2 Interview Guide Subsequent Meeting Example

What have you done in class during the past two weeks related to GMATE preparation?

Is this different in any way from what you've done in past terms?

Have you assigned specific homework related to the GMATE? Why? Why not?

How has your classroom approach changed as the GMATE draws more near?

Do your students seem concerned about the GMATE?

Are you concerned about your students' ability to perform well on the GMATE?

How do you communicate with your students about the GMATE during classes?

Observations (if applicable)

For each activity:

You chose to ________________ in your class on ___________ . How did you feel that went?

Did you choose that activity for any particular reason?

Is this an activity you rely upon regularly, or is this a new activity to you?

It seemed to relate directly/indirectly to GMATE item(s) __________.

Would you agree?

Was this done with purpose?

(If supplemented)

Did you design the supplement yourself?

Where did you find it, if not?

Do you share supplements with your colleagues?

Do they share supplements with you?

Appendix C3 Coding Schemes

Preliminary Coding Scheme	Final Coding Scheme
MAT (materials)	TXT (textbook)
	SUPP (supplements)
WHT (what was taught)	CONT (content)
	TST (test prep)
	TSK (tasks/activities)
LVL (students' level)	LVL
STKES (stakes of exam)	STKES
GMATE	GPREP (GMATE preparation)
	GVAL (GMATE value)
	GFEEL (feelings about the GMATE)
TME (teaching/prep time)	TME
HW (homework)	HW
FEEL (feelings)	ID (identity)
	PRO (professionalism)
	EVAL (teacher evaluations)
	JOB (job security/satisfaction)
HOW (how teachers taught)	ORG (classroom organization)
	TLK (who talks)
SEA (seasonality)	SEA

Appendix C4 Interview Coding Example

R: Have you assigned specific homework related to the GMATE?
C: A bit, but just the usual speaking journals that everyone does. (HW)
R: What are speaking journals?
C: I think probably just about everyone does them. Some people do them online, but I just have my students email me them. They record themselves doing some practice GMATE questions. Like a real test, kind of. (ORG/HW/TST)
R: How often do you ask students to do them?
C: Well, I only assign four, because I can't listen to a hundred recordings every week. It would drive me crazy [laugh]. I think other people do them weekly, but four is enough for me. (HW)
R: Why would it drive you crazy?
C: I mean, listening to tons of the same recordings again and again each week is hard enough, but students expect good feedback, too. It takes forever. (TME/JOB/EVAL)
R: How do you give them feedback?
C: I have a small sheet I give them with checkmarks on certain things for each task. I used to write everything, but this saves me a ton of time. (TME/SUPP) I saw I was writing the same thing again and again, and it ate up my nights and weekends. (TME/JOB) I don't know if students even read what I wrote.
R: Why do you think they may not read your feedback?
C: You're a teacher. You know. You tell a student about a mistake she's making, and then you see the same mistake in the next paper. The students who *do* read it definitely do better, I think. But that's not all of them. (FEEL?)
R: Do you give students other feedback?
C: If they want it. We have office hours, but students don't really come to talk with us. Probably a couple every term, at most. (TME)
R: Let's go back to the speaking journals. Who came up with that idea?
C: [Laughs] I don't know. Everyone's been doing them for years, I think. Since I came here, anyway. Probably someone like (removed). He's been here longer than almost anyone. (HW/SUPP)
R: Are you planning any other assignments in the next couple weeks related to the GMATE?
C: In class, you mean?
R: No. As homework.
C: Probably not. I do practice in class, so they'll get plenty of prep. It's easier to walk around and listen and give them advice in class. I think they like that, too. At least they write it on evals at the end of the term. (HW/EVAL/GPREP/TST/ORG/TLK)
R: What kinds of things do they write?
C: Mostly about feedback. That's the most common positive thing. I don't think the Korean professors do a whole lot of it. (ID/PRO/EVAL)
R: Any negative things you feel comfortable sharing?
C: Some things about the textbook. Some things about too much homework. I don't give a lot of homework, but some students think it's a lot. (EVAL/HW/TXT)

Appendix D1 Observation Scheme (completed)

Appendix D1 Observation Scheme

Teacher: D | Date: 10/27 | Visit No. 3 | Time: 10³⁰ – 1:45

Time	Activities/Episodes	Organization: Teacher T+Si	T+Sg	T+Sc	Student S+S2	S+Sg	Activity: Discussion Direct	Discussion Indirect	Presentation Direct	Presentation Indirect	Test Prep Direct	Test Prep Indirect	Materials: Textbook	Supplement
12'	Icebreakers – Q+A – related to discussion?				✓		(1) 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	✓	
11'	Pic description (lec)			✓			1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 (2) 3 4	1 2 3 4	✓	✓
12'	GMATE Pic Desc. practice				✓		1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 (2) 3 4	1 2 3 4	✓	
28'	Group Discussions (4-5 s)		✓ *				1 2 3 4	1 2 3 (4)	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	✓	
3'	HW explanation / + review			✓			1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4		

Handwritten notes:

* – More GMATE focus today than usual. Visit 4/5 → request GMATE 3/4 practice / lecture.

why still doing this kind of activity?

(I) – develop w/ test Q's? – why so long?

(I) lecture often?

(I) – always in pairs?

* T rotating groups

(I) why?

(I) – each class? – how often? – review too?

(I) when do you do presentation work?

(I) lots of textbook use → "always / often?"

(I) where? how chose? pics on projector

(I) why focus on Question 2?

178

Appendix D2 Observation Scheme (blank)

Teacher ______ Date ______ Visit No. ______ Time ______

Time	Activities/Episodes	Organization					Activity												Materials	
		Teacher			Student		Discussion				Presentation				Test Prep					
		T+Si	T+Sg	T+Sc	S+S2	S+Sg	Direct		Indirect		Direct		Indirect		Direct		Indirect		Textbook	Supplement
							1	2	1	2	1	2	1	2	1	2	1	2		
							3	4	3	4	3	4	3	4	3	4	3	4		
							1	2	1	2	1	2	1	2	1	2	1	2		
							3	4	3	4	3	4	3	4	3	4	3	4		
							1	2	1	2	1	2	1	2	1	2	1	2		
							3	4	3	4	3	4	3	4	3	4	3	4		
							1	2	1	2	1	2	1	2	1	2	1	2		
							3	4	3	4	3	4	3	4	3	4	3	4		
							1	2	1	2	1	2	1	2	1	2	1	2		
							3	4	3	4	3	4	3	4	3	4	3	4		

Appendix D3 Observation Scheme Item Definitions

Organization refers to the way the classroom activity was organized with respect to interactions. There are several subsections here, including teacher to individual student interactions (T+Si), teacher to small group of students (T+Sg), teacher to entire class (T+Sc), student to student in pairs (S+S2), and student to student in groups of more than two (S+Sg).

Activity refers to the type of tasks that occurred in the classroom. Subsections here include discussions (D), presentations (P), and test preparation (T), with each subsection further being broken down into being indirectly or directly related to the GMATE. If an activity was related to the GMATE, the tasks to which it related were noted in the coding sheet (1 relating to personal information, 2 having to do with the picture description, 3 connecting to process description, and 4 signifying opinion and/or advantages and disadvantages of something). For example, a discussion activity that indirectly related to tasks 3 and 4 on the GMATE was noted as “Di34” in my notes and appropriately ticked in the coding scheme, while a test preparation activity that directly related to the picture description task on the GMATE, such as practice pictures, was noted “Td2,” while I also marked the corresponding boxes on the observation scheme.

Materials refers to the type of learning materials used during a task. After initial observations, it became clear that there was far too wide a range of materials being used in classes, so I chose to divide this into only two subsections (textbook and supplement) while taking brief notes on the type of supplement used by teachers.

Appendix E1 Bilingual Consent Form

Washback Effects of a Korean University English Exit Exam

CONSENT FORM 동의서

I have been fully informed about the aims and purposes of the project.
본 당사자는 이 연구의 목표와 목적에 대해 충분히 안내 받았습니다.

I understand that:

there is no compulsion for me to participate in this research project and, if I do choose to participate, I may at any stage withdraw my participation
연구에 반드시 참여할 강제성은 없으며, 연구 도중에 참여를 원하지 않으면 언제라도 그만 둘 수 있습니다.

I have the right to refuse permission for the publication of any information about me
개인 정보를 공개하는 것에 대해 거절할 수 있는 권리가 있습니다.

any information which I give will be used solely for the purposes of this research project, which may include publications
제공된 정보는 오직 연구나 출판물의 발행의 목적을 위해서만 사용될 것입니다.

if applicable, the information, which I give may be shared between any of the other researcher(s) participating in this project in an anonymised form
이 연구에 익명으로 참여한 다른 연구자들이 제공된 정보를 공유할 수도 있습니다.

all information I give will be treated as confidential
제공된 정보는 기밀에 유지될 것입니다.

the researcher(s) will make every effort to preserve my anonymity
연구자들은 본인의 익명을 보호하기 위해 모든 노력을 다 할 것입니다.

....

(Signature of participant) (Date)
참가자 사인 날짜

....

(Printed name of participant)
참가자의 인쇄된 이름

Appendix E1 (cont.)

One copy of this form will be kept by the participant; a second copy will be kept by the researcher(s)
이 동의서는 참가자와 연구자가 각 한 부씩을 보관할 것 입니다.

If you have any concerns about the project that you would like to discuss, please contact:
이 연구에 있어서 궁금한 것이 있으면 연락 주십시오.

Jason A. Di Gennaro
Email: jason.digennaro@gmail.com
Contact phone number of researcher(s): 010-7298-3542
연구자의 연락처

OR

Dr. Li Li (Thesis Supervisor)
Email: li.li@exeter.ac.uk

Data Protection Act: The University of Exeter is a data collector and is registered with the Office of the Data Protection Commissioner as required to do under the Data Protection Act 1998.

정보 보호 조항: 1998 년에 생긴 정보 보호법에 따라 Exeter 대학교는 자료를 수집하고 the Office of the Data Protection Commisioner 에 등록하였습니다.

The information you provide will be used for research purposes and will be processed in accordance with the University's registration and current data protection legislation.

귀하가 제공하신 자료는 연구 목적으로만 사용 될 것이며 본 대학의 규정과 정보 보호법에 따라 처리될 것입니다.

Data will be confidential to the researcher(s) and will not be disclosed to any unauthorised third parties without further agreement by the participant.

자료는 연구자에게만 비밀로 유지되며 연구 참여자가 동의 하지 않은 제 3 자에게 공개되지 않을 것 입니다.

Reports based on the data will be in anonymised form.

논문에 나온 자료는 익명으로 기록될 것 입니다.

Any data that can identify you individually (e.g., video, voice, etc.) will be securely stored and destroyed within a period of 5 years after the completion of the research project.

참가자 개인을 특정 짓는 자료(예: 비디오, 목소리 등)는 논문 프로젝트가 완성된 후 5 년 동안 안전하게 저장되고 그 후에 파기될 것입니다.

Appendix E2 Teacher Information Sheet

Information Sheet
Washback Effects of a Korean University Exit Exam

You are being invited to participate in a research project which aims to investigate the effects that the GMATE assessment has on teachers' and students' teaching and learning practices, respectively. This research will be used to fulfil, in part, the requirements for the doctoral program at the University of Exeter, in which I am enrolled.

Before signing the accompanying consent form, please take a moment to read through the following information, which will give you a better understanding of the research questions and objectives of this project.

This project proposes three questions:

1) What are the perceived washback effects of the GMATE on teachers' teaching?
2) What are the perceived washback effects of the GMATE on students' learning?
3) What are the observed washback effects of the GMATE on classroom environments?

To answer these questions, I will be employing several different research methods, in order to maximize accuracy of the data and my subsequent interpretation and analysis. First, I will be using questionnaires for both students and teachers, in order to start building a rough understanding of your experiences and feelings regarding this particular exam. Next, I will be using a combination of semi-structured interviews and classroom observations to help build on your questionnaire answers. As teacher participants, I will be conducting scheduled classroom observations and interviews with you, at times that are most convenient for you, the participant.

Your personal information and data will be kept secure at all times, and will be destroyed within five years after the study is concluded. Please see the accompanying consent form for more specific information regarding anonymisation of your data and the Data Protection Act.

You are under no obligation to participate in this research, and will be at no disadvantage should you choose to abstain from participating. If you do choose to participate, however, please read the consent form accompanying this information sheet, and sign and date where appropriate. Your participation will greatly help me to complete my studies, and will further provide important data that may help create and conduct tests and testing policies in the future.

Thank you very much for your consideration and participation. If you have further questions regarding any component of this research project, please don't hesitate to contact me at either my personal email address: jason.digennaro@gmail.com or my University of Exeter email address: jad216@exeter.ac.uk.

Jason A. Di Gennaro

Appendix E3 Student Information Sheet (Korean)

한국 대학교 졸업 시험의
역류효과(Washback Effects)에 관한 연구 안내문

본 연구자는 University of Exeter 에서 박사 과정의 일환으로 GMATE 평가가 교수학습에 미치는 영향력에 관한 연구를 진행하고 있습니다.

다음의 내용은 이 연구의 문제와 목적을 제시하고 있습니다. 천천히 읽어보시고 동의서에 사인을 해주시면 됩니다.

연구 문제 제기

1) 인지하고 있는 GMATE의 역류효과가 교사의 교수에 어떤 영향을 주는가?

2) 인지하고 있는 GMATE의 역류효과가 학생의 학습에 어떤 영향을 주는가?

3) 관찰된 GMATE의 역류효과가 학습 환경에 어떤 영향을 주는가?

본 연구자가 위 문제에 답을 찾고 자료의 분석을 최대한 정확하게 하고자 몇 가지 다른 연구 방법을 택할 것 입니다. 첫 째, 저는 교사와 학생들의 이해를 돕기 위해서 GMATE 에 관한 내용을 설문지로 조사할 것 입니다. 다음으로 설문 조사의 결과를 바탕으로 반 구조화 면담과 수업 관찰을 할 예정입니다.

참여자의 개인 정보는 보호될 것이며 연구가 종료된 후, 5 년 뒤에 폐기될 것입니다. 귀하의 정보에 대한 익명성과 정보 보호법에 관한 더 자세한 정보를 원하시면 함께 첨부되어 있는 동의서에서 확인 가능 합니다.

귀하가 이 연구에 반드시 참여할 이유는 없으며, 참여하시는 것에 있어서 어떤 피해가 가지 않음을 약속 드립니다. 이 연구에 참여하기로 결정하셨다면, 본 안내문과 함께 첨부되어 있는 동의서를 꼭 읽어주시고 귀하의 사인과 동의하신 날짜를 적어주시기 바랍니다. 훗날 이 연구가 시험을 설계하고 정책을 도입하는데 많은 도움이 되었으면 좋겠습니다.

여러분의 협조에 진심으로 감사 드립니다. 만약 이 연구에 있어 더 궁금한 점이 있으시면 본 연구자의 이메일로(jason.digennaro@gmail.com 이나 jad216@exeter.ac.uk) 보내주시면 감사하겠습니다.

Jason A. Di Gennaro

Appendix E4 Student Information Sheet (English)

Information Sheet
Washback Effects of a Korean University Exit Exam

You are being invited to participate in a research project which aims to investigate the effects that the GMATE assessment has on teachers' and students' teaching and learning practices, respectively. This research will be used to fulfil, in part, the requirements for the doctoral program at the University of Exeter, in which I am enrolled.

Before signing the accompanying consent form, please take a moment to read through the following information, which will give you a better understanding of the research questions and objectives of this project.

This project proposes three questions:

1) What are the perceived washback effects of the GMATE on teachers' teaching?
2) What are the perceived washback effects of the GMATE on students' learning?
3) What are the observed washback effects of the GMATE on classroom environments?

To answer these questions, I will be employing several different research methods, in order to maximize accuracy of the data and my subsequent interpretation and analysis. First, I will be using questionnaires for both students and teachers, in order to start building a rough understanding of your experiences and feelings regarding this particular exam. Next, I will be using a combination of semi-structured interviews and classroom observations to help build on your questionnaire answers. As student participants, I will be conducting scheduled classroom observations and will ask you to fill out electronic questionnaires at different times during the term.

Your personal information and data will be kept secure at all times, and will be destroyed within five years after the study is concluded. Please see the accompanying consent form for more specific information regarding anonymisation of your data and the Data Protection Act.

You are under no obligation to participate in this research, and will be at no disadvantage should you choose to abstain from participating. If you do choose to participate, however, please read the consent form accompanying this information sheet, and sign and date where appropriate. Your participation will greatly help me to complete my studies, and will further provide important data that may help create and conduct tests and testing policies in the future.

Thank you very much for your consideration and participation. If you have further questions regarding any component of this research project, please don't hesitate to contact me at either my personal email address: jason.digennaro@gmail.com or my University of Exeter email address: jad216@exeter.ac.uk.

Jason A. Di Gennaro

Appendix F1 GMATE Item Examples (from Bishop & Carter, 2015, p. 109)

GMATE task 1

This question requires only a short reply because there are only 10 seconds to say your answer.

How many members are there in your family?
Where is your hometown?
What is your favorite kind of music?
Who is your best friend?
What are your hobbies?

GMATE task 2

This question has 15 seconds for thinking and 60 seconds for speaking. Your professor will provide you with some appropriate pictures to practice describing

GMATE task 3

This question has 20 seconds for thinking and 70 seconds for speaking.

Please describe in detail how to make a cup of tea.
Please describe in detail how to choose a gift for a friend.
Please describe in detail how to order a cup of coffee at a café.
Please describe in detail how to stay healthy in winter.

GMATE task 4

This question has 20 seconds for thinking and 80 seconds for speaking.

What are the advantages and disadvantages of wearing glasses?
What are the advantages and disadvantages of owning a car?
What are the advantages and disadvantages of going to university?
What are the advantages and disadvantages of being a vegetarian?
What are the advantages and disadvantages of wearing makeup?

Appendix F2 GMATE Rating Level Descriptions (from www.mate.or.kr)

Moderate High (MH)

직접적인 질문이나 개인 정보 요구에 응답할 수 있다. 길 안내, 물건 구입 등과 같은 일상적 상황에 필요한 질문을 할 수 있다. 문장을 좀 더 길게 구사할 수 있고, 문장을 연결하여 말할 수 있다. 어색한 표현이 나타나기도 하지만 어휘가 다양해진다.

Can respond to direct questions or requests for personal information. Can ask questions for everyday situations such as getting directions, buying things, etc. Can use longer sentences, and you can connect sentences. There are awkward expressions, but vocabulary is diverse.

Moderate Mid (MM)

일상적인 주제에 대하여 문장을 나열하거나 질문할 수 있다. 자신에게 익숙한 주제에 대한 의사 표현이 가능하다. 문장이 길어지기 시작하고, 어느 정도 문장을 연결할 수 있지만, 논리적이기보다는 단편적인 문장에 주로 의존한다.

Can list sentences or ask questions about everyday subjects. Can express oneself on a subject speaker is familiar with. Sentences begin to lengthen and can connect to some extent, but speaker relies heavily on the fragmented sentences rather than the logical ones.

Moderate Low (ML)

일상 생활에 필요한 의례적인 대화가 가능하다. 대화에 수동적으로 임하고, 단순한 문장을 구성하며 신상에 관한 직접적인 질문에 어렵게나마 의사 소통을 할 수 있다. 몇 가지 암기된 질문을 할 수 있다.

Possible to have ritual conversations necessary for everyday life. Can work passively in conversations, organize simple sentences, and communicate with difficulty to direct questions about one's identity. Can ask memorized questions.

Rudimentary (R)

단어를 나열하거나 암기한 문구로 매우 어렵게 의사를 표현할 수 있다. 직접적인 질문에 2-3 개의 단어나 관용어로 응답한다. 단어를 찾기 위해 자주 머뭇거리거나 학습된 표현들을 반복하는 경향이 있다.

Can express an opinion with difficulty using single words or memorized phrases. Answers direct questions with 2-3 words or phrases. There is a tendency to hesitate frequently to find words or to repeat learned expressions.

Bibliography

Abelmann, N., Kim, H., & Park, S. J. (2005). The Uneven Burden of Vitality: College Rank, Neoliberalism, and South Korea's "New Generation." In Young-Key, K. R., Grinker, R. R., & Larsen, K. W. (Eds.), *The Sigur Center Asia Papers: Korean Education,* 24, 33-52.

Akiyama, T. (2004*). Introducing EFL speaking tests into a Japanese high school entrance examination*. Unpublished PhD thesis. University of Melbourne.

Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing, 13*(3), 280–297.

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14,* 115-129.

Ambrosie, F., & Haley, P. W. (1988). The Changing School Climate and Teacher Professionalization. *NASSP Bulletin, 72*(504), 82-89.

Anderson, G., & Anderson, G. J. (1998). *Fundamentals of educational research*. Psychology Press.

Andrews, S. (1994). The washback effect of examinations: Its impact upon curriculum innovation in English language teaching. *Curriculum Forum, 1,* 44-58.

Andrews, S., & Fullilove, J. (1994). Assessing spoken English in public examinations- why and how? In J. Boyle & P. Falvey (Eds.), *English language testing in Hong Kong* (pp. 57-86). Hong Kong: Chinese University Press.

Andrews, S., Fullilove, J., & Wong, Y. (2002). Targeting washback – a case-study. *System(30)*, 207-223.

Au, W. (2007). High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis. *Educational Researcher, 36*(5), 258-267.

Auerbach, C., & Silverstein, L. B. (2003). *Qualitative data: An introduction to coding and analysis*. NYU press.

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Bachman, L. F. (2004). *Statistical analyses for language testing.* Cambridge: Cambridge University Press.

Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly,* 2, 1-34.

Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language testing*, *13*(3), 257-279.

Bailey, K. M. (1999). *Washback in language testing*. Princeton, NJ: Educational Testing Service.

Bishop, T., & Carter, K. (2015). *Express Yourself through Discussion and Presentation.* Kyung Moon Sa: Seoul.

Boeije, H. (2002). A Purposeful Approach to the Constant Comparative Method in the Analysis of Qualitative Interviews. *Quality and Quantity*, 36, 391-409.

Bogdan, R.C., & Biklin S.K. (1998). *Qualitative research for education: An introduction to theory and methods.* (3rd ed.) Boston: Allyn and Bacon

Brannen, J. (1992). *Mixing Methods: Qualitative and. Quantitative Research.* Aldershot: Avebury.

Brown, J. D. (1997). Designing surveys for language programs. In Nunan, D., and Griffee, D., (Eds.). *Classroom Teachers and Classroom Research (55-70).* Tokyo: Japan Association for Language Teaching.

Brown, J. D. (2001). *Using surveys in language programs*. Cambridge University Press.

Burchell, B., & Marsh, C. (1992). The effect of questionnaire length on survey response. *Quality & Quantity,* 26*,* 233-244.

Burns, R. B. (1997). *Introduction to research methods.* Addison Wesley Longman.

Burrows, C. (2004). Washback in Classroom-Based Assessment: A Study of the Washback Effect in the Australian Adult Migrant English Program. In Cheng, L., Watanabe, Y., & Curtis, A., *Washback in Language testing: Research Contexts and Methods*, (p.113-128). London: Lawrence Erlbaum Associates.

Carlsen, C. (2009). Crossing the bridge from the other side: the impact of society on testing. In L. Taylor & C. J. Weir (Eds.), *Studies in Language Testing 31: Language Testing Matters (344-356).* Cambridge: Cambridge University Press.

Chang, S. J. (2008). A Cultural and Philosophical Perspective on Korea's Education Reform: A Critical Way to Maintain Korea's Economic Momentum. *KEI Academic Paper Series, 3*(2).

Chen, L. M. (2002). Washback of a public exam on English teaching, Unpublished Ph.D. thesis. The Ohio State University

Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education, 11*(1), 38-54.

Cheng, L. (1998). Impact of a public English examination change on students' perceptions and attitudes toward their English learning. *Studies in Educational Evaluation, 24*(3), 279–301.

Cheng, L. (1999). Changing assessment: washback on teacher perceptions and actions. *Teaching and Teacher Education, 15,* 253-271.

Cheng, L. (2004). The Washback Effect of a Public Examination Change on Teachers' Perceptions Towards Their Classroom Teaching. In Cheng, L., Watanabe, Y., & Curtis, A., *Washback in Language testing: Research Contexts and Methods*, (p.147-170). London: Lawrence Erlbaum Associates.

Cheng, L. (2008). Washback, impact and consequences. In *Encyclopedia of language and education* (pp. 2479-2494). Springer US.

Cheng, L., Andrews, S., & Yu, Y. (2011). Impact and consequences of school-based assessment (SBA): Students' and parents' views of SBA in Hong Kong. *Language Testing, 28*(2), 221–249.

Cheng, L., & Curtis, A. (2004). Washback or Backwash: A Review of the Impact of Testing on Teaching and Learning. In Cheng, L., Watanabe, Y., & Curtis, A., *Washback in language testing: Research Methods and Contexts*. London: Lawrence Erlbaum Associates, Publishers.

Cheng, L., Klinger, D., Fox, J., Doe, C., Jin, Y., & Wu, J. (2014). Motivation and test anxiety in test performance across three testing contexts: The CAEL, CET, and GEPT. *TESOL Quarterly*, *48*(2), 300-330.

Cheng, L., Watanabe, Y., & Curtis, A. (2004). *Washback in language testing: Research Methods and Contexts*. London: Lawrence Erlbaum Associates, Publishers.

Cho, B. K. (2007). *Problems of High-Stakes Testing System: Students' Attitudes toward the English Section of the Korean University Aptitude Examination.* Saarbrucken, Germany: VDM Verlag Dr. Muller.

Choe, Y. H. (1987). *The Civil Examinations and the Social Structure in the Yi Dynasty Korea, 1392-1600.* Seoul: Korean Research Society.

Choi, I. (1999). Test Fairness and Validity of the TEPS. *Language Research, 35*(4), 571-603.

Choi, I. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing (25)*1, 39-62.

Chu, H., & Yeh, H. (2017). English Benchmark Policy for Graduation in Taiwan's Higher Education: Investigation and Reflection*. Journal of Language Teaching and Research, 8*(6), 1063-1072.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education*. Routledge.

Creswell, J. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, 3rd ed.* London: Sage.

Crotty, M. (2010). *The Foundations of Social Research: Meaning and Perspective in the Research Process.* Los Angeles: Sage.

Daly, N. P. (2011). English Is Not Math: The Negative Washback of High Stakes Testing on English Curricular Reforms in Taiwan. *English Teaching & Learning*, *35*(2), 47-89.

Davies, A. (1968). *Language testing symposium: A psycholinguistic approach.* Oxford: Oxford University Press.

Denzin, N. K., & Lincoln, Y. S. (2000). Introduction: The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 1-28). Thousand Oaks, CA: Sage.

Dewey, J. (1929). *Experience and Nature.* La Salle: Open Court.

Dey, I. (2003). *Qualitative data analysis: A user friendly guide for social scientists*. Routledge.

Ducasse, A. M., & Brown, A. (2009). Assessing paired orals: Raters' orientation to interaction. *Language testing*, *26*(3), 423-443.

East, M. (2015). Coming to terms with innovative high-stakes assessment practice: Teachers' viewpoints on assessment reform. *Language Testing*, *32*(1), 101-120.

Eckstein, M. A., & Noah, H. J. (1993). *Secondary school examinations: International perspectives on policies and practice*. Yale University Press.

Ernest, P. (1994). An introduction to research methodology and paradigms. School of Education, University of Exeter.

Fairclough, N. (1989). *Language and Power.* London: Longman.

Fox, J., & Cheng, L. (2016). Walk a mile in my shoes: stakeholder accounts of testing experience with a computer-administered test. *TESL Canada Journal*, *32*, 65-86.

Freeman, D. (1996). Redefining the relationship between research and what teachers know. In K. M. Bailey & D. Nunan (Eds.), *Voices from the language classroom: Qualitative research in second language education,* (pp. 88-115). Cambridge: Cambridge University Press.

Freire, P. (1993). *Pedagogy of the oppressed.* New York: Continuum International.

Fulcher, G. (2009). Test use and political philosophy. *Annual Review of Applied Linguistics*, *29*, 3-20.

Gable, G. G. (1994). Integrating case study and survey research methods: an example in information systems. *European journal of information systems*, *3*(2), 112-126.

Gall, M.D., Borg, W.R., & Gall, J.P. (2006). *Educational Research: An Introduction, 8th ed.* London: Pearson.

Garland, R. (1991). The mid-point on a rating scale: Is it desirable. *Marketing bulletin*, *2*(1), 66-70.

Glaser, B. (1965). The Constant Comparative Method of Qualitative Analysis. *Social Problems*,*12*(4), 436-445.

Green, A. (2007). *IELTS washback in context: preparation for academic writing in higher education.* Cambridge: Cambridge University Press.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of qualitative research*, *2*(163-194), 105.

Herman, J. L., & Golan, S. (1993). The effects of standardized testing on teaching and schools. *Educational measurement: Issues and practice*, *12*(4), 20-25.

Hiatt, F. (2011, October 24). In South Korea, too many college grads, too few jobs. *The Washington Post.* Retrieved from www.washingtonpost.com.

Hitchcock, G., & Hughes, D. (1995). *Research and the teacher: A qualitative introduction to school-based research*. Psychology Press.

Hughes, A. (1993). Backwash and TOEFL 2000. Unpublished manuscript, University of Reading.

Hwang, H. J. (2003). The impact study of high-stakes exams on teachings and students: A washback study of the university entrance exam at the secondary school level in South Korea. Quebec, Canada: McGill University.

Israel, M., & Hay, I. (2006). Research ethics for social scientists: Between ethical conduct and regulatory compliance.

Jeon, M., & Lee, J. (2006). Hiring native-speaking English speakers in East Asian countries. *English Today,* 4*,* 53-58.

Johnston, B. (2003). *Values In English Language Teaching.* London: Lawrence Erlbaum Associates.

Jung, H. (2008). The washback effects of English listening test in Korean CSAT on secondary school English learning and teaching. *The Linguistic Association of Korea Journal*, *16*(3), 143-162.

Kane, M. (2002). Validating high-stakes testing programs. *Educational measurement: Issues and practice*, *21*(1), 31-41.

Kang, C. (2007). Does Money Matter? The Effect of Private Educational Expenditures on Academic Performance. *National University of Singapore Working Papers, 0704*.

Kang, N. H., & Hong, M. (2008). Achieving Excellence in Teacher Workforce and Equity in Learning Opportunities in South Korea. *Educational Researcher, 37*(4), 200-207.

Kim, D., & O, K. (2002). Washback on 12th grade English classes of the English tests within Korean university entrance exams. English Education, *57*(3), 303-331.

Kim, E. (2011, July 24). Korean society groaning under academic inflation. *Yonhap News.* Retrieved from http://yonhapnews.co.kr.

Kincheloe, J. L., & McLaren, P. (2002). Rethinking critical theory and qualitative research. *Ethnography and schools: Qualitative approaches to the study of education*, 87-138.

Knight, P.T. (2002). Summative Assessment in Higher Education: Practices in disarray. *Studies in Higher Education, 27*(3), 275-286.

Koo, H. (1993). *State and Society in Contemporary Korea.* Ithaca, NY: Cornell University Press.

Koo, H. (2007). The Changing Faces of Inequality in South Korea in the Age of Globalization. *Korean Studies,* 31, 1-18.

Kvale, S. (1996). *Inter Views: An introduction to qualitative research interviewing*. Thousand Oaks, CA: Sage.

Lee, C., & Green, R. T. (1991). Cross-cultural examination of the Fishbein behavioral intentions model. *Journal of international business studies*, *22*(2), 289-305.

Lee, J. H. (2004). The School Equalization Policy of Korea: Past Failures and Proposed Measures for Reform. *Korea Journal, 44*(1), 221-234.

Lee, J. K. (2002). Japanese Higher Education Policy in Korea During the Colonial Period (1910-1945). *Education Policy Analysis Archives,* 10(14).

Lee, K. (1984). *A New History of Korea*. (Edward W. Wagner & Edward J. Shultz, Trans.). Cambridge: Harvard-Yenching Institute.

Lim, J. H. (2005). Class Reproduction and Competing Ideologies in Korean Education: A Critical Discourse Analysis on "School Collapse," 1999-2001. In Young-Key, K. R., Grinker, R. R., & Larsen, K. W. (Eds.), *The Sigur Center Asia Papers: Korean Education,* 24, 17-32.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage Publications, Inc.

Ling, C. Y., & Braine, G. (2007). The Attitudes of University Students towards Non-native Speakers English Teachers in Hong Kong. *RELC Journal, 38,* 257-277.

Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language testing*, *18*(4), 351-372.

Mackenzie, N., & Knipe, S. (2006). Research dilemmas: Paradigms, methods and methodology. *Issues in educational research*, *16*(2), 193-205.

Madaus, G. F. (1988). The Influence of Testing on the Curriculum. In Tanner, L. N. (Ed.), Critical Issues in the Curriculum, 83-121. 87th Yearbook of the National Society for the Study of Education, Part I. Chicago: University of Chicago Press.

Magno, C. (2010). Korean Students' Language Learning Strategies and Years of Studying English as Predictors of Proficiency in English. *TESOL Journal, 2,* 39-61.

McNamara, T. (2000). *Language Testing*. Oxford: Oxford University Press

McNamara, T., & Roever, C. (2006). *Language Testing: The Social Dimension.* Oxford: Blackwell.

Mertens, D. M. (2014). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods*. Sage publications.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational Measurement,* (p. 13-103). New York: Macmillan.

Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13*(3), 241-256.

Moodie, I. (2016). The anti-apprenticeship of observation: How negative prior language learning experience influences English language teachers' beliefs and practices. *System*, *60*, 29-41.

Muñoz, A. P., & Álvarez, M. E. (2010). Washback of an oral assessment system in the EFL classroom. *Language testing*, *27*(1), 33-49.

Nunan, D. (2003). The Impact of English as a Global Language on Educational Policies and Practices in the Asia-Pacific Region. *TESOL Quarterly,* 37(4), 589-613.

OECD. (2005). Teachers Matter. Attracting, developing, and retaining effective teachers. Overview. http://www.oecd.org/education/preschoolandschool/ 34990905.pdf.

OECD. (2010). Education at a Glance 2010: OECD Indicators. http://www.oecd.org/edu/ skills-beyond-school/educationataglance2010oecdindicators.htm.

Oh-Hwang, Y. (1993). Linkage between home environments, child psychosocial maturity, and child academic achievements. *Gifted International*, *8*(1), 32-37.

O'Leary, Z. (2004). *The essential guide to doing research.* London: Sage.

Pan, Y. (2014). Learner Washback Variability in Standardized Exit Tests. *The Electronic Journal for English as a Second Language, 18*(2).

Park, S. J. & Abelmann, N. (2004). Class and Cosmopolitan Striving: Mother's Management of English Education in South Korea. *Anthropological Quarterly,* 77(4), 645-672.

Patton, M. (1987). *How to use qualitative methods in evaluation.* London: Sage.

Pennycook, A. (1990). Towards a Critical Applied Linguistics for the 1990s. *Issues in Applied Linguistics, 1*(1), 8-28.

Pennycook, A. (2010). *Critical applied linguistics: a critical introduction.* New York: Routledge.

Perry, F.L. (2005). *Research in Applied Linguistics: Becoming a discerning consumer.* London: Lawrence Erlbaum Associates.

Phillipson, R. (2012). *Linguistic Imperialism.* Oxford: Oxford University Press.

Ponterotto, J. G. (2005). Qualitative research in counseling psychology: A primer on research paradigms and philosophy of science. *Journal of counseling psychology*, *52*(2), 126.

Popham, J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappa*, May, 679-682.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta psychologica*, *104*(1), 1-15.

Qi, L. (2004). Has a High-Stakes Test Produced the Intended Changes? In Cheng, L., Watanabe, Y., & Curtis, A., *Washback in Language testing: Research Contexts and Methods*, (p.171-190). London: Lawrence Erlbaum Associates.

Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing, 22*(2), 142–173.

Qi, L. (2007). Is testing an efficient agent for pedagogical change. *Foreign Language Teaching*, *6*(1), 49-54.

Qian, D., & Cumming, A. (2017). Researching English Language Assessment in China: Focusing on High-Stakes Testing. *Language Assessment Quarterly, 14*(2), 97-100.

Ramsey, P. H. (1989). Critical values for Spearman's rank order correlation. *Journal of educational statistics*, *14*(3), 245-253.

Rhee, S. (1995). The importance of creativity education in Confucian culture. In Zixiu, Z., & Jiannong, S. (Eds.). Maximizing human potential—Post conference in China meeting of the 11th World Conference on Gifted and Talented Children—Beijing, China August (5-8, 1995). Beijing, China: Institute of Psychology & Center for International Scientific Exchanges, Chinese Academy of Sciences, *71*(1), 5-8.

Richards, K. (2003). *Qualitative inquiry in TESOL*. Springer.

Robinson, J. (1994). Social Status and Academic Success in South Korea. *Comparative Education Review, 38*(4), 506-530.

Robson, C. (2002). *Real world research: A resource for social scientists and practitioner-researchers* (Vol. 2). Oxford: Blackwell.

Rodriguez, J. M., & Arellano, L. (2016). The Impact of High-Stakes Testing on Latina/o Students' College Aspirations. *Journal of Hispanic Higher Education*, *15*(2), 113-135.

Schmid, G. (2013). Youth Unemployment in Korea: From a German and Transitional Labour Market Point of View. *IZA Policy Paper No. 63.*

Seidman, I. (2006). *Interviewing as Qualitative Research.* New York: Teacher's College Press.

Seth, M. J. (2002). *Education Fever: Society, Politics, and the Pursuit of Schooling in South Korea.* Honolulu: University of Hawaii Press.

Seth, M. J. (2005). Korean Education: A Philosophical and Historical Perspective. In Young-Key, K. R., Grinker, R. R., & Larsen, K. W. (Eds.), *The Sigur Center Asia Papers: Korean Education,* 24, 3-15.

Shi, L. (2006). The Successors to Confucianism or a New Generation? A Questionnaire Study on Chinese Students' Culture of Learning English. *Language, Culture and Curriculum, 19*(1), 122-147.

Shih, C. (2007). A New Washback Model of Students' Learning. *The Canadian Modern Language Review, 64*(1), 135-161.

Shin, Y. H. (1978). The Establishment of the First Modern School in Korea. *Korean Social Science Journal*, 5(1), 127-141.
Shin, S., & Koh, M. S. (2005). Korean Education in Cultural Context. *Essays in Education*, vol. 14.

Shohamy, E. (1982). Affective Considerations in Language Testing. *The Modern Language Journal,* 66(1), 13-17.

Shohamy, E. (1993). The power of tests: the impact of language tests on teaching and learning. *NFLC occasional papers.* Washington, DC: The National Foreign Language Center.

Shohamy, E. (2000). Fairness in Language Testing. In A. Kunnan (Ed.), *Studies in Language Testing 9: Fairness and validation in language assessment, (p. 15-19).* Cambridge: Cambridge University Press.

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

Shohamy, E. (2003). Implications of language education policies for language study in schools and universities. *The Modern Language Journal*, *87*(2), 278-286.

Shohamy, E. (2007). Language tests as language policy tools. *Assessment in Education*, *14*(1), 117-130.

Shohamy, E. (2010). Cases of language policy resistance in Israel's centralized educational system. In K. Menken & O. Garcia (Eds.), *Negotiating Language Policies in Schools. Educators as Policymakers*, 182-197. New York: Routledge.

Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: washback effect over time. *Language Testing*, *13*, 298-317.

Siu, S. F. (1992). *Toward an understanding of Chinese American educational achievement* (Report No. 2). Center on families, communities, schools, and children's learning.

Smith, L. (1992). Ethical issues in interviewing. *Journal of Advanced Nursing*, 17, 98–103

Smith, M. L. (1991). Put to the Test: The Effects of External Testing on Teachers. *Educational Researcher, 20*(5), 8-11.

Solorzano, R. (2008). High Stakes Testing: Issues, Implications, and Remedies for English Language Learners. *Review of Educational Research, 78*(2), 260-329.

Sorensen, C. W. (1994). Success and Education in South Korea. *Comparative Education Review, 38*(1), 10-35.

Spratt, M. (2005). Washback and the classroom: the implications for teaching and learning of studies of washback from exams. *Language Teaching Research, 9*(1), 5–29.

Stecher, B., Chun, T., & Barron, S. (2004). The effects of assessment-driven reform on the teaching of writing in Washington State. *Washback in language testing: Research contexts and methods*, *71*.

Tan, M., & Turner, C. (2015). The Impact of Communication and Collaboration Between Test Developers and Teachers on a High-Stakes ESL Exam: Aligning External Assessment and Classroom Practices. *Language Assessment Quarterly,* 12, 19-49.

Tollefson, J. W. (1995). *Power and Inequality in Language Education*. New York: Cambridge University Press.

Tsagari, D. (2009). *The Complexity of Washback.* Frankfurt: Peter Lang.

Valli, L., & Buese, D. (2007). The Changing Roles of Teachers in an Era of High-Stakes Accountability. *American Educational Research Journal, 44*(3), 519-558.

Walford, G. (Ed.). (2001). *Doing qualitative educational research*. A&C Black.

Wall, D. (1997). Impact and washback in language testing. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education: Vol. 7: language testing and assessment,* (291-302). Dordrecht: Kluwer Academic.

Wall, D. (2000). The impact of high-stakes testing on teaching and learning: can this be predicted or controlled? *System, 28,* 499-509.

Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory.* Cambridge: Cambridge University Press.

Wall, D., & Alderson, J. C. (1993). Examining washback: the Sri Lankan Impact Study. *Language Testing, 10,* 41-69.

Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing, 13*(3), 318-333.

Watanabe, Y. (2000). Washback effects of the English section of Japanese university entrance examination on instruction in pre-college level EFL. *Language Testing Update, 27,* 42-67.

Watanabe, Y. (2004). Methodology in Washback Studies. In Cheng, L., Watanabe, Y., & Curtis, A., *Washback in Language testing: Research Contexts and Methods*, (p.19-36). London: Lawrence Erlbaum Associates.

Watanabe, Y. (2013). The National Center Test for University Admissions. *Language Testing, 30*(4), 565-573.

Wise, S. & DeMars, C. (2005) Low Examinee Effort in Low-Stakes Assessment: Problems and Potential Solutions. *Educational Assessment, 10*(1), 1-17.

Wiseman, S. (1961). *Examinations and English education.* Manchester: Manchester University Press.

Xie, Q. (2015). Do component weighting and testing method affect time management and approaches to test preparation? A study on the washback mechanism. *System*, *50*, 56-68.

Xie, Q., & Andrews, S. (2013). Do test design and uses influence test preparation? Testing a model of washback with Structural Equation Modeling. *Language Testing*, *30*(1), 49-70.

Yoo, J. H. (1973). The System of Korean Local Government. In Nahm, A. (Ed.), *Korea Under Japanese Colonial Rule: Studies of the Policy and Technique of Japanese Colonialism.* Kalamazoo, MI: Western Michigan University, Institute for International and Area Studies, Center for Korean Studies, 55-59.

Yoon, M. (2014, July 27). Credential inflation hurting job-seekers. *The Korean Herald.* Retrieved from www.koreanherald.com.

Yu, G. (2007). Students' voices in the evaluation of their written summaries: Empowerment and democracy for test takers? *Language Testing, 24*(4), 539-572.

Zhan, Y., & Andrews, S. (2014). Washback effects from a high-stakes examination on out-of-class English learning: insights from possible self theories. *Assessment in Education: Principles, Policy, and Practice, 21*(1), 71-89.

Zou, S., & Xu, Q. (2017). A Washback Study of the Test for English Majors for Grade Eight (TEM8) in China—From the Perspective of University Program Administrators. *Language Assessment Quarterly, 14*(2), 140-159.