

High-Stakes Language Testing and Young English Language Learners: Exploring L2
Motivation and L2 Test Validity in South Korea

by

John F. Haggerty

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Arts

Applied Linguistics

Carleton University

Ottawa, Ontario

December, 2010

© John F. Haggerty



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-79568-2
Our file Notre référence
ISBN: 978-0-494-79568-2

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

Abstract

This thesis explores the L2 attitudes, beliefs, and self-conceptions of young language learners in South Korea, based on six learner-background characteristics, in an attempt better understand their L2 motivational patterns. In addition to considering gender, educational level, and self-reported L2 proficiency, this study will also consider the time learners spent preparing for L2 tests, the level of L2 test completed (middle-school or university-entrance), and the frequency of university-entrance level test completion. The inclusion of the last three research questions is an explicit attempt to widen conceptions of L2 motivation to more properly include the effects of L2 testing, especially in test-driven educational environments. A questionnaire was developed and administered to middle school students enrolled in a private language school in Seoul (N=341). An exploratory factor analysis (EFA) approach was utilized in order to identify factors for further comparison of subgroups. The results indicated that middle school students differed significantly on a majority of L2 motivational factors depending on: 1) their perceived L2 proficiency level; 2) the amount of time they spent preparing for L2 tests; and 3) the completion of university-entrance level L2 tests. The implications of these findings are discussed in relation to some of the major motivational theories advanced in the fields of psychology and second language acquisition (SLA) and the potential implications this has for modern notions of validity in language testing. Finally, some recommendations are offered to improve the administration of high-level, high-stakes tests that affect English language learners (ELLs) around the world.

Acknowledgements

I would like to first thank my thesis advisor, Dr. Janna Fox, for her invaluable guidance in navigating the troubled waters of field research in a foreign country. I benefitted tremendously from her wealth of experience in language testing and her wise words of encouragement when things did not go quite as I had planned. She was often the only sympathetic ear I could find once my closest friends grew weary of my ramblings. Without her, I am quite sure I could not have completed this research. I would also like to thank my graduate studies supervisor, Dr. Devon Woods, for having the faith to allow me to do off-campus research and for inspiring me to follow my heart and pursue research where I felt it was necessary most. His dogged pursuit to uncover what lies beneath our beliefs, assumptions, and knowledge has had a far greater impact on this inquiry than he may have realized. I am also greatly indebted to Joan Grant for her excellent administrative support and for fully explaining all the options I had available to me. A special thanks to Karen Rishel for getting necessary paperwork to the correct people on campus when I could not be there myself. I would also like to thank Young Jin for not only being an invaluable translator, but also a dear friend whose insight into education and motivation in Korea helped me to better organize and understand my research. I would also like to thank the Kiervin family for opening up their home and providing support during my thesis defence in Ottawa, Canada. Finally, this research would not be possible without the support of Seong Hye. When I started this study she was my fiancée, but when I finished she was my wife. I think this speaks the loudest to not just how important she was in this study, but as it turns out, in my life as well.

Table of Contents

Abstract.....	ii
Acknowledgements.....	iii
Table of Contents.....	iv
List of Tables.....	vi
List of Figures and Illustrations.....	vii
List of Symbols, Abbreviations and Nomenclature.....	viii
CHAPTER ONE: INTRODUCTION.....	1
CHAPTER TWO: VALIDITY IN LANGUAGE TESTING.....	7
2.1 The ‘Early’ Years: Criterion-related and Content Validity.....	9
2.2 The Ascendance of Construct Validity.....	12
2.3 Positivism, Psychometrics, and Validity.....	13
2.4 The Messick Model.....	15
2.5 The Bachman and Palmer Model.....	18
2.6 Kane’s Argument-Based Approach.....	20
2.7 McNamara and the Social Context of Language Testing.....	22
2.8 Consequential Validity: A Role for Motivation?.....	24
CHAPTER THREE: MOTIVATION RESEARCH.....	27
3.1 Motivation in Psychology.....	29
3.1.1 Achievement Motivation Theory.....	30
3.1.2 Goal-Setting and Goal-Orientation Theory.....	31
3.1.3 Attribution Theory.....	32
3.1.4 Self-Efficacy.....	33
3.1.5 Self-Worth Theory.....	34
3.1.6 Self-Determination Theory.....	35
3.1.7 Weiner’s Social Motivation.....	36
3.2 Motivation in SLA.....	37
3.2.1 Gardner and Lambert’s ‘Integrativeness’.....	37
3.2.2 1990 to the Present Day: Rethinking and Revisualizing L2 Motivation.....	38
3.2.3 Dörnyei’s L2 Motivational Self-System.....	40
3.2.4 A Greater Appreciation of Young ELLs in L2 Motivation.....	43
3.2.5 A Greater Role for Language Testing in L2 Motivation Research.....	45
3.3 Development of L2 Motivational Constructs.....	46
3.3.1 Dörnyei -Inspired Constructs.....	47
3.3.2 Pilot Study Constructs.....	50
3.4 Learner-background Characteristics (Variables).....	51
3.4.1 Gender.....	51
3.4.2 Education Level.....	52
3.4.3 Self-reported L2 Proficiency.....	53
3.4.4 L2 Test Preparation.....	54
3.4.5 Level and Frequency of L2 Test Completion.....	54

CHAPTER FOUR: RESEARCH METHOD	56
4.1 Participants	57
4.2 Setting	57
4.3 Instrument	58
4.4 Procedure	59
4.5 Data Analysis	60
CHAPTER FIVE: RESULTS	64
5.1 Exploratory Factor Analysis Results	64
5.1.1 Factor 1: Ought-To L2 Self + Instrumentality (prevention).....	67
5.1.2 Factor 2: Attitudes to L2 Community	68
5.1.3 Factor 3: Attitudes to L2 Learning	69
5.1.4 Factor 4: Ideal L2 Self.....	69
5.1.5 Factor 5: Attitudes to L2 Testing	70
5.1.6 Theoretical Constructs Not Replicated by EFA Procedure.....	70
5.2 Factor Scores	71
5.2.1 Correlation among Factor Score Variables	71
5.2.2 Results for Gender and Education Level.....	72
5.2.3 Results for Self-Reported L2 Proficiency	73
5.2.4 Results for L2 Test Preparation.....	74
5.2.5 Results for Level and Frequency of L2 Test Completion	77
CHAPTER SIX: DISCUSSION	80
6.1 L2 Motivational Factors Identified	80
6.2 Gender and Education Level	82
6.3 Self-Reported L2 Proficiency	84
6.4 L2 Test Preparation.....	86
6.5 Level and Frequency of L2 Test Completion	87
CHAPTER SEVEN: CONCLUSIONS AND RECOMMENDATIONS.....	90
7.1 Limitations.....	90
7.2 Conclusions about L2 Motivation	91
7.3 Conclusions about Validity in Language Testing.....	95
REFERENCES	100

List of Tables

Table 1: Theoretical Constructs Utilized for Present Study	49
Table 2: Pattern Matrix for 5-Factor Solution (EFA)	66
Table 3: Pearson Product-Moment Correlations for Factor Score Variables	72
Table 5: Factor Scores and Self-Reported L2 Proficiency Level (N=329)	73
Table 4: Factor Scores and L2 Test Preparation (N=309)	75
Table 6: Factors and Level of L2 Test Completion (N=332)	78

List of Figures and Illustrations

Figure 1: Messick's Validity Matrix.....	16
Figure 2: Sampled Items from Taguchi Study in Chinese Context.....	50
Figure 3: Sampled Items from Pilot Study in Korean Context.....	51
Figure 4: Theoretical Constructs and EFA Factors Identified.....	64
Figure 5: Scree Plot for 5-Factor EFA Solution	67

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
AMT	Achievement Motivation Theory
ANOVA	Analysis of variance
BAK	Beliefs, assumptions and knowledge
CBT	Computer-based test
CFA	Confirmatory factor analysis
EFA	Exploratory factor analysis
EFL	English as a foreign language
ELL	English language learner
GPA	Grade point average
Jr. G-TELP	Junior General Test of English Language Proficiency
iBT	Internet-based test
KMO	Kaiser-Meyer-Olkin measure of sampling adequacy
L2	Second language
PBT	Paper-based test
PAF	Principal axis factoring
PELT (elementary)	Practical English Level Test for Elementary English
SEM	Structural equation modelling
SDT	Self-Determination Theory
SLA	Second language acquisition
SPSS	Statistical Package for the Social Sciences
TEPS	Test of English Proficiency
TLU	Target language use
TOEFL	Test of English as a Foreign Language
TOEIC Bridge	Test of English for International Communication (Bridge)
TOSEL	Test of Skills in English Language
Tukey's HSD	Tukey's 'Honestly Significant Different' post hoc criterion

It is difficult to find a person who does not have a testing story that relates to how a single test affected and changed his or her life, for good or for bad. The experiences of taking tests are remembered by test takers for years after the events have taken place.

Elana Shohamy, 2001

Chapter One: Introduction

For almost a decade now, I have been an EFL teacher in South Korea. A large amount of that time has been spent preparing middle school language learners for various language tests, particularly the Test of English as a Foreign Language (TOEFL) in its paper-based (PBT), computer-based (CBT) and now its internet-based (iBT) formats. Originally, I thought of this work as providing valuable assistance to young students that would improve their academic opportunities and enhance their future careers. I suppose I also thought of it as a necessary source of income and, as such, refrained from thinking about it critically. Over the years, I became increasingly concerned about the potential effects (both positive and negative) of this type of test preparation, and ultimately, what the ‘testing stories’ of my young students would eventually become.

This thesis offered me the opportunity to dig a little deeper into the L2 motivational attitudes, beliefs, and self-conceptions of young Korean language learners and try to better understand the complex network of influences that shaping and re-shaping it. It also gave me an opportunity to come to terms with my own role as a test-preparation teacher for young learners. The language test I helped prepare young students for, the Test of English as a Foreign Language (or TOEFL), was cognitively-demanding. This test challenged my young learners to understand complex cognitive concepts and

intricate relationships intended for older learners and they were expected to do all of this in a second language when, quite often, they had not yet learned these concepts in their native language. This test, as was often impressed upon me by Korean colleagues, parents, and students themselves, also played a dramatic role in whether middle school students could get accepted to more prestigious high schools. English language performance was, and is, an important consideration in the high-school application process. The high school one attended would greatly influence the university one could go to, which in turn, would affect what potential career one could expect to have. For the great majority of students I spoke with throughout my decade in the educational context, there was a strong perception of an ‘educational domino effect’ that generally started at middle school – if not sooner. The stress associated with all important tests, particularly in the third and final year of middle school, was quite palpable in the classroom. The TOEFL, as well as the Test of English proficiency (TEPS) which was introduced to the scene later, were indeed ‘high-stakes’ language tests for the middle-school students I taught.

In South Korea more generally, the importance of English language learning and testing is powerful and omnipresent. The implementation of the 7th National Curriculum in 2000 (which further expanded the role of English education in public schools), the proliferation of private language schools throughout the country, the high percentage of family income spent on English language education, and the continued rise of English language requirements as gatekeeping devices for education and employment, are all testaments to the significance and magnitude of English language learning in South Korea (for a fuller discussion, see Nam, 2006; Choi, 2008). For many, the results of

language tests can dramatically influence the opportunities available to them and their ability to parlay these opportunities into success in a highly competitive education and employment market. As such, there is extreme pressure to perform on English tests even at quite young ages, some English language learners (ELLs) taking advanced English proficiency tests as early as elementary school (Choi, 2008, p. 53). These larger political, social, and educational issues may not be completely understood by the average middle-school learner in South Korea, but it is highly likely they will be affected by them. The main objective of this thesis is to better understand how this may be influencing their L2 motivation.

The TOEFL, created by the Educational Testing Service in the U.S., has been administered in South Korea for years as a benchmark (amongst other factors) for acceptance to most universities and foreign language high schools (Choi, 2008, p. 45). More recently, the TEPS created by Seoul National University in South Korea has risen in prominence to compete with the TOEFL in the language testing market. Both of these tests purport to measure the academic skills necessary to perform in English at the university level. These are two of the major tests used in Korea to measure academic English proficiency. There are no specific recommendations regarding the appropriate age for ELLs to begin preparing for or taking the test. The creators and administrators of these tests do not explicitly state that their test should not be taken by young ELLs. These are decisions presumably left up to test users and test takers. Despite the reality that young ELLs are taking these cognitively-demanding tests as early as elementary school, a fairly glaring example of test misuse, test designers have not adequately investigated the validity implications this may have for their tests.

However, if modern notions of validity are to be taken seriously, much more investigation into the nature and impact of high-stakes tests on younger ELLs is needed. While there are many facets of validity that need to be addressed with respect to younger ELLs, this study will focus on the social consequences of test use (Messick, 1989) by exploring the L2 attitudes, beliefs, and self-conceptions of young language learners in South Korea, based on six learner-background characteristics, in an attempt to better understand young learner L2 motivational patterns. Since previous motivational studies conducted with young learners have found significant effects with regards to gender, educational level (age), and self-reported proficiency, this research will first consider the following research questions:

- Are L2 motivational factors influenced by gender?
- Are L2 motivational factors influenced by educational level?
- Are L2 motivational factors influenced by self-reported L2 proficiency?

While there have been many L2 motivation variables researched and written about, what has been lacking in many L2 studies has been a more explicit investigation of L2 testing and how it may motivate or de-motivate young language learners. Therefore, in attempting to redress this imbalance, three additional research questions were considered:

- Are L2 motivational factors influenced by the time spent preparing for L2 tests?
- Are L2 motivational factors influenced by the level of L2 test completion (middle-school or university-entrance)?
- Are L2 motivational factors influenced by the frequency of university-entrance level L2 test completion?

The inclusion of the last three research questions is an explicit attempt to widen the current conception of L2 motivation to more properly include the effects of L2 testing, especially within test-driven educational environments. This thesis also hopes to add to a growing understanding of the factors influencing L2 motivation among young ELLs. By creating a more explicit link between the consequences of L2 test use in terms of L2 motivational attitudes, beliefs and self-conceptions of young language learners, this thesis also aspires to demonstrate the importance of considering the wider social context in the test validation process. Implicit throughout this thesis is the belief that the social consequences of language testing are an integral part of a continual test validation process, and therefore, the ultimate responsibility of test designers.

Before investigating these important questions, it is necessary to gain a better appreciation of the historical evolution of the issues involved. Understanding where we have been often provides valuable information about where we are and where we might need to go. Chapter Two provides a brief overview of the evolution of the concept of validity over approximately the last half century in order to provide some context for a validity argument that remains as highly contentious today as when it was first proposed by Samuel Messick in the 1980s. Chapter Three explores some of the more influential theories of human motivation from the fields of psychology and second language acquisition (SLA) in order to better understand and interpret the L2 motivational constructs that have informed this investigation. Chapters Four and Five describe the research methodology utilized to investigate L2 motivation in this context and the results obtained. Chapter Six discusses the major findings in light of some of the research questions considered for this study. Finally, Chapter Seven offers some conclusions

based on the motivational theories examined as well as my own interpretation of the data. The potential implications of these results for modern notions of validity in language testing will also be addressed. To close out this thesis, some recommendations will be offered for improving the creation and administration of high-level, high-stakes L2 tests that could potentially help to minimize the negative consequences on young ELLs suggested by this study.

Chapter Two: Validity in Language Testing

The field of language testing has gone through a considerable amount of change in its relatively brief history. Of course, this is only brief in relation to the histories of other more established fields that can more effortlessly trace their lineage deep into the past. Certainly the *field* of language testing cannot claim as historical a pedigree as the field of philosophy or mathematics. However, as Spolsky (2007) has reminded us, a discussion of the history of language testing that begins at some point in the latter half of the 20th century, perhaps with the seminal work of Carroll (1961) or Lado (1961), reveals a lack of ‘historical sense’ (Spolsky, 2007, p. 12). In fact, the testing of language stretches at least as far back as biblical times with the use of the Shibboleth test to identify linguistic markers in order to distinguish (and all too often extinguish) members of closely related tribes (Spolsky, 1995; McNamara, 2005). Unfortunately, the analysis of validity presented in this chapter suffers from precisely this kind of short-sightedness as it is heavily weighted towards theoretical developments that have occurred over only the last five decades. Nevertheless, this short period has experienced significant change and has inspired a great deal of research, including the research reported in this thesis. Given this, it is hoped the oversight will be forgiven.

A great deal of recent debate in language testing has revolved around the nature and scope of validity and the appropriate method of argument to be used to support test use. This debate reflects the complex and contentious issues that continue to surround the development, administration, interpretation, and social use of language tests in a surfeit of local, national and international contexts. The proliferation of language testing, used

for an increasing variety of purposes, has caused many to question what test ‘validation’ really means. Our position on this important question reveals a great deal about our beliefs, and underlying assumptions, about the role of language test designers, users, and other stakeholders in an international testing market.

This chapter will begin with a very brief review of two ‘early’ concepts utilized in the test validation process: *criterion-related* (predictive and concurrent) and *content validity*. These two approaches to validation remained dominant in language testing until at least the 70’s and continue to hold significant influence today. A third major concept, *construct validity*, was originally developed to deal with instances in which criterion or content validation was deemed inadequate, as was the case with many psychological tests (Cronbach & Meehl, 1955). Beginning with urgent calls for test designers to “make a place for it in their methodological thinking” (ibid., p. 300), construct validity would eventually rise in importance in the 80’s and 90’s and begin to challenge test designers to support all test score interpretations with theoretical rationale and empirical evidence. This will be followed by a brief discussion of the dominant role played by psychometrics in language testing and how significant theoretical changes in approaches to language acquisition and use have influenced the field of language testing to move beyond some of its positivistic assumptions.

To bring us to the present day, this chapter will turn to some of the major contributions made to our understanding of validity by four influential modern test theorists: Samuel Messick, Lyle Bachman, Michael Kane, and Tim McNamara. Their conceptions of validity continue to challenge test designers and users to acknowledge the assumptions inherent in every decision made, more fully acknowledge the wider social

context of language testing, and begin to incorporate the social consequences of test use into the test validation process.

Finally, this chapter will conclude with a discussion of the practical limits of validity and the potential for an increasing role for motivation research in high-stakes test validation processes. Major theorists in the field of language testing have been calling for greater responsibility since the 1950's, when the concept of construct validity was originally developed. Since that time, there has been increasing demands for test designers to move beyond psychometric analysis of the internal components of tests and begin to investigate the social consequences of tests on individuals and society (e.g., Messick, 1989, 1995, 1998; Shepard, 1993, 1997; Moss, 1998, McNamara, 2001; McNamara & Roever, 2006). Concepts such as 'use-oriented testing' and 'critical language testing' have challenged test designers to address questions such as "what happens to test-takers taking tests, to the knowledge created by tests, [and] to teachers preparing their students for tests..." amongst others (Shohamy, 2008, p.363). This has challenged test designers to begin to address some of their most difficult validity issues.

2.1 The 'Early' Years: Criterion-related and Content Validity

Three major concepts of validity, (content, criterion-related, and construct) have been most dominant in the history of test validation (Crocker & Algina, 1986) and have traditionally been associated with specific testing objectives (Messick, 1989, p. 6). Two of these, criterion-related and content validity, were developed in the 'early' years of language testing and continued to be heavily relied upon until at least the late 70's (Kane, 2001). In very general terms, these concepts are concerned with: 1) the ability of test

scores to concur with or predict performance on some external *criterion* (e.g., the scores obtained on an established mathematical aptitude test or scores obtained in a mathematics course), and 2) the relevance of test *content* to the subject area being assessed (e.g., basic mathematical concepts). In addition to claims of reliability (through the use of reliability coefficients that measured consistency from one testing event to another), early test validation procedures would almost exclusively rely upon these methods, quite often completely independent of each other (Kane, 2001).

In order to establish criterion-related validity, test makers required an external criterion (or measure) that could be used as a point of comparison for test scores. If the scores on any given test could be shown to correspond to this external measurement, the test could be said to have criterion-related validity. A strong correlation indicated that the test had concurrent or predictive strength and was thus ‘validated.’ This positive result could then be used as a point of reference to support the validity of future test performances as well. No attempt was generally made to acknowledge the assumptions that lay behind the choice of test tasks, test content, or the matching external criteria because “the validity of the criterion, defined here in terms of ‘task performance,’ was taken for granted” (Kane, 2001, p. 320). In reality of course, there were numerous unsupported inferences involved in the matching of external criteria to test scores, and these became increasingly contentious when criterion-related arguments were extended over time and context (Messick, 1989, p. 7). Often, the criterion chosen for comparison, for example grade point average (GPA), involved a number of variables (e.g., social networks of support or financial ability) that the test did not measure at all (Fox, 2004). In addition to these significant limitations, there was an even greater obstacle confronted

when, as was often the case, no external criterion was available for comparison. To overcome this, another validity argument would be put forth based on the content selected for the test.

In the absence of an external criterion, test designers would often turn to ‘experts’ in the field, invariably academics, to examine test content and assess its relevance and representativeness to determine its *content* validity (Messick, 1989; Kane, 2001). While a seemingly reasonable method for establishing the appropriate content of a subject area, there were inherent problems with this approach, especially when relied upon exclusively. First of all, judgements, even those of the ‘expert’ kind, are “fallible and may imperfectly apprehend domain structure or inadequately represent test structure, or both” (Messick, 1989, p.7). Even if suitable experts could be agreed upon and content could be selected that was relevant to, and representative of, the content domain, a classical content validity framework completely sidestepped vital issues such as the selection of test method, the internal and external score relationships, the appropriate inferences that could be made about test performances, and the social consequences that resulted from test use. This lack of appreciation for potential sources of irrelevant test variance, that in fact invalidate any test, led Messick to conclude that “content validity does not qualify as validity at all” (ibid., p. 7). Regardless of their limitations, these two early test validation approaches would continue to be heavily relied upon well into the 70’s (Kane, 2001). In addition to their convenience and practicality in the test validation process, their dominance could also be attributed to the theoretical preeminence of positivism that pervaded most every other field of inquiry and, directly connected to this, the overwhelming acceptance of psychometrics as the primary analytical technique.

2.2 The Ascendance of Construct Validity

Owing its existence to the need to justify the interpretations of clinicians and psychologists in the early 50's (Cronbach, in Kane, 2001, p. 321), construct validity was originally viewed in testing as simply an add-on to criterion-related and content validity, often utilized when the more established approaches proved inappropriate or insufficient (Kane, 2001). The original concept, expanded upon by Cronbach and Meehl (1955), utilized a hypothetico-deductive approach whereby theoretical constructs, through "correspondence rules," were connected to a complex nomological network (or theory) and are "validated by checking the empirical laws against data" (Kane, 2001, p. 321). This approach was primarily concerned with providing theoretical argumentation and evidence in order to validate test constructs.

In the 60's, there was little indication that "criterion or content models were to go away or be subsumed under construct validity" (Kane, 2001, p. 322). It would take growing dissatisfaction with the assumptions inherent in these earlier models, and increasing awareness of the theoretical developments occurring in concepts of language acquisition and use, to push construct validity to the forefront of the test validation process. These factors lead to more critical approaches that increasingly questioned the unacknowledged assumptions characterizing test validation at the time. Of particular concern to some was the often piecemeal and selective way validity evidence was collected and presented (Messick, 1989) amounting to validity evidence being used as needed, much like a 'toolkit' (Kane, 2001, p. 323). In addition, it was becoming more apparent that language tests, as they always had been, were used for social and political

purposes and that this could not continue to be conveniently ignored in the test validation process.

Not long after construct validity became part of the 1954 Technical Recommendations (American Psychological Association, 1954), Loevinger (1957) recognized the unified nature of construct validity in which lines of evidence refer to test “content, its internal structure, and relation to outside variables” (p.689), which for her, represented the “whole of the subject, from a systematic, scientific, point of view” (p. 641). More than three decades later, Samuel Messick (1989) would present a unified vision of validity that could only be fully satisfied when the theoretical construct was defined, test relevance and utility was established, the value implications inherent in the interpretation of test scores was explicitly stated, and the social consequences of test use were investigated.

2.3 Positivism, Psychometrics, and Validity

Traditional approaches to validity and reliability were greatly assisted, and heavily reinforced, by significant advances made in the fields of mathematics and statistics in the first half of the 20th century. These revolutionary methods permitted complex relationships amongst individual item responses and overall test performances to be analyzed over time, and from context to context, in increasingly sophisticated ways. It would not be long before psychometrics became firmly entrenched in the field of language testing, solidifying statistics as the main tool in the analysis of validity. For the first time, test designers could apply widely-accepted, although not widely-understood, statistical procedures to explore whether their tests were reliably and ‘validly’ measuring

what they were purporting to measure. Perhaps more than any other development in the history of language testing, this has had the greatest impact on improving the consistency and, at least to this extent, the fairness of language testing.

This mathematical revolution, not coincidentally, occurred in an era dominated by positivistic thought and a strong belief in the value of the scientific method. This paradigm, affecting virtually every field to its core, encouraged test designers to employ complex mathematical techniques, often understood by only a select few in the field, in order to provide the support needed to validate test scores. Traditional validation concepts could now be augmented by widely accepted scientific methods and, for many, mystical mathematical formulas. Given the often ad-hoc approach to validation apparent in the past (Loevinger, 1957), this represented enormous progress. It encouraged the application of consistent methods of reliability and validity measurement as well as a common language for reporting the results.

The availability of these new statistical techniques fit quite nicely with new theoretical developments. Robert Lado in his groundbreaking book *Language Testing* published in 1961 argued that language could be broken up into its constituent parts (or *discrete points*) and that language learners could then be tested on their knowledge of these points. This encouraged test designers to develop a battery of questions that offered mathematically-friendly measures of language knowledge often comprised of multiple choice or similarly closed-ended test formats. These kinds of tests would come to dominate the approach to the testing of language in the 60's.

However, quite dramatic changes in the way theorists conceived of language were occurring in the 1970's, most notably those influenced by Dell Hymes's (1971) and his

proposed concept of communicative competence, a response to (and criticism of) the dominant Chomskian approach. This would begin to have dramatic effects on conceptions of what it actually meant to know, use, and ultimately test language. Psychometrics, while very powerful in the analysis of responses to individual test items, had considerable more difficulty assessing socially-inspired conceptions of language, for example 'sociolinguistic competence' or 'strategic competence' (Canale & Swain, 1980). In fact, due to its theoretical assumptions, psychometrics had obscured the social context of language testing, a concealment that may very well have been deliberate (McNamara & Roever, 2006). It would not be until psychometrics began to lose some of its imperious role in the assessment of validity that questions about the underlying assumptions of validation practices could begin to gain strength. The concept of construct validity that was originally developed in the early 50s would now begin to take on new meaning and importance in the 70's. However, it would take the efforts of a number of major test theorists, some of them psychometricians themselves, to push the concept of validity even further.

2.4 The Messick Model

Perhaps the most influential modern concept of validity in language testing has been put forward by test theorist and psychometrician Samuel Messick (1989, 1995) in which he argued for a significant expansion of the concept of validity. His progressive 'validity matrix' (see Figure 1 below) merged four major facets of testing into an inseparable symbiotic relationship, comprised of construct validity, test relevance and

utility, the value implications involved in test design and interpretation, and the social consequences of test use.

	Test Interpretation	Test Use
Evidential Basis	Construct Validity (CV)	CV + Relevance/Utility (R/U)
Consequential Basis	CV + Value Implications (VI)	CV + RU + VI + Social Consequences

Source: Messick, 1995, p. 748.

Figure 1: Messick's Validity Matrix

According to Messick (1989), this unified concept “integrates considerations of content, criteria, and consequences into a construct framework for testing rational hypotheses” (p. 8). Some have incorrectly assumed that this matrix permits the consideration of one cell to the exclusion of others, suggesting that construct validity research can somehow occur in a “social and value vacuum” (McNamara, 2001, p. 336). However, this is clearly not what Messick had intended. As a ‘unified concept’, he was arguing that validity in language testing involves value judgements from the very first decision made, the implication being that “both meaning and values, as well as both test interpretation and use, are intertwined in the validation process” (Messick, 1995, p. 749). These value judgements have real-world social consequences that need to be properly considered in order for a unified vision of validity to be properly satisfied. Messick’s

unified vision, placing the social consequences of testing at the apex, represented a revolution in thinking about validity and still continues to stir up a great deal of debate about the practical limits of test designers' social responsibility.

While it was clear that Messick's model would challenge the field to question their theoretical assumptions at every stage in the test validation process, it was less clear, at least to various critics of the model, how responsible test designers should be for the consequences of the use of their tests. For Messick, there was no confusion; any reasonable concept of validity would need to include an examination of how the test played out, in terms of construct representation and relevance, in the actual social context in which the test was used. Failure to include this consideration was a failure to consider how the interpretations of test scores were being used, the results of which could directly impact the validity of the test. This was especially important if any negative consequences of test use could be traced back to the constructs themselves.

However, some critics of a unified vision of validity, correctly or incorrectly, questioned whether test designers could, or even should, be expected to examine the consequences of the test use. While few would question the position that tests, especially those involving high-stakes, had significant impacts on individuals and society, there was reluctance to make this part and parcel of the concept of validity and a necessary component of the test validation process. For Messick, its inclusion was vital. However, it was not that all consequences, whether intended or not, needed to be properly accounted for, but rather that the social consequences be investigated in relation to the constructs developed by the test. This was the only way to fully ensure that test scores were actually measuring what they were purporting to measure.

Despite the ‘unified’ nature of Messick’s validity concept, he would later clarify that this does not “imply that validity cannot be usefully differentiated into distinct aspects to underscore issues and nuances that might otherwise be downplayed or overlooked” (Messick, 1995, p. 5). However, he argued that each component, “whether labeled content validity, criterion-related validity, or whatever – cannot stand alone in validity arguments (Messick, 1998, p. 37). For Messick, the separation could only be temporary, and in the end, accumulated evidence needed to be “integrated into an overall validity judgement to sustain score inferences and their action implications” (Messick, 1995, p. 8).

2.5 The Bachman and Palmer Model

Another influential reformulation of the test validation process has been put forth by Bachman and Palmer (1996) in their proposed model of ‘test usefulness’. Six test qualities, comprised of reliability, construct validity, authenticity, interactiveness, impact, and practicality, were proposed as a pragmatic method for evaluating the strengths and weaknesses of a test. While impossible to maximize all these qualities, this framework could enable test designers to assess the overall usefulness of a test utilized for a specific purpose. Within the framework, construct validity was isolated as one of a number of other factors involved in test design, implementation, interpretation and use. The social consequences of test use, instead of unified under construct validity, was treated as a separate variable labelled ‘impact.’

Bachman and Palmer’s model purported to provide a more reasonable and practical way for test designers and users to assess various components of test

'usefulness.' By avoiding the controversy surrounding a fully unified version of construct validity, they hoped to provide a taxonomy of test qualities that could be evaluated, more or less in isolation, in order to determine the relative contribution of each quality to the overall 'usefulness' of a test. In this process, each quality could add or detract from test usefulness, but it was the combined effect of all test qualities that would determine if the test should be utilized for a specific purpose. In order to prevent a disproportionately negative consequence resulting from one test quality, minimum acceptable levels needed to be established for them all although few guidelines were provided for how these minimum levels could be set. Bachman & Palmer (1996) applied their model to a number of real-world examples in order to demonstrate how their test qualities could be put to work in the test validation process.

Despite the supposed practicality of the model for test designers, the model did not appear to gain much popularity in practice (Davies, in McNamara, 2006, p. 35). While theoretically valuable for its explicitness in terms of test qualities and the importance of establishing a connection to the target language use (TLU) situation, specific guidelines about how to determine minimum acceptable levels or how to assess the relative contributions of each quality to the overall assessment of test usefulness were far less explicit. Bachman (2005) later conceded that taxonomies of test qualities (including his own) essentially provide a "list of independent qualities" to consider and do not "bring us any closer to a coherent theory of test use" (p. 4). Instead, he suggested the utilization of an 'assessment use argument' in the format of a Toulman argument (consisting of claims, warrants, and rebuttals) in order to more explicitly address the connection between validity and test use. This approach mirrored to some extent another

approach to validation developed by Michael Kane that would bring the validity argument itself center stage in test validation efforts. It is to his ideas to which this thesis will now turn.

2.6 Kane's Argument-Based Approach

Michael Kane (1992, 2001) has discussed a number of problematic issues that have resulted from the adoption of a unified model of construct validity. One of the more unfortunate side effects of a unified vision, utilizing a distinction originally coined by Cronbach (1988), has been the bifurcation of test validation efforts into more idealized *strong* programs and more pragmatic *weak* programs. Strong programs, due to their rigorous theoretical requirements, became unwieldy for many test designers. Weak programs often resulted in highly disorganized and even opportunistic 'wastebaskets' of validity evidence (Cronbach, in Kane, 2001, pp. 326-327). Eschewing the label 'construct validity' entirely, Kane proposed an 'argument-based approach' to test validity that focused instead on the interpretative argument that must inevitably entail "the test score as a premise and the statements and decisions involved in the interpretation as conclusions" (Kane, 1992, p. 527). In order to validate an interpretation of test performance, the argument must acknowledge the assumptions and inferences on which a score interpretation is based and upon which decisions are made. Essential to the validity argument is the identification of competing interpretations and the gathering of sufficient evidence to support the inferences and assumptions on which test scores are based. Reminiscent of Bachman and Palmer's pragmatism, Kane also concedes that the validity argument cannot be absolute in any sense, but instead one that is 'highly plausible' based

on the clarity and cohesion of the arguments and the plausibility of its assumptions (Kane, 1992, pp. 527-528).

In order to understand the kinds of interpretative arguments that are regularly made in support of tests, Kane (1992) provides six general categories of inferences: observation, generalization, extrapolation, theory-based, decision, and technical. Validity evidence is collected to address the various kinds of inferences made in order to interpret a test score for a particular use. Highly effective evidence addresses the weakest inferences and assumptions and serves to strengthen the overall validity argument. In the end, interpretive arguments for any given test display a number of distinct characteristics: they are “artifacts [in that they are created, not factual], they change with time, they may need to be modified for particular examinees or circumstances, and they are more or less plausible” (Kane, 1992, p. 533). Kane makes it clear that this is an *approach* to validity and not a *type* of validity. This approach provides a reasonable process for assessing the arguments made in support of test score interpretation and the ways test scores are ultimately used. It is also an evolving process, one that continually challenges the assumptions and inferences made in the interpretation and use of test scores, responds to competing interpretations, and as a result, continually strengthens the validity argument.

The strength of Kane’s approach in providing a framework for analyzing the interpretation of test scores has been acknowledged by other test theorists (for example, Bachman, 2005); however, there has also been concern expressed over an apparent lack of attention paid to the social consequences of test use. McNamara (2006) argued that Kane’s approach, while accepting the importance of consequences in principle, “does not provide a methodology for reflecting on them or investigating them” (p. 47). Even

Bachman's (2005) attempt to address this limitation, according to McNamara (2006), treats consequences too narrowly in the interests of making the validation process more manageable (p. 47). It is to the full social context of language testing, including the social consequences of test use, that the next major test theorists challenges us to properly acknowledge.

2.7 McNamara and the Social Context of Language Testing

McNamara's dissatisfaction with the level of validation apparent in language performance testing was clearly expressed in his first major publication:

The validity of second language performance assessments involve more than content- and criterion-related aspects of validity; the larger issue of construct validity has been insufficiently considered. In particular, empirical evidence in support of the claims concerning the validity of second language performance tests has in general been lacking. (McNamara, 1996, p. 7)

In many ways, McNamara's work has been an ongoing effort to redress this situation. While he lauded the efforts of many test theorists, especially Bachman whom he describes as a heroic 'Aeneas' in the field (McNamara, 2006, p. 35), he ultimately felt that they didn't go far enough to properly address the full social context of language performance and performance testing. He claimed that there was an implicit "commitment to working from within, at most questioning the usefulness of tests, and refraining from going further in examining the social implications of test use" (McNamara & Roever, 2006, p. 41).

McNamara has consistently and cogently argued for the need to more properly engage social theory in language testing. Doing this enables us to fully appreciate the complex social context involved in how tests are designed, what they purport to measure, and what is actually done with their results. His most recent book written with Carsten Roever, *Language Testing: The Social Dimension* (2006), aims to widen the social lens through which language test theorists and designers should be looking. The authors explicitly challenge the field to meaningfully engage social theory and accept the full implications that are entailed, as daunting as this task may be:

The problem ... is that we need an adequate social theory to frame the issues that we wish to investigate. The most productive and relevant of these theories are not only relatively unfamiliar to language testers but in fact challenge many of the fundamental epistemological and ontological assumptions of the field and this requires intellectual stamina and confidence to negotiate. (McNamara & Roever, 2006, p. 17)

For McNamara, there is a need to return to the challenge most clearly embodied in Samuel Messick's unified vision of validity. Attempts to disentangle the consequences of test use from the concept of validity, either through more practical concepts such as test usefulness or more reasonable approaches such as argument-based architectures, have not adequately addressed the reality that language testing is always a social act with social consequences, both influenced by and influencing the social context. For McNamara, perhaps Messick's most important legacy is not the inclusion of test consequences in his validity matrix, but the recognition that our values are intertwined in every aspect of the testing process and, as such, the constructs utilized reflect our socially-constructed values

and can never be taken for granted. For language testing, this “demise of the notion of the possibility of a value-free science, in this case psychometrics, means that values are seen as having a central role in determining the meaning of test-score interpretations” (McNamara, 2006, p. 40). This, perhaps more than any other development, has encouraged a greater consideration of the social consequences of language testing and has led to vehement calls for more ethical responsibility for test designers and increasing professionalization of the field.

2.8 Consequential Validity: A Role for Motivation?

In many ways, those who investigate the concept of validity stand near a crossroad initially reached over three decades ago, if not earlier. We can choose to focus exclusively on the test properties alone, ignoring the social environment that is influencing, and being influenced by, the wider social context. Alternatively, we can choose to take the road less traveled and fully embrace the wider social context of which we are an integral part while at the same accepting the full implications that this entails. To be sure, steps have been made in this direction, yet these have generally been tentative steps, and ones that have, for the most part, not significantly altered language testing in practice.

Can the social consequences of test use become an integral part of the test validation process? For those who hold a very strict traditional view of validity, (e.g., Wiley, 1991; Popham, 1997), the answer would likely be a resounding no. For these theorists, considerations of validity should be restricted to issues involving the accuracy of score based inferences (Popham, 1997, p. 10). Test validation that includes a

consideration of the consequences of test use entail complex questions that unnecessarily confound the validity concept and are best left for other sectors of society to deal with. Test designers should not be expected to carry out such extended analyses in order to assess whether their intended consequences are in fact realized nor should they be held responsible for any unintended consequences that may result from the use of their tests. However, for those who hold a more expansive view of validity (for example, Shepard, 1993, 1997; Moss, 1998), test designers have a duty to consider the wider social consequences involved in the use of their tests whenever and wherever possible and this needs to be integral part of the test validation process itself. In order to drive this agenda, it is suggested that tests must also demonstrate ‘consequential validity.’

The hope that guides this paper is that, even if one cannot accept the social consequences of test use as an integral component of validity, one would hopefully accept the view that evidence of the negative social consequences of a test, however obtained, should be of considerable importance to any responsible test developer. If this has the effect of changing testing practice for the better, then the objective is achieved. In an ideal world, test developers would take greater initiative in the investigation of social consequences, and assume full responsibility for the consequences of test use if for no other reason than reliance on other sectors of society to properly deal with these issues has proven to be largely ineffective. However, the social and political reality of our international marketplace precludes this, at least for the time being. In the meantime then, it falls on other concerned parties to investigate the social consequences of test use in the hope of positively affecting testing practice. In this vein, this paper aims to examine the potential social consequences of L2 high-stakes testing on young ELL L2 motivation. It

is strongly felt that young ELL L2 motivation, and more specifically, the potential impacts of high-stakes L2 testing, should play a much more significant role in any investigation of the social consequences of high-stakes testing. In order to shed light on this complex construct, the following chapter will trace the development of motivation in the field of psychology and its more recent application to second language research.

Chapter Three: Motivation Research

This chapter, much like the previous one, briefly sketches some of the major conceptual developments that have occurred over the last half century, this time concerning the equally slippery and contentious concept of motivation. Cognitive psychologist Bernard Weiner (1992) defines motivation as “why human and subhuman organisms think and behave as they do” (p. 1) and offers two opposing metaphors to be used as heuristic devices in the classification of various theories of motivation that have been promulgated over the last century or so. At one extreme, motivation theorists who have viewed humans as essentially *machines* have tended to emphasize our “inflexible behaviour, lack of volitional responding, predetermined reactions to output, use of energy, an aim or function, and coordinated subparts or a structure that is instrumental to a function” (p. 149). At the other extreme, theorists who have viewed humans as essentially *Godlike* have tended to emphasize our complete rationality and knowledge and assume that all humans “know all the goals available to them, know the likelihood of each goal, calculate how those likelihoods change depending on what caused success or failure, and then select the goal with the highest personal utility” (p. 288). In reality, most modern theories of motivation incorporate aspects from both metaphors, but this heuristic device serves as a constant reminder that our view of motivation reflects basic epistemological assumptions about the nature of our existence: these are crucial assumptions that should never be taken for granted.

At this point, I feel it is necessary for me reveal some of my own beliefs and assumptions regarding motivation. First, I believe motivation encompasses our beliefs,

attitudes, and self-conceptions and these are greatly influenced by our past experiences. However, motivation is also a continuous process that involves sustained action towards an objective and continual re-evaluation of past experiences. I have always viewed motivation as something that is influenced by our biological makeup, but is far more intensely influenced by the multitude of semiotic signs and symbols surrounding us from an early age, the opinions and pressures of those who are most significant in our lives, the types of goals we set for ourselves, and the reasons we attribute for success and failure. I believe our motivations change depending on the particular context we are in and according to the various stages of our life. However, our actions are not always logical or rational to others or even to ourselves. Indeed, we are not always aware of what motivates (or de-motivates) us, which makes the investigation of this complex concept exceedingly difficult. Finally, I have a strong belief that many motivational tendencies are developed in childhood and adolescence and become difficult (though certainly not impossible) to alter in adulthood. Thus, the investigation of childhood motivation is especially important to me as I believe it has direct and lasting effects on a child's adult life. As I continue to investigate the thoughts of motivational theorist in the fields of psychology and SLA, my own beliefs and assumptions continue to be re-shaped.

This chapter will begin with a cursory review of some of the more influential theories of motivation emanating from the field of psychology, from which L2 motivation researchers have regularly harvested their concepts. This will lead to a brief review of L2 motivation research, a field of inquiry whose origin has largely been attributed to the seminal research of Robert Gardner, Wallace Lambert and their associates in Canada (Dörnyei, 2001). More recently, Zoltán Dörnyei (2009), reviving

some of Gardner's original concepts as well as his quantitative-intensive approach, has developed a very ambitious 'L2 Motivational Self System' that, with the help of sophisticated statistical techniques, has demonstrated a degree of consistency across a number of international contexts.

This will be followed by a discussion of two issues pertinent to this thesis. First, the sometimes ambiguous role of young ELLs in L2 motivation research has led to some confusion about how concepts should be applied to learners of various ages. Secondly, there is an unfortunate paucity of attention paid in L2 motivation research concerning the impact of language testing, especially in test-intensive meritocracies (Ross, 2008). Even though there has been very compelling arguments made for the enormous power of tests on language learners (for particularly poignant examples, see Shohamy, 2001, pp. 7-14), the impact (or washback) of changes to national language tests on educational environments (e.g., Cheng, 2005) and the negative impact of high-stakes language tests on individual second language learners (Fox & Cheng, 2007; Fairbairn & Fox, 2009), most L2 motivation studies subsume language testing under some other construct or concept, or ignore its role entirely. Finally, the process that was used to develop the L2 motivational constructs that informed this study will be explained.

3.1 Motivation in Psychology

Investigation into motivation has largely mirrored the dominant epistemological assumptions prevalent at any given time. In the field of psychology, whose researchers have contributed immensely to our understanding of an individual's motivation, early concepts gradually shifted from an examination of predominantly biologically-based

drives inherent in us all (e.g., Freudian concepts) to more behavioural-mechanistic notions that view outside pressures as determining our motivation to act, a relationship represented most clearly in the work of B.F. Skinner (1965) and the precepts of behaviourism. These concepts largely view the individual as lacking the ability to control their motivation; remaining largely at the mercy of biological or external forces. More recently, this has given way to cognitive-mediational or constructivist perspectives marking the increased acceptance of the individual as a rational thinking being who sets goals and alters behavior in continual mediation with the social environment (Eccles, Wigfield & Schiefele, in Dörnyei, 2001, p. 19). These modern concepts of motivation, largely concerned with the cognitive mechanisms at play within an individual's mind, have had a significant influence on theorists investigating L2 motivation.

3.1.1 Achievement Motivation Theory

Few would question that there is some kind of innate desire in all humans to accomplish certain tasks. When we see a baby crawling on the floor attempting to reach some distant object, or a child struggling to take those very first steps, it is hard not to be impressed by what seems to be a very natural impulse to achieve something. Many cognitive-based theories of motivation assume we all are 'innately active' and have an "inborn curiosity and an urge to get to know [our] environment" (Dörnyei, 2001, p. 20) and thus view the fact that we are motivated to act as essentially unproblematic. What concerns many researchers who operate within an 'expectancy-value' framework are the mechanisms involved in directing and shaping our inherent desires. Achievement Motivation Theory (AMT), originally articulated by Atkinson (Atkinson, 1964; Atkinson

& Raynor, 1974), was a fairly 'early' attempt to explicate the major forces influencing our motivation to achieve tasks. Four inter-connected components of motivation were theorized, consisting of 1) expectancies of success (what are my chances for success?), 2) incentive values (what will I get if I succeed?), 3) need for achievement (do I really need to succeed?), and 4) fear of failure (what will happen if I don't succeed?). According to AMT, the sum total of these four components work together to either encourage or discourage our achievement-oriented behaviour. This theory provided a persuasive framework for understanding motivation and would remain dominant for decades. Despite the many developments that have occurred in motivation research up until the present day, the influence of various components of this theory can still be seen in many modern conceptions.

3.1.2 Goal-Setting and Goal-Orientation Theory

Goal-setting has generally been considered an important component of motivational behaviour, and for Locke and Latham (1990) it is in fact absolutely necessary since "all human action is caused by purpose, and, for action to take place, goals have to be set and pursued by choice" (Dörnyei, 2001, p. 25). According to this theory, the goals we set have the most positive effect on our motivations when we believe the goal is attainable and important, and our performance generally improves when the task to be completed is both specific and difficult, vaguely-defined distal goals being far less motivational. Elaborating on these relationships, Goal-Orientation Theory, developed explicitly to better understand the type of learning that happens in school settings (Dörnyei, 2001, p. 27), puts forth two general contrasting rationales that underlie learning

behaviour and performance judgements. Those who exhibit a *mastery orientation* attempt to learn content for its own sake without much regard for the external rewards that may, or may not, come along with achievement. Their long-term objectives are not as adversely affected by short-term failures because their judgements are largely self-referenced (Ames, 1992, p. 262). Conversely, those who exhibit a *performance orientation* are far more concerned with short-term, norm-referenced achievements that can be publicly recognized and can, in turn, increase one's self-worth. Ames, echoing other researchers utilizing this concept, has argued that a mastery goal orientation encourages a "motivational pattern likely to promote long-term and high-quality involvement in learning" (ibid., p. 263). Due to these apparent benefits, these concepts have been influential in attempts to transform pedagogy to encourage more intrinsically-inspired and autonomous learner styles.

3.1.3 Attribution Theory

For those who believe we are guided by our accumulated experiences and our rational assessment of our various successes and failures, motivation is a process that continually evolves throughout our lives. Thus, at various points in our life we 'take stock' of our past experiences, whether consciously or unconsciously, and assign reasons for why things happened the way they did. These reasons (i.e., attributes) become a central concern for Attribution Theory. According to proponents of this theory, often associated with the work of Bernard Weiner (1992), our beliefs about the causes of our past failures play a crucial role in shaping our future behaviour. Of particular importance is the degree of control we believe we possess over the cause, a perception that will

greatly influence our decision to engage in similar behaviour in the future. According to Graham (in Dörnyei, 2001, p. 22), common attributions in a classroom environment can include ability, effort, task difficulty, luck, mood, family background, and the help or hindrance from others. If the attribute assigned for failure is ‘stable and uncontrollable,’ as in the attribute of *ability* (e.g., I studied hard, but I simply can’t do it), the greater likelihood we will avoid similar tasks in the future. However, if the attribute we assign is something under our control and unpredictable, as in the attribute of *effort* (e.g., I didn’t study hard because I was sick), then the negative impact of failure will be minimized and we will be more likely to engage in the task again. Implicit in this theory is the assumption that individuals make rational decisions based on their perceptions of cause and effect. It does not necessarily matter if the cause-effect relationship is actually true, only that the individual subjectively *believes* it to be. A similar theory that bases many of its precepts on the power of our subjective beliefs is the concept of self-efficacy.

3.1.4 Self-Efficacy

Largely associated with the work of Bandura (1997), self-efficacy concerns “people’s beliefs about their capabilities to exercise control over their own level of functioning and over events that affect their lives” (Bandura, in Dörnyei, 2001, p. 23). Those who maintain a strong belief in their ability display more effort and deliberation in learning tasks and are more likely to remain on task in the face of obstacles (Zimmerman & Schunk, 2008; Gorsuch, 2009). However, self-efficacy is not a trait that simply exists in some but not in others; rather, it is developed through an individual’s interactions within their immediate social context. Bandura’s social cognitive theory holds that we are

all ultimately “self-organizing, pro-active, self-reflective, and self-regulating” (Pajares, in Gorsuch, 2009, p.509) and as such, we make judgements of our various abilities in order to make decisions about our future course of action. These beliefs do not arise out of ‘thin air’, but instead are based on our experiences in specific domains and on specific tasks (Gorsuch, 2009, p. 509). Self-efficacy is not simply an emotional response to a learning obstacle, but a reflective judgement about one’s ability to perform a task whereby “cognition is invoked and actively used to plan and regulate and reflect upon participation in learning opportunities” (ibid., p. 509).

3.1.5 Self-Worth Theory

Covington’s (1992) Self-Worth Theory argues that the desire to have a positive self-image is a basic need that all individuals possess. It is fundamental to motivational behaviour, particularly in school settings, that individuals maintain a positive image of their worth to themselves and to others around them. In social situations where the opinions of others can directly threaten our self worth, we may reduce our effort because it would be far less threatening to acknowledge that our failure is due to a lack of concern or effort than due to our lack of ability. This may explain the tendency of some students in school to show little to no interest in their performances. This may happen because in many schools “self-acceptance comes to depend on one’s ability to achieve competitively” which is a direct result of a pervasive tendency in society to “equate accomplishment with human value” (Covington, in Dörnyei, 2001, p. 24). Thus, the importance of maintaining one’s own self-worth becomes paramount to educational achievement.

3.1.6 Self-Determination Theory

Starting with the assumption that goals become ‘psychologically potent’ because they reflect our needs, self-determination theory (SDT) explores “the degree to which people are able to satisfy their basic psychological needs as they pursue and attain their valued outcomes” (Deci & Ryan, 2000, p. 227). Three essential needs that improve the extent to which we are ‘self-determined’ are *competence* (a feeling we can do something well), *relatedness* (a feeling of connectedness to others), and *autonomy* (a feeling that we are in control of what we do). Essential in our attempt to satisfy our basic needs is the degree to which we value goals intrinsically (because we choose and want to do it) or extrinsically (because others want us to do it).

However, there has been some dissatisfaction expressed about the ability of the intrinsic/extrinsic dichotomy so dominant in motivation research to capture the complexity involved in being self-determined (Dörnyei, 2001). It has been argued that extrinsic motivation does not always work against its more positively-viewed intrinsic counterpart. Deci and Ryan (1985) instead proposed a more intricate model for extrinsic motivation that theorized degrees of regulation placed on a continuum whereby at one extreme, regulations were completely determined and controlled by others, and on the other they were completely internalized and at harmony with the individual’s values, needs and identity. Deci and Ryan argued that extrinsic rewards, when sufficiently internalized, could even become part of our intrinsic motivation (Dörnyei, 2001, p. 29). SDT, sharing some affinity with Vygotskian sociocultural conceptions of learning and identity formation now quite in vogue in educational research, represented a significant

departure from more traditional cognitive approaches in psychology that often ignored the intimate influence that society has in the shaping of motivation. However, there would be others that would take this even a step further.

3.1.7 Weiner's Social Motivation

The great majority of motivational theories that have been explored in the field of psychology have been individualistic in their approach. Within these motivational frameworks, whether developed within motivational psychology or social psychology, the focus has largely been on the individual cognitive processes involved with only scant attention paid to the wider social context in any real sense. More recently, motivation in psychology has taken a decidedly 'social turn' as many researchers have begun to incorporate the sociocultural context into their concepts signifying a "growing recognition that all these environmental dimensions have a certain amount of influence on one's cognition, behaviour, and achievement" (Dörnyei, 2001, p. 30). One such attempt to incorporate the sociocultural context has been put forth by Weiner (1994) in theorizing the existence of both 'social' motivation, which requires the psychological presence of significant others, and 'personal' motivation which does not (p. 557). Although he acknowledges that these two concepts can be very difficult to separate in practice, they are useful theoretical constructs for analytical research into motivation. This has been only one example in psychology to place more emphasis on "the study of motivation that stems from the sociocultural context rather than from the individual" (Dörnyei, 2001, p. 30). It remains to be seen how able and willing the field of psychology is to incorporate the wider social context into its core concepts. While the preceding

review has only been a brief review of some of the major motivational concepts explored in psychology, it will become apparent in the next section how influential they have been, and continue to be, in the exploration of L2 motivation.

3.2 Motivation in SLA

3.2.1 Gardner and Lambert's 'Integrativeness'

The expansion of motivation research into the field of second language acquisition has largely been attributed to the pioneering work of Canadian researcher Robert Gardner and his colleagues around the start of the 1960s (Dörnyei, 2001). Gardner and his colleagues attempted to tackle the concept of motivation as it relates to the acquisition of a second language. However, the role of context would become far more influential for these theorists as they argued that the process involved in learning a L2 was not the same as other classroom subjects. The L2 learner is “faced with the task of not simply learning new information ... which is part of his *own* culture but rather of *acquiring* symbolic elements of a different ethnolinguistic community” (Gardner, in Dörnyei, 2001, p. 47, italics in original). This being the case, a social-psychological approach, recognizing the influence of the social context in the development of learner orientations, would play a far greater role in L2 motivation studies right from the start. Although this early model was still largely an individualistic cognitive-based approach, the effects of environmental variables on L2 motivation would become much more central to the investigation.

Gardner and Lambert's (1972) original model of L2 motivation proposed two *orientations* that language learners generally possessed: *integrative* or *instrumental*. Questionnaires were developed which purported to measure the level of integrative motivation a learner possessed, defined in 1959 as the "willingness to be like valued members of the target community" (Gardner & Lambert, in Dörnyei, 2001, p. 49) or alternatively, their level of instrumental motivation, representing more pragmatic concerns to enhance opportunities to obtain various external rewards. In many ways, these concepts appeared to mimic the intrinsic/extrinsic dichotomy so prevalent in psychology; however, the explication of these concepts would become far more complex and intricate, involving highly sophisticated quantitative techniques that would continue to reshape the model for decades to come (see Dörnyei, 2001; 2009, for a fuller discussion). Ultimately, these concepts would develop into what Gardner (1985) coined his 'socioeducational model.' However, the 90's would witness increasing calls for alternative concepts and methodological approaches that could be more properly applied in the classroom and in an increasingly globalized educational context (Dörnyei, 2001).

3.2.2 1990 to the Present Day: Rethinking and Revisualizing L2 Motivation

Gardner and Lambert's has had a 'heavy influence' on L2 motivation research, at least until the early 1990s (Crookes & Schmidt, 1991, p. 471). Indeed, many studies until the 90's continued to explore the statistical strength of early L2 motivational constructs, as well as newly-theorized ones, that largely utilized quite similar quantitative techniques (heavily influenced by factor analysis). Despite the social-psychological approach being used, these abstract concepts of motivation, far removed from the language learning

situations which spawned them, dissatisfied a growing number of theorists, researchers, and practitioners in their attempts to understand L2 motivation in increasingly complex and varied learning contexts.

This would lead some researchers to call for more ‘education friendly’ L2 motivational constructs that more properly considered the educational context involved in shaping L2 motivation (Dörnyei, 2001). For many, there was a need for greater recognition of the “immediate instructional context” and how decisions about educational tasks, evaluations and rewards influence motivation (Stipek, 1996). A pioneering article by Crookes and Schmidt (1991) called for more explicit investigation of the learning environment. This was necessary because of what Au (1988) referred to as the ‘causality hypothesis’ characteristic of the socioeducational model. Drawing support from a number of other authors (e.g., Strong, 1984), they questioned the non-problematized causative link made between attitude and motivation, noting that “achievement might actually be the cause instead of the effect of attitude” (Crookes & Schmidt, 1991, p.474). At best, attitudes represent ‘distal factors’ that do not necessarily relate directly to motivation in the strict sense, much less sustained motivational behaviour.

There was also increasing dissatisfaction expressed over the applicability of the concept of ‘integrativeness’ beyond the context in which it was conceptualized. For many, the concept was an ‘enigma’ that ‘simply did not make much sense’ in many other language learning contexts (Dörnyei, 2009, p. 23). A number of researchers expressed doubts about the relevance of the concept in contexts where contact with members of the L2 culture was far less frequent and meaningful (e.g., Warden & Lin, 2000; Yashima, 2002). Even in contexts where contact with the target culture was regularly possible, it

was felt that ‘integrativeness’ failed to capture the lived experiences of many L2 learners attempting to cope, linguistically and culturally, with their new environments. A deeper contextual understanding of L2 motivation was necessary that addressed the effect (and affect) of power, conflict, and identity in shaping the level of commitment L2 learners were able or willing to ‘invest’ (Norton, 2000).

Since these criticisms have been expressed, the concept of integrativeness has “played a rapidly diminishing role in L2 motivational research” (Dörnyei, 2009, p. 24). In its place, there have been a growing number of studies that have explored alternative motivational concepts utilizing more qualitative approaches to the study of motivation in an attempt to provide more robust, emergent, motivational concepts (e.g., Williams & Burden, 1999; Norton, 2000). However, there have also been a number of other studies that have continued along the original social-psychological, quantitative path while attempting to adapt previous models (e.g., Noels, Pelletier, Clément & Vallerand, 2000; Dörnyei, 2009). The approaches and methods now available for modern L2 motivation researchers are certainly eclectic. Navigating these complex and sometimes contradictory motivational concepts is a formidable task for anyone approaching the field for the first time. Yet, they also provide a number of strategic vantage points from which to view the concept of motivation, which when considered together provide much greater insight than was ever before possible.

3.2.3 Dörnyei’s L2 Motivational Self-System

Zoltán Dörnyei, one of the leading L2 motivational researchers over the last few decades, has developed a model of L2 motivation that borrows from concepts developed

in mainstream psychology, Gardnerian concepts, as well as other L2 motivational concepts that have been developed over the course of six decades. Attempting to overcome some of the criticisms that have been levelled against Gardner's socioeducational model, while not throwing out the 'conceptual baby' with the bathwater (MacIntyre, MacKinnon & Clément, 2009), Dörnyei shed the integrativeness/instrumental label and in its place, has proposed concepts inspired by the burgeoning field of self-research in psychology. The subfield of 'possible selves' and 'future self-guides' offered an intriguing alternative for Dörnyei in the exploration of L2 motivation while allowing him to sidestep the theoretical baggage associated with the concept of 'integrativeness' (Dörnyei, 2009).

Borrowing on the work of Higgins (1987) and Markus and Nurius (1987), and more specifically on their exploration of 'future self' imagery in motivational behaviour, Dörnyei utilizes the concepts of 'ideal self' and 'ought-to self' because they can be applied more meaningfully to a variety of language learners in a multitude of learning contexts. These 'self states', it has been proposed, are experienced by language learners as reality (Dörnyei, 2009, p. 16) and play a significant role in sustained motivational behaviour. According to the explanations of these concepts originally proposed by Higgins (1987), 'ideal self' refers to the characteristics we would like to possess in the future while the 'ought-to self' generally represents the characteristics that significant others would like us to have. Not to be directly equated with the intrinsic/extrinsic dichotomy, as was also the case with SDT (Deci & Ryan, 1985, see above), these concepts were not theorized as necessarily being in conflict or being mutually exclusive. Instead, there can be, and often is, significant overlap in these concepts. Since, we are

greatly influenced by our external context, this can cause what may have been an ‘ought-to’ conception initially, to become part of our ‘ideal’ self, in effect removing any potential conflict (Dörnyei, 2009, p. 14).

The degree to which future self-guides are in harmony, or are ‘additive,’ can have a significant impact on the motivational behaviour of individuals. Even feared possible selves can have positive motivational impact if they are effectively counterbalanced with more desirable future selves. To maximize their positive motivational impact, our future possible selves also need to be activated, plausible, and reasonably vivid, something that clearly does not exist for all language learners. While our ‘ideal L2’ and ‘ought-to L2’ selves constitute two major components of Dörnyei’s model, another component, the ‘L2 learning experience’ completes the picture. This was added to “recognize the motivational impact of the main components of the classroom learning situation” (Dörnyei, 2009, p. 29), including the impact of the teacher, the curriculum, the peer group, and the experience of success; however, it has been left to future research to investigate the “self aspects involved in this bottom up process” (ibid., p. 29).

These concepts have been utilized in a series of studies, which Dörnyei (2009) has confidently noted, have been conducted in five different countries with over 6000 participants, the results of which have provided “solid confirmation for the proposed self system” (p. 31). The research reported in this thesis will draw heavily on the concepts developed by Dörnyei and his colleagues, more specifically a Dörnyei-inspired model employed in a study of three Asian contexts (Japan, China, and Iran) involving nearly 5000 participants (Taguchi, Magid & Papi, 2009). However before turning to a description of this model, it is necessary to address two pertinent issues that have not

received significant attention within L2 motivation studies, an oversight that will hopefully change in the near future.

3.2.4 A Greater Appreciation of Young ELLs in L2 Motivation

It has not always been clear in L2 motivation research whether the theoretical concepts under study are expected to operate differently depending on the ages or cognitive levels of individuals. While cognitive levels are certainly part and parcel of most theories of learning (e.g., Piaget's theory of cognitive development or Vygotsky's cultural-historical psychology), the major theories developed in motivational psychology and L2 motivation offer little guidance in terms of how these theories could apply to young learners, if indeed they should. In one study that more explicitly investigated changes in motivation as students progress from elementary school to middle school, Harter, Whitesell and Kowalski (1992) discovered that changes in perceived competence was associated with intrinsic motivation. Those students who felt their competence decreased after the transition were more likely to experience a reduction in their intrinsic motivation. The authors speculate that this could be a result of an increased focus on external evaluations that foster more objectified self-evaluations based on social comparison. This may have the effect of bolstering "the perceived competence of the most competent students while potentially undermining that of the less competent" (p. 801). A few studies which have specifically investigated age effects with respect to L2 motivation are also worth a closer look.

Williams and Burden (1999) examined the attributions of success and failure expressed by young learners of French in the U.K. Their small-scale grounded approach

found that the youngest learners involved in their study (10-12 years old) tended to rely heavily on verbal teacher feedback in assessing their achievement and progress; however, students in secondary school (13-15 years old) tended to rely much more heavily on feedback in the form of marks, grades or test results. They found support for the notion that as young ELLs matured, they expressed greater variability in their attributions for success and failure, including more external factors. Secondary school participants appeared to possess more clear perceptions of ability based on their external scores. They concluded that learner attributions were highly affected by their social environment and argued that when “the emphasis in the classroom is on achieving high marks, then marks, rather than an internal sense of the development of linguistic or communicative competence, will become the benchmark of success or failure” (Williams & Burden, 1999, p. 199).

Kormos and Csizer (2008) investigated Dörnyei-inspired concepts across three different education groups (secondary school, university, and adult learners) and discovered lower mean values for an idealized self amongst secondary school students that “might be explained with reference to the fact that students’ self-image goes through considerable changes in the period of adolescence ... and therefore their Ideal L2 Self is also under transformation at this age” (p. 346). For adults, the self-image is far more stable. The relationships discovered across these educational groups lead the researchers to conclude that “theories of L2 motivation do not only have to take into account the setting in which students acquire the language but also the age of the learners” (p. 349).

It would appear that we need to be extremely sensitive in applying motivational concepts to learners of various ages. While it does add a level of complication to any

analysis, it is most certainly a necessary one, and may in fact be the key to understanding L2 motivational processes over time. Since this research is primarily concerned with the study of young ELLs aged 12-15, there is need to be cautious when interpreting any result on the basis of theoretical constructs which may operate differently for older learners than younger ones.

3.2.5 A Greater Role for Language Testing in L2 Motivation Research

To close out this section, it is also worth considering a more specific role for language testing in L2 motivation research. In most studies, if language testing is considered at all, it is most often subsumed under a more general category of feedback, educational environment, attitudes to the learning environment, or some other more general construct. However, in testing-intensive environments, often characteristic of Asian educational systems (Ross, 2008), it is worth considering whether this really does justice to the potential motivational effect that language testing has on learners, especially younger ones. In one study that more explicitly investigated the influences of language testing on 626 Japanese high school students, Sakai and Kikuchi (2009) concluded that teacher competence, teaching style, and school facilities were far less influential than learning content and test scores in demotivating students. Given the significant power that tests seem to hold over language learners and the language learning environment (Shohamy, 2001; Cheng, 2005) it is unfortunate that there is still a dearth of L2 motivational research that more explicitly investigates the potential impact of language tests on L2 motivation.

Language tests, as part of the overall assessment of ELL language ability, progress, and performance, undoubtedly play a pivotal role in the overall learning experience of ELLs (Fairbairn & Fox, 2009). High-stakes L2 testing ‘episodes’ are likely to be somewhat more traumatic for some than others (Fox & Cheng, 2007) and perhaps more dramatic for young ELLs than adult learners (Haggerty, in press). These are issues which require greater attention in L2 motivational studies in the future. This thesis hopes to begin to redress this imbalance by: 1) establishing L2 motivational factors to better explore L2 attitudes, beliefs and self-conceptions of young language learners utilizing previously-researched L2 motivational constructs, and 2) investigating differences in these L2 motivational factors on the basis of six learner-background variables, including three that specifically address their L2 testing experience. Hopefully, this will inspire further research into the potential effects of L2 testing on L2 motivation. It is now necessary to explain the procedure used to establish the theoretical L2 motivational constructs that guided this study.

3.3 Development of L2 Motivational Constructs

One of the main objectives of this thesis is to examine any significant differences that exist in the L2 motivation of young ELLs on the basis of learner-background characteristics, most importantly, their L2 testing experiences. In order to develop potential constructs that could prove fruitful in this context, it was decided to incorporate what has already been discovered in other contexts that share some of the same educational characteristics. A large-scale study has investigated a number of L2 motivational concepts in three Asian countries: Japan, China and Iran (Taguchi, Magid &

Papi, 2009, hereafter referred to as the 'Taguchi study'). While predominantly focused on adult learners (N= 3420), the Taguchi study also included middle school learners in China (N=244). This study is of relevance to Korea because of some of the educational similarities that both Japan and China share with South Korea, in particular their exam-oriented approach to education (Taguchi et. al., 2009, p. 69). In South Korea, the assessment of English for university entrance purposes has a significant impact on what secondary school teachers think and do (Li, 1998). In the case of China, there is also an additional similarity with regards to the high importance placed on English in which 'knowing English has become a must for anyone who wants to compete in a global marketplace' (Taguchi et al., 2009, p. 69). For these reasons, the theoretical constructs developed in the Taguchi study, particularly those investigated in the Chinese context, directly informed most of the theoretical constructs developed for this investigation. A brief explanation of all the constructs used for this study is therefore necessary before moving on to an explanation of the method utilized and the results obtained.

3.3.1 Dörnyei -Inspired Constructs

The theoretical constructs utilized for this investigations (8 of 10) were modelled on questionnaire items included as part of a large scale study of L2 motivation in Japan, China and Iran (i.e., the Taguchi study). In that study, Structural Equation Modelling (SEM) was utilized to assess the appropriateness of Dörnyei's L2 Motivational Self-System in EFL contexts other than Hungary (where much of the research on Dörnyei's system had been conducted). They concluded that the Hungarian findings demonstrated 'external validity' having demonstrated similar patterns in three contexts that were

'strikingly different' from the original research context (Taguchi et. al., 2009, p. 68). They also concluded that their analysis "confirmed the validity of the entire tripartite L2 Motivational Self System" and helped them to "understand certain cross-cultural differences in different educational contexts" (ibid., p. 88). They also found support for dividing the construct of 'instrumentality' (understood generally as our desire to do things because of the effects of our external environment) into two distinct types, *promotional* and *preventional*, which ultimately depended on the "extent of internalisation of external incentives" (ibid. p. 88), a result appearing to support Deci & Ryan's (1985) more complex approach to intrinsic/extrinsic motivation discussed earlier (see Ch. 3). Overall, their findings strongly supported the utility of the L2 Motivational Self-System constructs in understanding L2 motivation in a variety of EFL contexts.

For the purposes of this study, 8 of the 10 theoretical constructs that were used to model questionnaire items were directly inspired by the Taguchi study. In their study, these constructs obtained alpha scores of .70 or greater (see Table 1 below) indicating adequate internal consistency. The authors also provided the exact wording for all the questionnaire items they used to investigate these constructs in each context. This represented a fairly extensive pool of constructs and items. In order to maintain as much consistency as possible, I decided to focus solely on the questionnaire items used in the Chinese context because, as previously stated, I believed this context shared a number of similarities to the Korean context. Table 1 below compares the total number of questionnaire items used for each construct in the Taguchi study and the total number of items used for the present study (see Appendix A for a list of the Taguchi questionnaire items that were excluded from the present study).

Table 1: Theoretical Constructs Utilized for Present Study

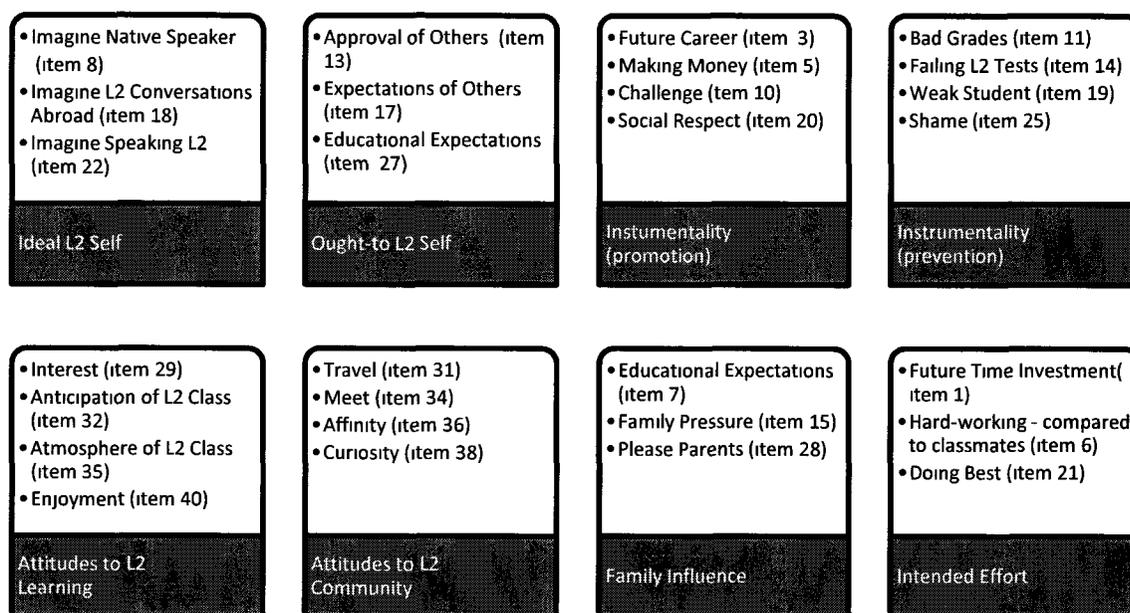
Factor Name	Taguchi Study (China)		Present Study (Korea)
	# of Items	α	# of Items
<i>Taguchi Study Constructs*</i>			
1. Ideal L2 Self	5	0.83	3
2. Ought-to Self	7	0.78	3
3. Instrumentality (prevention)	5	0.84	4
4. Instrumentality (promotion)	8	0.78	4
5. Attitudes to L2 Learning	4	0.81	4
6. Attitudes to L2 Community	4	0.76	4
7. Family Influence	5	0.70	3
8. Intended Effort	6	0.75	3
TOTAL	44		28
<i>Constructs from Pilot Study **</i>			
9. Attitudes to L2 Testing			8
10. L2 Social Value			4
TOTAL			40

* Taguchi, Magid & Papi, 2009

**Haggerty (in press).

Figure 2 below provides a graphical representation of the eight Taguchi study constructs used in this study (for the complete wording of items, please refer the questionnaire in Appendix B). If the purpose of this study had been to replicate the results of the Chinese study, it would have been preferable to include all the constructs and variables used. Instead, I decided to sample from the Taguchi study for two main reasons: 1) the constructs had been previously tested and were found to be reliable in the Chinese context, and 2) there was a need to keep the questionnaire fairly short due to concerns about the ability of middle school students to maintain focus (a problem experienced in

the pilot study conducted for this thesis). Decisions about which items to exclude from the analysis were based on their potential inapplicability to young learners, their potential unintelligibility to young learners, or their perceived repetitiveness.



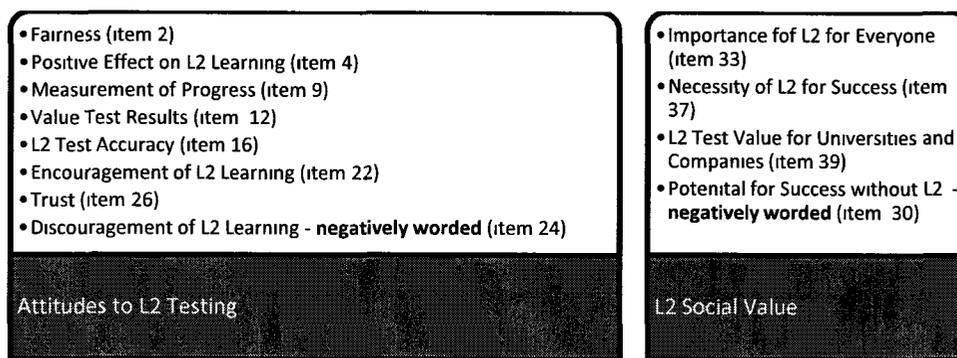
Source Taguchi, Magid, & Papi, 2009

Figure 2: Sampled Items from Taguchi Study in Chinese Context

3.3.2 Pilot Study Constructs

Two additional constructs, ‘Attitudes to L2 Testing’ and ‘L2 Social Value’ were added based on a pilot study of testing and learning attitudes conducted in preparation for this thesis (see Haggerty, in press). I included these theoretical constructs in order to further investigate their explanatory potential in this context as well as explore their relationships to other more established constructs. A total of 8 items were included for the construct ‘Attitudes to L2 Testing’ because this was an important construct for the purposes of this thesis and had only been pre-tested once on a fairly small sample

(N=65). The construct ‘L2 Social Value’ included four items that demonstrated strong internal consistency in the pilot study ($\alpha = .91$) and were thought to hold potential relevance in better understanding ELL motivation. Figure 3 below provides a graphical representation of the two pilot study constructs used in this study (for the complete wording of items, please refer the questionnaire in Appendix B).



Source Haggerty (in press)

Figure 3: Sampled Items from Pilot Study in Korean Context

3.4 Learner-background Characteristics (Variables)

In order to investigate any differences that may exist in L2 motivational beliefs, attitudes, and self-perceptions among groups of participants, the following learner-background characteristics were included. A total of six variables were considered by this study. The rationale for the inclusion of these learner characteristics follows.

3.4.1 Gender

The effects of gender, particularly in adolescence, have frequently been discovered in motivation studies conducted in the field of psychology. For example, in an

investigation of the 105 adolescents (12-15 years old) enrolled in a summer enrichment program in Canada, Pychyl, Coplan, and Reid (2002) discovered that only the female students' (as a group) were significantly influenced by the differing parenting styles of their mothers and fathers. It was suggested that "fathers have a more direct effect on their daughters' development of procrastination, while the mother's effect is mediated through the self-system" (p. 282). In the area of L2 motivation studies, noticeable gender effects have also been discovered. For example, Kissau (2006) in a study of 490 grade 9 students in Ontario, Canada, discovered consistent gender effects across a number of L2 motivational variables and concluded that many of these were influenced by society. Henry (2009) in a study of L2 motivation among 169 Swedish L2 learners found that girls L2 self-concepts appeared to strengthen over time while the boys appeared to weaken. Based on the results of studies such as these, it was deemed necessary to consider the effects of gender in this study. I had no pre-conceived idea of how this variable might be associated with L2 motivational factors based on my experience in the educational setting.

3.4.2 Education Level

In addition to gender, some L2 studies have also found that motivation levels can differ depending on the age or education level of learners. Williams & Burden (1999) in a small-scale study of British young learners of French discovered that there was a greater range of attributions given for success or failure as the learner increased in age and concluded that the teacher, the classroom environment, and the wider educational context all play a role in this process. As discussed earlier, Kormos & Csizer (2008) in their study

of 623 Hungarian students discovered that motivational “factors play a different role in L2 motivation at different ages” and some factors may not “even meaningful in a particular setting or for a specific age group” (p. 350). Based on my personal experience in the setting, middle school students in their third year are under a great deal of pressure to perform in all their subjects. This is because their test results are a vital component in the admission decisions of most of the more prestigious high schools in the country. For these reasons, I suspected that education level might have some influence on L2 motivational factors and therefore decided to include this variable in the analysis.

3.4.3 Self-reported L2 Proficiency

In the motivation literature, there has been a great deal of investigation into how motivational attributes affect the level of L2 proficiency ultimately achieved by language learners (for a more recent example, see Sparks, Patton, Ganschow & Humbach, 2009). While this is an important educational issue, it is equally instructive to investigate how ELL *perceptions* of proficiency play a symbiotic role in shaping and re-shaping motivation. A recent study in the Asian context by Law (2009) has found that ELL beliefs about ability and intelligence play a significant role in their reading comprehension. Students who held a belief that effort and ability were essentially fixed and not under their control performed more poorly than students who did not make such a strong connection. In another study conducted by Dörnyei and Kormos (2000) it was found that ‘linguistic self-confidence,’ which included favourable conceptions of language aptitude, progress, and a belief in success, was significantly correlated with improved oral performance. For this study, this variable was included to explore its

association with L2 motivational factors. It was hypothesized that ELLs who had higher perceptions of L2 proficiency would also have more positive L2 attitudes, beliefs, and self-conceptions.

3.4.4 L2 Test Preparation

Test-taker strategies have been investigated for well over two decades (Cohen, 1984), have added useful approaches and techniques in the investigation of test-taker processes such as verbal reports and think aloud protocols (Anderson, 1991), which have proven their utility in ongoing test validation processes (Cohen, 2007). Yet, the amount of time and effort expended in test preparation has not received a great deal of attention. In contexts where L2 testing is omnipresent, such as is the case in many Asian contexts, the amount of time spent preparing for tests can be extensive. This could have a dramatic influence on various aspects of L2 motivation. Therefore, I decided to include this variable as well.

3.4.5 Level and Frequency of L2 Test Completion

There are no studies that I am aware of that have explicitly investigated the level of L2 test completed by young ELLs and the effects this may have on L2 motivation. This variable was included because it was believed by me that middle school students who had completed a higher level test (i.e. one intended for university admission purposes) would have improved L2 attitudes, beliefs and self-conceptions. This was based on my own experience within the educational context and further suggested by the results of the pilot study conducted in preparation for this thesis (Haggerty, in press). In

addition, I felt that the number of times (or frequency of) university-entrance level L2 tests completed could also influence L2 motivational factors. Students may experience a shift in their motivation as they participate in more high-stakes testing events. I did not have a clear pre-conceived idea of this influence.

These six variables, along with the ten L2 motivational constructs described above, provided the basis for addressing the research questions for this study. To reiterate, these questions included:

Are L2 motivational factors influenced by:

- gender
- educational level?
- self-reported L2 proficiency?
- time spent preparing for L2 tests?
- level of L2 test completion (middle school or university-entrance)
- frequency of university-entrance level test completion

Now that the constructs and variables have been explained, this thesis can turn to the method used to collect the data and analyze the results.

Chapter Four: Research Method

In order to address the research questions set for this paper, the following research method was followed. Since this research utilized an exploratory factor analysis (EFA) approach, followed up by factor score comparisons among groups of respondents, the data analysis section will provide the theoretical rationale for identifying the factors and the calculation of factor scores.

It should be noted that the method chosen was a strictly quantitative approach to the investigation of L2 motivation. Questionnaire research offers the advantage of collecting a great deal of data in a relatively short time. It also gives a researcher the ability to understand general patterns in a large group of participants and is thus very suitable for an exploratory study of a specific context. Of course, a qualitative study would have provided a richer and more detailed account of the individual factors influencing ELL L2 motivation and was considered in the design phase of this research. However, it became too unwieldy for me to consider in this context. Due to a conflict between university ethics requirements that require the signature of every parent, and the reluctance of all the parents I approached to commit their children to such intimate and time-consuming research, a qualitative approach was not feasible for this study (whether on its own or as part of a mixed-method approach). Hopefully, future researchers with much greater access to the educational context than I had will be able to add a more qualitative understanding of the young ELL L2 motivational patterns discovered in this study.

4.1 Participants

The participants for this study consisted of a convenience sample of middle school students (N= 341) enrolled in an English private language school in Seoul, South Korea. The ages of the participants ranged from 12 to 15 with a mean age of 14.3 years. A total of 141 (41.3%) were male and 200 (58.7%) were female. All participants indicated that they were currently in middle school. Of these, 59 (17%) were in their first year of middle school, 60 (18%) were in their second year and 222 (65%) were in their third and final year of middle school. Students were asked to complete questionnaires at their school during class time.

4.2 Setting

The setting was a private language school (also referred to as an academy or '*Hagwon*' in South Korea) to which parents pay an education fee to enrol their children for specialized language study. English private language schools specialize in preparing students for various English tests as well as developing general English proficiency skills. Students generally attend these schools for two to four hours a day, two to five times a week after their regular school hours. Class sizes tend to be much smaller than most Korean public schools. These schools are very common in South Korea as many can be spotted throughout every large city in the country. It is not known what percentage of Korean middle school students attend a private language school, so it is possible that the students involved in this study were more privileged than average middle school students (although the socioeconomic area was described by the principal and a number of teachers as average). As a result, the results of this study can only be applied to this

setting and cannot be generalized to the wider population. However, there is also no compelling reason to suspect that these students were significantly different than other middle school students in Korea.

4.3 Instrument

The instrument used for this study was a questionnaire that, in addition to the collecting information about participants' gender, grade, L2 proficiency, L2 test preparation, level of L2 test completion, and frequency of L2 test completion, included a total of 40 questionnaire items (28 statements and 12 questions) that elicited participants' opinion based on a 6-point Likert scale (1= strongly disagree to 6=strongly agree for statements; 1=not at all to 6=very much for questions). The questionnaire as administered can be viewed in the appendix (see Appendix B). All questionnaire items were provided in Korean first and then immediately followed by English. This was done as it was felt that most students would feel more comfortable reading questionnaire items in their native language. All English questionnaire items were translated into Korean once and then 'back-translated' (Brislin, in Dörnyei & Taguchi, 2010, p.51) into English. Then, a team-based approach was utilized to assess the suitability of the translated questions (ibid., p.50). Two native-Korean teachers and two native-Canadian teachers (including myself) were used in this process. The focus was to maximize the potential of each question to be understood by middle school students. Some items were re-translated as a result of this process. 28 of the English-worded items came directly from the Chinese questionnaire used in the Taguchi study (see Table 1) and the other 12 items came from

the pilot study as described in the previous section. Two items were negatively worded in order to check for consistency in participant responses.

4.4 Procedure

Permission for this research was initially sought from three private language schools in the Seoul area. A letter of information was sent to the principals and academic coordinators of these academies fully explaining the nature of the research. Having received a positive response from two of these academies, I arranged an interview with each school (the principal and a few teachers) to discuss the research in more detail. After the interview, one school elected not to participate and one school agreed. They expressed interest in viewing the results of the research and were told that they would receive a copy of the research upon completion. After informed written consent was given by the principal of the school, ethics clearance was given by Carleton University's research ethics board (see Appendix C for a copy of the ethics certificate). It was decided to administer the questionnaire in each classroom with the primary researcher (myself) present along with the classroom teacher and an interpreter in order to properly explain the nature of the research to the students. The questionnaires were administered over a four-day period in order to obtain responses from as many classes and students as possible. Due to a time conflict, one class was missed. However, the overall response rate achieved was quite good at just over 86% (341/395) of all possible participants (i.e. total number of middle school students registered at the school).

Before handing out questionnaires, students were given a brief explanation of the research and told that their participation was completely voluntary, they could quit at any

time, and that their responses would remain completely anonymous. This information was given to them in their native language. They were encouraged to ask any questions about the research before commencing. An interpreter was on hand to deal with any language issues during the administration. Students were observed for any distress while completing the questionnaire. Upon completion, questionnaires were placed in a manila envelope and sealed. The total time taken to administer the questionnaire ranged from approximately 10-20 minutes. Students were thanked for their assistance and in a few classes, after being given another opportunity to ask questions or make comments, a brief question and answer period followed.

4.5 Data Analysis

The data were inputted and analyzed with the SPSS statistical package (v.16). During the data input stage, a total of 22 questionnaires were excluded because 1) students only answered a few initial questions and then appeared to abandon the questions or 2) it was obvious that students had not read the questions (e.g. by creating a zigzag pattern of responses). When a reliability analysis was conducted on all questionnaire items, it indicated very strong internal consistency ($\alpha = .92$).

Individual items were checked for normality through visual inspection of histograms (with overlaid normality curves) and Q-Q plots. Upon inspection, the majority of variables appeared reasonably normally distributed, except for two variables which were highly negatively skewed and platykurtic (item# 3 and #30). Further inspection of the descriptive statistics for these variables (see Appendix D) confirmed that they deviated substantially from normality, and on this basis, they were removed from further

analysis. An examination of the descriptive statistics for the remaining variables revealed that a majority of variables were somewhat negatively skewed and/or platykurtic although there were no other extreme deviations.

An exploratory factor analytic (EFA) approach was utilized for this investigation. Even though many of the theoretical constructs used in this study had been tested in other contexts and had been found to be fairly reliable, they had not been tested in this specific context (a Korean private language school) and for these particular participants (Korean middle school students). Therefore, it was more preferable to allow constructs to emerge from the data instead of forcing constructs upon them as would be the case in confirmatory factor analysis (CFA). Further supporting the use of an EFA approach, this would be a convenience sample and, as a result, there would be no attempt to generalize the findings to a wider population.

In terms of an appropriate sample size for EFA, there have been a number of conflicting recommendations given. At minimum, Gorsuch (1983) has suggested that there be a ratio of five participants for each variable used and no less than 100 participants in the sample (for a fuller discussion of this recommendation and others, see MacCallum, Widaman, Zhang & Hong, 1999; Gorsuch, 1997). This research achieved a modest ratio of approximately 8.5:1 (N=341, 40 initial questionnaire items). Since this analysis was exploratory, this ratio was felt to be acceptable. A number of additional statistics were examined in order to confirm the appropriateness of this dataset for factor analysis. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy for this dataset was .896, which was quite satisfactory (Hutcheson & Sofroniou, in Field, 2005, p. 640). Bartlett's test of sphericity, which measures whether there are enough relationships

amongst the variables to conduct factor analysis, was significant ($p < .001$). An inspection of the anti-image correlation matrix revealed that all the diagonal values were above the bare minimum of .50 (Field, 2005, p. 659). In fact, 19/38 variables achieved values above .80 (17 of which were above .90). Finally, communalities for all variables were inspected and most variables (35/38) produced communalities in the low to moderate range (.40 to .70) which are fairly common in the social sciences (Costello & Osborne, 2005, p. 4). Overall, the average communality for all variables was .54.

Factors were extracted based on the principal axis factoring (PAF) method. This method was chosen as it is much less sensitive to assumptions of univariate and multivariate normality than other methods (Fabrigar, Wegener, MacCallum & Strahan, 1999; Costello & Osborne, 2005). The rotation method chosen was oblique since this allows for the extracted factors to correlate with each other (as opposed to an orthogonal rotation which does not). This was decided because the constructs being investigated were psychological in nature, and as such, there should be a fairly strong expectation that they will correlate to some degree (Fabrigar et al., 1999, p. 282). The promax procedure was selected after comparing the results obtained through other rotational methods. The promax procedure offered the most clearly interpretable results and the fewest cross-loaded variables. To guide factor selection, two standards were set: 1) they obtained eigenvalues of 1 or more (The Kaiser criterion), and 2) they were located at or above a point of inflexion (or 'elbow') on a scree plot (Cattell, 1966). A data reduction procedure was followed in which weakly-loaded and/or cross-loaded items ($>.30$) were systematically eliminated.

After the final factor solution was identified, factors scores for all participants were calculated using the regression method available in SPSS (v.16) which allows factor scores to correlate (Field, 2005, p. 646). Since factor scores can become new variables in the analysis of various subgroups (DiStefano, Zhu, & Mindrilla, 2009), the distribution of factor scores must be analyzed to ensure they meet the assumptions of the statistical tests to be used. Factor scores (and their distributions within various groups) were further analyzed for their suitability for correlation analysis and analysis of variance (ANOVA). Inspection of the factor score distributions indicated all were reasonably normally distributed. Levene's test for homogeneity of variance indicated the assumption was met for the majority of subgroups under analysis. In situations where subgroups violated this assumption, Welch's F and the Games-Howell post hoc statistic was utilized.

Chapter Five: Results

5.1 Exploratory Factor Analysis Results

A total of 5 EFA factors (based on the 10 original L2 motivational constructs) emerged from the EFA procedure utilized. Figure 4 below presents a visual representation of the role played by the 10 theoretical constructs and in the identification of the EFA factors. 6 of 10 original constructs played a role in the creating the L2

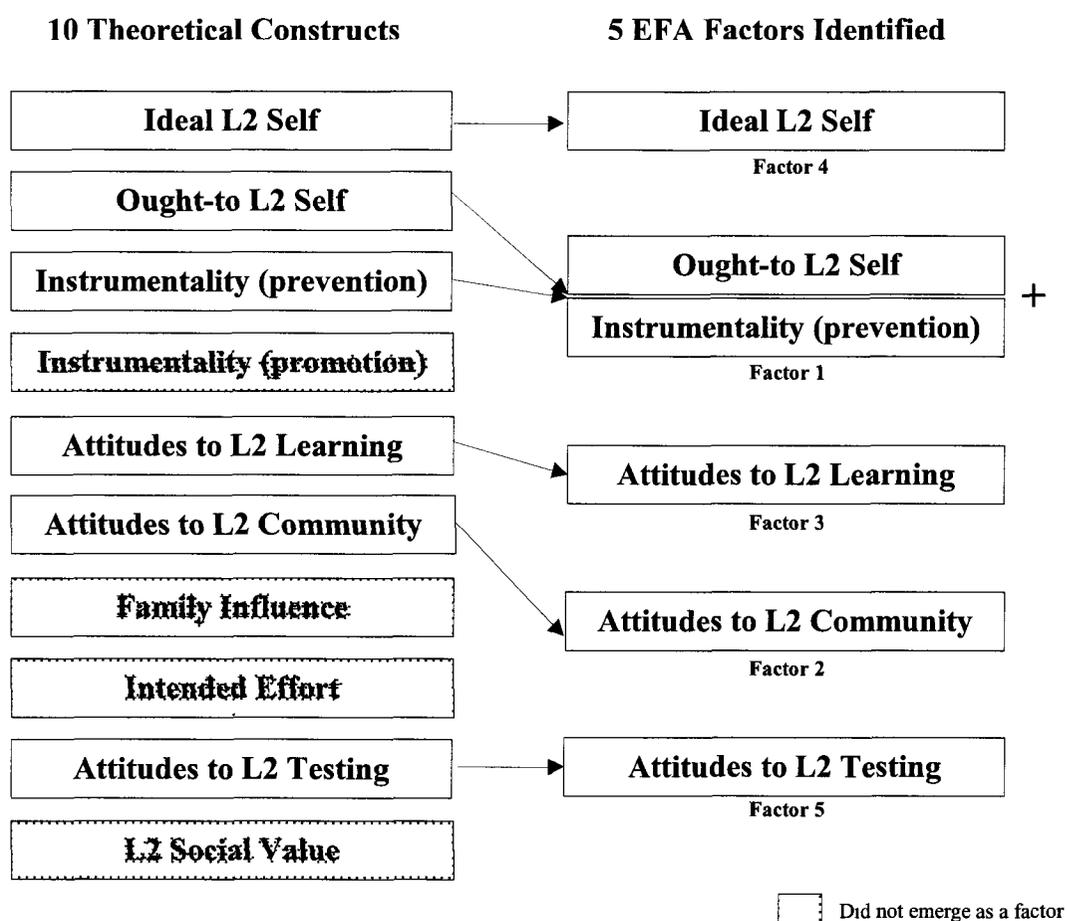


Figure 4: Theoretical Constructs and EFA Factors Identified

motivational factors for this study. As the figure shows, four of the five EFA factors were identified based on their theoretical constructs directly. One factor was identified based on two of the constructs that played a major role in its emergence, 'Ought-to L2 Self' and 'Instrumentality (prevention).'

Table 2 on the page below is the pattern matrix produced by SPSS (v.16). This lists the five extracted factors along with their respective item loadings, eigenvalues, percentages of explained variance, and Cronbach alphas. In total, these factors accounted for over 61% of total variance. The naming of the factors was fairly straightforward, with only one factor failing to approximate its theoretical construct.

At first, the EFA procedure extracted a total of 6 factors based solely on the Kaiser criterion (eigenvalue over 1). This is the default setting for which SPSS (v.16) conducts factor analysis. In order to determine the final number of factors to be included for analysis, the scree plot was also inspected (see Figure 2 below). A clear point of inflexion was found at Factor 3, suggesting a 3 factor solution. However, the line did not appear to 'even out' until after Factor 6. Based on this visual inspection, there appeared to be some justification for considering additional factors.

As this was an exploratory study, instead of eliminating potentially meaningful factors, it was decided to further investigate factors 4, 5 and 6. Reliability analyses of these factors indicated adequate internal consistency for factors 4 ($\alpha = .83$, 3 items) and 5 ($\alpha = .73$, 4 items), but weak internal consistency for factor 6 ($\alpha = .63$, 3 items). As a result, it was decided to keep factors 4 and 5, and exclude factor 6 from further analysis. This resulted in a five-factor EFA solution. A description of each identified factor will be provided before moving on to an analysis of factor scores.

Table 2: Pattern Matrix for 5-Factor Solution (EFA)

<i>Item #</i>	<i>Statement/Question</i>	<i>Factor 1</i>	<i>Factor 2</i>	<i>Factor 3</i>	<i>Factor 4</i>	<i>Factor 5</i>
25	Studying English is important to me, because I would feel ashamed if I got bad grades	.80				
19	Studying English is important to me because, If I don't have knowledge of English, I'll be considered a weak student	.77				
27	Studying English is important to me because an educated person is supposed to be able to speak English	.71				
14	Studying English is necessary for me because I don't want to get a poor score or a fail mark on English proficiency tests	.70				
28	Being successful in English is important to me so that I can please my parents/relatives	.69				
20	Studying English is important to me in order to attain a higher social respect	.68				
13	Studying English is important to me in order to gain the approval of my peers/teachers/family	.66				
11	I have to study English because I don't want to get bad marks in it	.61				
36	Do you like the people who live in English-speaking countries?		.95			
38	Would you like to know more about people from English-speaking countries?		.82			
34	Do you like meeting people from English-speaking countries		.75			
31	Would you like to travel to English-speaking countries?		.39			
29	Do you find learning English really interesting?			.84		
40	Do you really enjoy learning English?			.77		
32	Do you always look forward to English classes?			.69		
1	I would like to spend lots of time studying English			.55		
18	I can imagine myself living abroad and having a discussion in English				.90	
8	I can imagine myself speaking English as if I were a native speaker of English				.85	
22	I imagine myself as someone who is able to speak English				.62	
16	My score on English tests reflects my English ability					.68
9	English tests allow me to see how well I am doing in English					.64
2	I think the English tests I have taken so far are fair					.62
26	I trust my English test scores					.58
<i>Eigenvalues</i>		7.19	4.33	1.80	1.40	1.22
<i>Percentage of Variance Explained</i>		27.7%	16.7%	6.9%	5.4%	4.7%
<i>Cronbach Alpha</i>		.86	.86	.83	.86	.73

Extraction Method Principal Axis Factoring

Rotation Method Promax with Kaiser Normalization

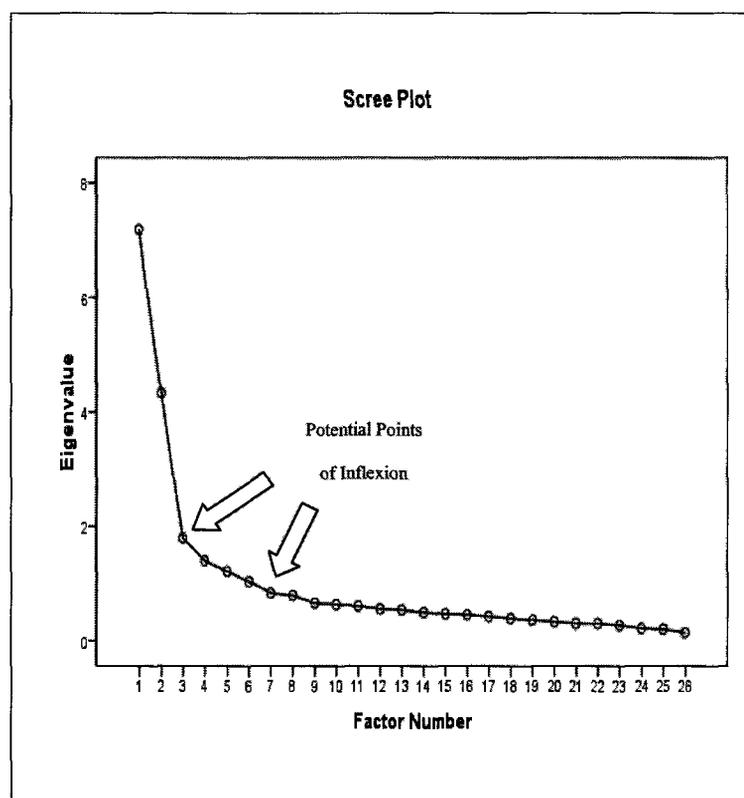


Figure 5: Scree Plot for 5-Factor EFA Solution

5.1.1 Factor 1: Ought-To L2 Self + Instrumentality (prevention)

This factor consisted of a combination of variables from two major constructs investigated in the Chinese context by Taguchi, Magid and Papi (2009). Four of the seven variables that loaded onto this factor (items 25, 19, 14, and 11) are from the Taguchi study construct ‘Instrumentality (prevention)’ and two others (items 27 and 13) are from their construct ‘Ought-to L2 Self.’ One additional variable (item 28) is from the Taguchi study construct ‘Family Influence.’ The Cronbach alpha obtained for this factor was good ($\alpha = .86$) indicating reasonable internal consistency. A closer inspection of

the wording of these items reveals they are all generally focused on the motivational influences of the surrounding social and educational environment. These include the fear of getting bad grades in comparison to others (items 25, 14, and 11), meeting the expectations of others (items 19 and 27), and the desire to gain approval from others (items 28, 20, and 13). For many participants there were strong associations among these items. Instead of providing a new name for this construct, it was decided to base this on the two theoretical constructs that played the largest role in its emergence: ‘Ought-to L2 Self’ + ‘Instrumentality (prevention).’

5.1.2 Factor 2: Attitudes to L2 Community

This factor replicated its theoretical construct based on the four items sampled from the Taguchi study construct, ‘Attitudes to L2 Community.’ Conceptually, this factor includes participants’ affinity for members of the L2 community (item 36) as well as their desire to know more about them (item 38), meet them (item 34), or travel to their country (item 31). The Cronbach alpha obtained for this factor was good ($\alpha = .86$) indicating reasonable internal consistency. Compared to the other variables loading significantly on this factor ($>.30$), participants’ desire to travel to an L2 country (item 31) had the weakest loading (.39). This suggests that, for some participants, the desire to travel was not necessarily directly associated with their positive attitudes towards L2 community members. However, the overall results indicated that these items were strongly associated for many participants.

5.1.3 Factor 3: Attitudes to L2 Learning

This factor consisted of three variables from the Taguchi study construct ‘Attitudes to L2 Learning’ and one item from their construct ‘Intended Effort.’ Conceptually, this factor includes participants’ interest and enjoyment in learning English (items 29 and 40), their anticipation of English class (item 32), and their desire to spend a lot of time studying English (item 1). The Cronbach alpha obtained for this factor was good ($\alpha = .83$) indicating reasonable internal consistency. Although the desire to spend a lot of time studying English (item 1) was based on another theoretical construct ‘Intended Effort,’ it was included due to its strong loading and its conceptual similarity to the other items (i.e., we tend to want to spend more time doing what we enjoy).

5.1.4 Factor 4: Ideal L2 Self

This factor replicated its theoretical construct based on three items sampled from the Taguchi construct ‘Ideal L2 Self.’ Conceptually, this factor includes participants’ ability to imagine themselves: living abroad and having English conversations (item 18); being able to speak as a native speaker of English (item 8); and as someone who can speak English (item 22). The Cronbach alpha obtained for this factor was good ($\alpha = .86$) indicating reasonable internal consistency. Additionally, these items clustered together on the original pattern matrix (see Appendix E), and with every extraction and rotation method attempted without any significant crossloadings ($>.30$) to or from other extracted factors.

5.1.5 Factor 5: Attitudes to L2 Testing

The final construct identified for this analysis was composed of 4 items originally used in a pilot study conducted on L2 testing and learning experiences in preparation for this thesis (Haggerty, in press). Conceptually, this factor includes participants' beliefs about L2 test scores in terms of accuracy (item 16), measurements of progress (item 9), fairness (item 2), and trustworthiness (item 26). The Cronbach alpha obtained for this factor was satisfactory ($\alpha = .73$) indicating adequate internal consistency. Although eight items were part of the original theoretical construct, four of these failed to load solely on this factor or were very weakly loaded ($<.30$). These were removed from further analysis.

5.1.6 Theoretical Constructs Not Replicated by EFA Procedure

A total of ten theoretical constructs guided the selection of questionnaire items for this exploratory study (see Figures 2 and 3). Four of these constructs were replicated as theoretically expected while two constructs appeared to conflate into one EFA factor: 'Ought-to L2 Self' + 'Instrumentality (prevention).' Four theoretical constructs failed to emerge from this EFA procedure. These included the Taguchi study constructs 'Instrumentality (promotion)', 'Family Influence', and 'Intended Effort' as well as the pilot study construct 'L2 Social Value.' Although some of the variables for these constructs clustered together as expected, they did not attain the statistical or theoretical strength necessary for identification. This was most likely a result of the paucity of variables included for these constructs (3 or 4) rather than a reflection of their lack of explanatory power. Follow-up research should include more items for each construct in order to properly assess their utility in better understanding young ELL L2 motivation.

5.2 Factor Scores

For each of the five factors identified, factor scores for each participant were calculated utilizing the regression method included in the 'Factor' procedure of SPSS (v.16). Participants were then separated into a number of subgroups so that comparisons could be made on the basis of their factor scores. For dichotomous groups, Pearson product-moment correlations were used while ANOVA was used to compare 3 groups or more. Given that some of the subgroup sample sizes to be tested were unequal, the assumption of homogeneity of variance was vital to obtain meaningful ANOVA results (Field, 2005, p. 324). Levene's test of homogeneity of variance for most of the subgroup comparisons indicated they satisfied this assumption. For the subgroup comparisons that did not meet this assumption, Welch's F and the Games Howell post hoc criterion were utilized instead.

5.2.1 Correlation among Factor Score Variables

Table 3 below reports the Pearson product-moment correlations for all factor score variables. All variables significantly correlated with each other ($p < .01$). However, many of the strongest associations discovered were correlated with 'Attitudes to L2 Learning,' this factor being highly associated with 'Attitudes to L2 Community' ($r = .684$, $p < .01$), 'Ideal L2 Self' ($r = .554$, $p < .01$), and 'Attitudes to L2 Testing' ($r = .501$, $p < .01$). There was also a strong association discovered between 'Attitudes to L2 Community' and 'Ideal L2 Self' ($r = .601$, $p < .01$). Interestingly, 'Ought-to L2 Self + Instrumentality (prevention)' significantly correlated with 'Ideal L2 Self' ($r = .215$, $p < .01$), although it was the lowest of the correlations.

Table 3: Pearson Product-Moment Correlations for Factor Score Variables

<i>Factors</i>	<i>Factor 1 Ought-to L2 Self + Instrumentality (prevention)</i>	<i>Factor 2 Attitudes to L2 Community</i>	<i>Factor 3 Attitudes to L2 Learning</i>	<i>Factor 4 Ideal L2 Self</i>	<i>Factor 5 Attitudes to L2 Testing</i>
Factor 1 Ought-to L2 Self + Instrumentality (prevention)	1.000	.322**	.277**	.215**	.399**
Factor 2 Attitudes to L2 Community		1.000	.684**	.601**	.411**
Factor 3 Attitudes to L2 Learning			1.000	.554**	.501**
Factor 4 Ideal L2 Self				1.000	.343**
Factor 5 Attitudes to L2 Testing					1.000

** Correlation is significant at the .01 level

5.2.2 Results for Gender and Education Level

Pearson point-biserial correlation coefficients were calculated for gender and participants' factor scores. A statistically significant gender effect ($r=.109$, $p<.05$) was discovered for the factor 'Attitudes to L2 Community' indicating that female students in this sample tended to hold more positive attitudes on this dimension. However, there were no other significant gender effects discovered for any other factor. In terms of education level, no significant effects ($p<.05$) were discovered among the education groups on any of the factors. However, when 1st and 2nd year middle school students ($N=119$) were grouped together and compared with 3rd year middle school students ($N=229$) a significant effect ($r=-.156$, $p<.01$) was discovered for 'Attitudes to L2 testing' indicating that participants in their 3rd year of middle school tended to hold more negative attitudes about L2 testing than their younger cohorts. There were no other significant education level effects ($p<.05$) discovered for any other factor.

5.2.3 Results for Self-Reported L2 Proficiency

Respondents were asked to report their L2 proficiency level. Of the participants who responded to this question, 22% (74/339) reported their L2 proficiency level to be ‘beginner’ or ‘low-intermediate’, 45% (154/339) reported ‘intermediate’, and 33% (111/339) reported it to be ‘high-intermediate’ or ‘advanced’. Table 5 below reports the results of a one-way ANOVA of L2 proficiency groups and their scores on each of the 4 factors that satisfied Levene’s test for homogeneity of variance. Results indicated there were significant mean differences in these groups on three of four factors ($p < .001$). There were no significant differences found among proficiency groups for the factor ‘Ought-to L2 Self + Instrumentality (prevention)’, $F(2, 329) = 2.32, p > .05, \eta^2 = .01$. Post hoc analyses using Tukey’s HSD indicated there were significant mean differences

Table 4: Factor Scores and Self-Reported L2 Proficiency Level (N=329)

<i>Factor Scores</i>		<i>Sums of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>Sig</i>	<i>Eta Squared (η^2)</i>
Ought-to L2 Self + Instrumentality (prevention)	Between Groups	4 180	2	2 090	2 322	100	01
	Within Groups	294 381	327	900			
	Total	298 561	329				
Attitudes to L2 Community	Between Groups	21 438	2	10 719	12 594	000	07
	Within Groups	278 302	327	851			
	Total	299 740	329				
Attitudes to L2 Learning	Between Groups	35 113	2	17 557	22 175	000	12
	Within Groups	258 893	327	792			
	Total	294 007	329				
Attitudes to L2 Testing	Between Groups	23 623	2	11 812	16 594	000	09
	Within Groups	232 764	327	712			
	Total	256 387	329				

between the 'beginner to low-intermediate' and 'intermediate' groups on 'Attitudes to L2 Community' ($p < .001$), 'Attitudes to L2 Learning' ($p < .001$), and 'Attitudes to L2 Testing' ($p < .001$). A significant difference also existed between the 'intermediate' and 'high intermediate to advanced' groups on 'Attitudes to L2 Testing' ($p < .01$).

The factor 'Ideal L2 Self,' for which homogeneity of variance could not be assumed, Welch's F indicated that there were significant differences among the group means, $F(2, 167.495) = 21.024, p < .001$. Post hoc analysis using Games-Howell indicated there was a significant difference between the 'beginner to low-intermediate' and the 'intermediate' group ($p < .001$) and between the 'intermediate' and 'high-intermediate to advanced' group ($p < .001$).

Overall, these results suggest that participants' perception of their L2 proficiency were associated with their attitudes about the L2 community, L2 learning, and L2 testing, as well as on their conception of an 'Ideal L2 Self' and that these attitudes were most negative among participants who reported their L2 proficiency to be 'beginner to low-intermediate.' For the significant ANOVA results for which Levene's test of homogeneity of variance was satisfied (see Table 4), Eta squared values suggested medium effect sizes for 'Attitudes to L2 Community' ($\eta^2 = .07$) and 'Attitudes to L2 Testing' ($\eta^2 = .09$), and an effect size approaching large for 'Attitudes to L2 Learning' ($\eta^2 = .12$).

5.2.4 Results for L2 Test Preparation

Data was collected regarding the average number of hours participants spent on publically-available English test preparation (not English tests completed in their public

school). Of those who responded to this question, 25% (81/318) reported spending '0 hours per week' preparing for such tests, 28% (90/318) reported spending '1-3 hours per week', 25% (78/318) reported spending '4-6 hours per week', and 22% (70/318) reported spending a total of '7 hours or more per week.'

Table 4 below reports the results of a one-way ANOVA of L2 test preparation groups and their scores on each of the 4 factors that satisfied Levene's test for homogeneity of variance. Results indicated there were significant differences in these groups on all four factors ($p < .01$). Post hoc analyses using Tukey's HSD indicated there were significant mean differences between the '0 hours per week' and '4-6 hours per week' groups on 'Attitudes to L2 Learning' ($p < .01$) and 'Attitudes to L2 Testing' ($p < .01$). There were also significant differences between the '0 hours per week' and '7

Table 5: Factor Scores and L2 Test Preparation (N=309)

<i>Factor Scores</i>		<i>Sums of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>Sig</i>	<i>Eta Squared (η^2)</i>
Ought-to L2 Self + Instrumentality (prevention)	Between Groups	14 785	3	4 928	5 552	.001	.05
	Within Groups	271 613	306	888			
	Total	286 398	309				
Attitudes to L2 Community	Between Groups	17 226	3	5 742	6 550	.000	.06
	Within Groups	268 273	306	877			
	Total	285 499	309				
Attitudes to L2 Learning	Between Groups	22 715	3	7 572	8 980	.000	.08
	Within Groups	258 025	306	843			
	Total	280 741	309				
Attitudes to L2 Testing	Between Groups	29 941	3	9 980	14 073	.000	.12
	Within Groups	217 015	306	709			
	Total	246 956	309				

hours or more per week' groups on all four factors ($p < .001$).

The factor 'Ideal L2 Self,' for which homogeneity of variance could not be assumed, Welch's F indicated that there were significant differences among the group means, $F(3, 168.796) = 11.697, p < .001$. Post hoc analysis using Games-Howell indicated there was a significant difference between the '0 hours per week' and the '7 hours or more' group ($p < .001$) and between the '1-3 hours per week' and '7 hours or more' group ($p < .001$).

These results suggest that the number of hours spent preparing for external English tests were significantly associated with attitudes towards the L2 community, L2 learning, and L2 testing. In addition, they appeared to play a role in respondents' perceptions of 'Ought-to L2 Self + Instrumentality (prevention)' and 'Ideal L2 Self.' These attitudes and self-conceptions were most negative among participants who reported not spending any time per week preparing for external English tests. According to Cohen (in Volker, 2006, p. 662) effect sizes (Eta squared) can be viewed as 'small' ($\eta^2 \geq .01$), 'medium' ($\eta^2 \geq .06$), or 'large' ($\eta^2 \geq .14$). For the significant ANOVA results for which Levene's test of homogeneity of variance was satisfied (see Table 5), these results suggested a small effect size for 'Ought to L2 Self + Instrumentality (prevention)' ($\eta^2 = .05$), medium effect sizes for 'Attitudes to L2 Community' ($\eta^2 = .06$) and 'Attitudes to L2 Learning' ($\eta^2 = .08$), and an effect size approaching large for 'Attitudes to L2 Testing' ($\eta^2 = .12$).

5.2.5 Results for Level and Frequency of L2 Test Completion

Participants were also asked to report the type and number of L2 tests they had completed. Participants were only asked to report their completion of publically-available English tests (not English tests taken in their public school). For the complete spelling of the test acronyms used for these tests, please refer to the Table of Abbreviations (p. viii). A complete explanation of these tests and their intended purposes is beyond the scope of this thesis (for a fuller explanation of all these tests and their purposes, please see Choi, 2008). For this analysis, English tests were separated into two main test groups: 1) middle-school level tests (PELT, TOEIC Bridge, TOSEL, and Jr. G-TELP) that are explicitly intended for assessments of elementary or middle school students for a variety of purposes including diagnostic, progress, and achievement assessment, and 2) university-entrance level L2 tests (TOEFL and TEPS) defined for this study as fairly cognitively-demanding measurements of English proficiency that are accepted for admission to major universities in South Korea.

For middle-school level tests, 72% (247/341) of respondents reported they had not completed any of these tests while 28% (94/341) reported that they had. For university entrance level tests, 67% (229/341) of respondents reported they had not completed any of these tests while 33% (112/341) reported that they had. Pearson point-biserial correlation coefficients were calculated between each dichotomous group and their respective factor scores and the results are listed in Table 6 below. Significant correlations were discovered for the middle-school level test subgroups on two of five factors: 'Attitudes to L2 Community' ($r=.119$, $p<.05$) and 'Attitudes to L2 Learning' ($r=.125$, $p<.05$).

For the university-entrance level test subgroups, significant differences were discovered on four of five factors: ‘Attitudes to L2 Community’ ($r=.166$, $p<.01$), ‘Attitudes to L2 Learning’ ($r=.210$, $p<.01$), ‘Ideal L2 Self’ ($r=.174$, $p<.01$), and ‘Attitudes to L2 Testing’ ($r=.202$, $p<.01$). The strength and direction of these correlations indicated that middle school students who had completed a university-entrance level L2 test generally held more positive L2 attitudes, beliefs, and self-conceptions. However, there was no significant difference found for the factor ‘Ought-to L2 Self + Instrumentality (prevention)’ for both test subgroups.

Table 6: Factors and Level of L2 Test Completion (N=332)

<i>L2 Test Level Completed</i>		<i>Ought-to L2 Self + Instrumentality (prevention)</i>	<i>Attitudes to L2 Community</i>	<i>Attitudes to L2 Learning</i>	<i>Ideal L2 Self</i>	<i>Attitudes to L2 Testing</i>
Middle School Level (no/yes)	Correlation Coefficient	062	119*	125*	097	025
	Sig (2-tailed)	261	030	023	076	665
University Entrance Level (no/yes)	Correlation Coefficient	027	166**	210**	174**	202**
	Sig (2-tailed)	620	002	000	001	000

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

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

Participants were also asked to report the number of times they had completed a university-entrance level L2 test. Amongst those who had reported completing these tests, 57% (64/112) had completed them 1-3 times, 29% (32/112) 4-6 times, and 14% (16/112) 7 or more times, including two students who completed these types of tests a

remarkable 12 times. One-way ANOVA results indicated there were no significant differences in these groups on all 5 factors ($p < .05$). Test frequency groups were not significantly different in their factor scores for 'Ought-to L2 Self + Instrumentality (prevention),' $F(2, 108) = 2.296, p > .05$, 'Attitudes to L2 Community,' $F(2, 108) = 1.231, p > .05$, 'Attitudes to L2 Learning,' $F(2, 208) = .781, p > .05$, and 'Attitudes to L2 Testing,' $F(2, 108) = .886, p > .05$. For the factor 'Ideal L2 Self,' for which the assumption of homogeneity of variance was not satisfied, Welch's F indicated there were no significant mean difference among these groups, $F(2, 41.745) = 41.745, p > .05$. These results suggest the number of times university-entrance level tests had been completed by middle school students was not significantly associated with any of their scores on the five EFA factors identified.

Having presented the results of the EFA procedure (the factors identified), and the results of the statistical comparison of factor scores for each learner-background characteristic (the variables), the next chapter will discuss some of the implications for these findings in relation to the research questions established for this study.

Chapter Six: Discussion

This chapter will discuss the implications of the research results in terms of the research questions that have been established for this study. Each research question will be listed under its appropriate heading.

6.1 L2 Motivational Factors Identified

The 5-factor EFA solution identified in this study provides additional support for the utility of constructs derived from Dörnyei's L2 Motivational Self System in a wide variety of educational contexts. Despite sampling a fairly low number of items (3 or 4) from the Taguchi study, the five EFA factors that emerged in this study displayed a high degree of concordance with their theoretical antecedents. These constructs appear to hold explanatory power in understanding young learner L2 motivation in this context and provide a foundation for further exploration using alternate research methods (perhaps CFA, mixed methods, or a strictly qualitative approach).

Four of the factors that emerged from the EFA procedure, 'Attitudes to L2 Learning,' 'Attitudes to L2 Community,' 'Ideal L2 Self,' and 'Attitudes to L2 Testing' were identified based on their respective theoretical constructs. These factors generally replicated their theoretical constructs with only minor differences. However, one factor did not replicate its theoretical construct as precisely. The factor identified as 'Ought-to L2 Self + Instrumentality (prevention), explaining the greatest amount of variance (27.7%), was composed of variables from a number of constructs. Of the eight items

comprising this factor, four originated from the construct ‘Instrumentality (prevention),’ 2 came from ‘Ought-to L2 Self,’ one came from ‘Instrumentality (promotion), and one came from ‘Family Influence.’ Perhaps, had more items been sampled from each construct, they would have loaded onto separate factors instead of clustering together as they did in this study. However, it is also possible that any EFA method utilized would have had difficulty separating these closely related concepts regardless of the number of items included.

In fact, there is good reason not to expect the constructs ‘Instrumentality (prevention)’ and ‘Ought-to L2 Self’ to distinguish themselves significantly from each other. Taguchi, Magid, and Papi (2009) separated the construct of ‘Instrumentality’ into its ‘prevention’ and ‘promotion’ aspects, inspired by a distinction made by Higgins (1998), in order to assess their respective contributions to ‘Ideal L2 Self’ and ‘Ought-to L2 Self.’ In all three Asian contexts studied they found that ‘Instrumentality (prevention)’ was highly correlated with ‘Ought-to L2 Self’ (Taguchi et. al., 2009, p. 78). For this reason, it would not be surprising for the two constructs to conflate to some degree. Further investigation, employing more items per construct, would be necessary to determine if they can be meaningfully separated for young ELLs in this context. Nevertheless, the five factors identified in this study mirrored a number of previously researched L2 motivational constructs and provided a solid foundation for the further analysis of subgroup differences.

There was also significant correlation among all 5 EFA factors. This is to be expected given that each construct is attempting to probe related facets of L2 motivation. It is not all that surprising to discover that attitudes to L2 learning are associated with

attitudes to L2 communities or L2 testing. There is likely a strong symbiotic relationship among these variables for most ELLs. However, it is interesting to note that ‘Ought-to L2 Self + Instrumentality (prevention) + ‘Ideal L2 Self’ were also found to be significantly correlated. As Taguchi, Magid, and Papi (2009) discovered in their study of the Chinese context, there appears to be a complex connection between conceptions of one’s future self and one’s desire to satisfy the expectations of significant others. They suggest that this may be because many people in China “view themselves not only in individual terms ... but also as a direct extension of their family (Taguchi et al., 2009, p. 80). The desire to satisfy the expectations of others may become significantly internalized so that distinctions between what one wants for oneself and what one is pressured to do by others become strongly associated. The results of this study provide support for a similar relationship in the Korean context.

6.2 Gender and Education Level

Research Question #1

Are L2 motivational factors influenced by gender?

Overall, gender effects were minimal as scores on only one factor, ‘Attitudes to L2 Community,’ appeared to be significantly affected. This result indicates that female participants tended to have a greater desire to learn more about L2 communities. This interest may be influenced by a desire to travel or to learn more about foreign cultures. However, the lack of any gender effect on the other four factors indicates that female and

male participants did not differ greatly on the majority of L2 motivation factors under investigation.

Research Question #2

Are L2 motivational factors influenced by education level?

In terms of an education level (or age) effect, the results indicated that participants did not differ greatly based on their grade level on any of the factors under study. To further explore any potential grade (age) difference, students in their first and second year were combined into one group and compared with students in their third and final year. These groups did not differ greatly on four of five L2 motivational factors, but there was a difference discovered for ‘Attitudes to L2 Testing,’ third year middle school students reporting more negative attitude and beliefs. This difference may be due to the increased stress that many 3rd grade middle school students experience. This may cause them to have more negative attitudes towards L2 tests, particularly those who are unable to do a higher level L2 test.

In my experience, students who have not demonstrated the English proficiency necessary (as demonstrated in class and on more minor English tests completed in school) are generally not encouraged (by themselves, and often by others as well) to do a university-entrance level English test like the TOEFL or the TEPS. Those who have not demonstrated the ability to cope with such a demanding test often feel more stress as the deadline to complete the high school application process approaches. The high school application process begins near the end of the third year of middle school. Middle school

students' English test scores are a significant part of this process, particularly if they aim to enter a foreign language high school. Those who are unable to complete a university-entrance level test may react more negatively to the L2 testing environment. The increased stress surrounding this period could help account for why there are more negative attitudes to L2 testing for students in their third year of middle school than those in earlier grades.

6.3 Self-Reported L2 Proficiency

Research Question #3

Are L2 motivational factors influenced by young ELL perceived L2 proficiency?

The results of this study strongly indicated that perceptions of L2 proficiency play very dramatic roles in a number of L2 attitudes, beliefs and self-conceptions. However, in this study, the factor 'Ought-to L2 Self + Instrumentality (prevention)' was not significantly associated with perceived L2 proficiency level. This suggests that all students (beginner to advanced) were well aware of the importance that others have placed on the learning task and the necessity of meeting these expectations in order to avoid the negative consequences of failure.

However, perceived L2 proficiency was significantly associated with students' attitudes and beliefs towards L2 communities, L2 learning, and L2 testing, as well as their conception of an 'Ideal L2 Self.' Middle school students who had the lowest perceptions of their L2 proficiency tended to hold the most negative opinions on all these L2 motivational aspects. It may be tempting for some to conclude that this negativity is entirely the fault of the student. If students had more faith in their ability, they would be

able to maintain a more positive attitude towards the subject matter. However, this is exactly the kind of stagnant and myopic view of the language learner that Firth and Wagner (1997) and others have warned leads to distorted conceptions of what is really happening in the social context. It is vital to consider the attitudes and behaviour of individuals in combination with their local and extended social environment as ‘a topic of investigation’ itself (Firth & Wagner, 1997, p. 288). To accomplish this, it is necessary to consider intervening factors (both individual and social) that may be influencing this perception of L2 proficiency and may also be having a concomitant affect on L2 motivation.

For these participants, influences on perceptions of L2 proficiency would very likely include any feedback they had received from teachers in their classrooms or from family members based on their performance in the subject. Of course, there are a host of other factors that likely play a contributing role in perceptions of L2 proficiency, for instance the students’ interest or enjoyment of the subject, the informal feedback received from teachers and parents, or the student’s goals associated with the learning task. However, in the assessment of one’s own proficiency in a subject, two factors are likely to influence many language learners: 1) the level of proficiency generally expected by significant others and 2) the level of proficiency demonstrated by cohorts. Social norms have a dramatic influence on one’s perception as they provide an objective way to compare. In this context, which is heavily exam-based, one of the most influential sources for social comparison may be the results of tests. In my personal experience, the results of exam scores are crucial to perceptions of proficiency and this has a greater

effect on younger ELLs who are often less equipped to interpret the meaning of test results.

6.4 L2 Test Preparation

Research Question #4

Are L2 motivational factors influenced by the time spent preparing for L2 tests?

In this study, those who reported spending the most time in L2 test preparation (apart from their regular school tests) also had the most positive L2 motivational attitudes, beliefs and self-conceptions. Unfortunately, the results themselves provide little guidance in explaining potential reasons for this relationship. Those who spend a lot of time in L2 test preparation may have more clearly defined goals, or may develop them as a result, and this could be influencing their positivity. It is also possible that these students have more confidence in their L2 proficiency and believe they have a greater chance at completing a high level L2 test successfully. Further investigation is needed to determine if this relationship exists outside of this specific context and to investigate potential causes. My suspicion is that these students have been strongly encouraged to prepare for high-level English tests (by themselves, by their parents and teachers, and by the larger educational environment) and have welcomed, or rationalized, the opportunity because they have accepted the importance of the learning task. In my experience, there have been very few students who expressed an intrinsic interest or enjoyment of test preparation, although this is another variable to consider in future analysis.

6.5 Level and Frequency of L2 Test Completion

Research Question #5

Are L2 motivational factors influenced by the level of L2 test completion?

When students were divided on the basis of whether they had completed a university-entrance level test completion, some striking results emerged. Those who had completed one of these tests at least once reported significantly more positive L2 attitudes, beliefs, and self-conceptions. These students were more positive on 4 of the 5 L2 motivational factors.

Research Question #6

Are L2 motivational factors influenced by the frequency of university-level L2 test completion?

In terms of how many times middle school students completed a TOEFL or TEPS, there were no significant associations discovered. The most dramatic differences in L2 motivational factors occurred between the middle-school group who had not taken either of these tests and the group that had completed at least one of them. For this sample at least, it did not appear that repeatedly taking these tests magnified any differences discovered in the L2 motivational factors based on L2 test completion alone.

While these results may surprise some educators, they are likely to make more intuitive sense to those with extensive experience in highly test-intensive educational environments in Asia. In my own personal experience as an EFL educator in South Korea, it became quite apparent to me over the years that students who were able to do

one of these tests considered it somewhat of a status symbol. Many students perceived that this distinguished them from many others and gave them a far better chance of gaining access to more prestigious educational and occupational resources in the future. For many students, it gave them a sense of accomplishment because they had been able to accomplish something that so many others had not. Alternatively, students who felt they were unable to do one of these tests, and so did not, generally accepted that this could have a negative impact on their future. This was the message being given to many students from their surrounding educational and social environment.

Given the exploratory method utilized for this research, it is not possible to confidently state any causal relationships involved in the L2 motivational relationships discovered. However, it might assist in the design of follow-up studies to consider how various theories of motivations might attempt to explain the relationships discovered in this study. Perhaps, some students were more or less innately motivated to achieve as Achievement Motivation Theory might have us believe. Perhaps, some failed to set specific, important, and attainable goals according to tenets of Goal-Setting Theory. It is possible that some developed a 'performance' orientation instead of a 'mastery orientation' (Goal-Oriented Theory) and thus become paralyzed on the basis of their weak results. Maybe, some attributed their failures to their own inability and thus had given up hope as Attribution Theory might speculate. Some may have internalized a belief that they cannot accomplish the task (Self-efficacy Theory) or attempted to maintain their sense of 'self-worth' by rejecting any interest in the task itself. Additionally, some may have not been given the necessary autonomy in order to become more 'self-determined.'

All the above motivational theories could potentially help to explain a part of the L2 motivational patterns discovered in this study. However, many these theories would also tend to focus largely on the individual and pay little to no attention considering the wider social environment surrounding and impinging upon us all. Without this consideration, insight into factors affecting L2 motivation would be drastically incomplete. The danger of focusing on the psychology of the individual language learner is that the greater social context is often pushed to the sidelines when in fact it needs to be center-stage. Without adequate consideration of the wider social context and how this interacts with individuals, an incomplete picture is likely to emerge. In this context, the results of this study suggest that high-stakes L2 testing is playing a role in L2 motivational attitudes, beliefs, and self-conceptions. Given that these are young ELLs, there is an urgency in determining whether the existence of such a high-stakes test, coupled with pressure to take a high-level test, is adversely affecting a great many L2 learners who are not able to do so, and if so, the extent of this impact. These results suggest that there is an effect, but it falls to future research to determine its detailed nature, extent, and significance.

Chapter Seven: Conclusions and Recommendations

7.1 Limitations

There are a number of limitations which need to be mentioned before offering some conclusions and recommendations. One of the main objectives of this study was to explore the potential factors influencing young ELL L2 motivation in the South Korean context; however, being a convenience sample consisting of middle school students in one private language school in Seoul, the results cannot be generalized to a wider population. Although many parents in South Korea enrol their children in private language schools, there are also many who likely do not possess the financial capability to do so. The relationships discovered in this study may be dramatically different for middle school students who do not have access to these educational resources. A randomly selected sample, including middle school students who have not attended private language schools, would be needed in order to generalize these findings to the larger Korean educational context. In addition, the number of items utilized for some of the theoretical constructs developed for this study was the bare minimum (3) recommended for factor analysis (MacCallum, Widaman, Zhang, & Hong, 1999). This was done primarily to keep the questionnaire brief for middle school students. However, in retrospect, it would have been preferable to reduce the number of constructs under investigation. Further research should attempt to incorporate more items (at least 5) for each construct in order to more properly assess their explanatory power for younger ELLs in this educational context. Finally, as this research relied exclusively on questionnaire data, it represents only a snapshot of the attitudes and opinions of

respondents. Future research of a more qualitative and longitudinal nature is needed to capture richer and deeper understandings of learner attitudes, beliefs and self-conceptions, how this affects motivated behavior, and how this changes over time and space.

7.2 Conclusions about L2 Motivation

The results of this study do not clearly refute or deny any of the theories of motivation discussed in Chapter Three. Unfortunately, the respondents' voices echoing behind the data, as they generally are in strictly quantitative methods, are largely faint and somewhat ambiguous. Although a significant amount of data was collected and some interesting patterns discovered, few explanations for these patterns can be gleaned from the data itself. It will of course fall to future research to establish more definitively whether the relationships discovered in this study exist in other similar contexts, the extent of influence they have on the educational environment and the individual language learner, and ultimately, what can be done to help minimize any negative consequences that have resulted. Undoubtedly, a mixed-method approach would be preferable in tackling the enormously complex network of L2 motivational influences surrounding young ELLs in this context. In the meantime, it is helpful to speculate about what the results of this study could mean in terms of L2 motivation.

The major findings of this research were threefold. It was found that the amount of time young ELLs spent in L2 test preparation, their self-reported L2 proficiency level, and their level of L2 test completed (middle-school or university-entrance) were all strongly associated with the middle-school students' L2 attitudes, beliefs, and self-

conceptions. Those middle school students who prepared more for L2 tests (apart from their public school), had higher self-reported L2 proficiency, or who had completed a university-entrance level L2 tests, also had more positive scores on a majority of L2 motivational factors. For potential explanations, as it is for all who conduct research into the human condition, we must look to theory for guidance. My perusal of the various motivational theories described in Chapter Three has led me to the conclusion that they all have something unique to offer in the piecing together of this motivational puzzle and no one motivational theory would be able to account for all of the patterns observed.

However, I found the most explanatory power for the interpretation of these results in Self-Efficacy Theory. In order to better define this concept, it is instructive to quote the words of the thinker credited with its origin:

This core belief [self-efficacy] is the foundation of human motivation, well-being, and accomplishments. Unless people believe they can produce desired effects by their action, they have little incentive to act or to persevere in the face of difficulties. Whatever other factors serve as guides and motivators, they are rooted in the core belief that one has the power to effect changes by one's actions. (Bandura, 2006, p. 3)

For me, the strength of this statement lies in its simple intuitive truth. There simply appears to be a lot of strength in the power of one's belief and this is often difficult to separate from one's assumptions or even one's 'knowledge.' This was greatly impressed upon me by the concept of beliefs, assumptions, and knowledge (BAK) articulated by Devon Woods (1996) to better understand teachers' verbalizations. The interwoven nature of BAK evolves "in the face of conflicts and inconsistencies ... gain[ing] depth and breadth as varied events are interpreted and reflected upon (Woods, 1996, p. 212). In

my own experience preparing young ELLs for high-stakes tests, I was often left wondering what conflicts and varied events were influencing the early development of my students' BAK. The issue that has continued to intrigue me is not so much about the importance of these beliefs, but instead how these beliefs arise in the first place, how they become more negative or more positive, and how they may change from context to context over the course of one's life. As Bandura's (2006) social cognitive theory postulates, a complete appreciation of human behavior requires an "integrated causal system in which socio-structural influences operate through psychological mechanisms to produce behavioural effects" (p. 5). Clearly, the interaction between the individual and the social environment is crucial to a more complete understanding of L2 motivation.

For the middle school students in this study, their classroom environment represents a significant component of their socio-structural environment, one that they are generally not free to walk away from. They must deal with this environment in one way or another. The transition from elementary to middle school is a dramatic step in most students' lives (Zimmerman & Cleary, 2006). To be sure, there are a host of new environmental variables a middle school student must negotiate, including the existence of an entirely new peer group, a new set of educational expectations, and perhaps a more disciplined and routinized school structure. However, in my experience in both of these settings, one of the most salient differences I have noticed between a middle and elementary school classroom in South Korea has been the increased importance placed on the results of tests. The results of tests take on new meaning for many middle school students and this may encourage many to engage in a greater degree of social comparison, whether they admit to it or not (Schunk & Meece, 2006). These normative

comparisons have an immediate effect on cognitive perceptions of ability. As Schunk and Meece (2006) have noted “students own performances offer the most reliable guides for gauging self-efficacy In general successes raise and failures lower self-efficacy...” (p. 73). As the importance and challenge of these tests rises, they have a concomitant effect on the self-efficacy of many individuals, in both positive and negative directions.

This study found that the middle-school students who had completed a university-entrance level L2 test generally held more positive L2 attitudes, beliefs, and self-conceptions. This is perhaps surprising to some with little experience in test-intensive educational environments. For students in these environments, being able to do one of the most challenging tests available to them is likely to be perceived as a significant accomplishment, one that may lead to improved educational opportunities in the future. The inability to do a highly challenging test may negatively affect the learning trajectory of students throughout middle school. The increased stress resulting from a social expectation that one should complete the most challenging test available (in this case the TOEFL and TEPS) may significantly transform the educational environment and have both positive and negative influences L2 attitudes, beliefs and self-conceptions of middle school students on the basis of those who can perform (and thus take the test), and those who cannot (and thus avoid the test). These conclusions are of course speculative, but they explain the exploratory results obtained in this study fairly well and align with my own personal experience in the setting. This is by no means the only possible interpretation of the data. Indeed, this interpretation even if meaningful may only apply to this specific context. It will take much further investigation, employing deeper and richer investigative techniques in a wider variety of settings in order to better understand the

dynamic relationship between L2 testing and L2 motivation. The remainder of this thesis will consider the implications of these findings for the validation process of high-stakes tests and offer a few recommendations for improving the administration of such tests.

7.3 Conclusions about Validity in Language Testing

The second chapter of this thesis examined the development of validity over the last half century and considered a role for L2 motivation research in the L2 test validation process. This would entail expanding the traditional concept of validity beyond considerations of test content, the appropriate criteria to be used, and an assessment of the validity of the construct on an evidential basis alone. It would require an acceptance of the value implications involved in the use of test scores and a consideration of the consequences of test use (Messick, 1989). These are considerations which have been acknowledged by a number of prominent test theorists (e.g., Kane, 2001; Bachman, 2005, McNamara & Roever, 2006), but have been resisted by many in the field as being too burdensome or impractical in reality (e.g., Wiley, 1991; Popham, 1997). However, the results of this study have highlighted the benefit and necessity of moving beyond the internal makeup of the test itself and to more properly acknowledging the social context in which all tests live and breathe. This is especially urgent when there are indications, such as those suggested by this study, that high-stakes L2 testing designed for university-entrance purposes may be creating an educational culture among ELLs that may motivate some students at the expense of others. In addition, excessive reliance on these large-scale standardized tests does not generally provide satisfactory diagnostic feedback in order to guide learning, improve instruction, or gauge progress (Jang, 2009). There are a

number of results that should cause concern for responsible test creators, users, and administrators.

First, there should be concern over the fact that a surprisingly high number of young ELLs are taking a L2 test that is designed for learners at higher educational levels. In this study, 33% (112/341) of middle school students had taken either a TOEFL or TEPS at least once, a number of them taking it more than once. As these tests are designed for students preparing for entrance to university, there should be some investigation into the appropriateness of such testing on young learners. Additionally, there should be concern over the amount of time that some students are devoting to external L2 test preparation. In this study, some middle school students were spending a great deal of time preparing for external English tests. If a young ELL spends a significant amount of time preparing for English tests that are not designed for their age level, the potential effect that this could have on their overall educational development should be investigated.

There should also be concern expressed over the level of self-reported proficiency that was reported by the students in this study. Since this may be based on their ability (or inability) to do an English test that is too cognitively demanding for them, a middle school student's basis of comparison in order to judge their own proficiency may become highly distorted. They may develop a belief that they are unable to do something because the bar has been set far too high. Even though some appear able to barely reach it, (undoubtedly, with an enormous amount of effort and assistance) the effects on those who may be unable, could be significantly influencing their L2 motivation. The results of this study suggest that the use (or more accurately, misuse) of university-entrance level

L2 tests in this educational context could be having negative consequences on the great majority of young ELLs (67% in this study) who did not complete one of these tests.

In South Korea, testing is heavily relied upon as a measurement of a student's ability beginning at a young age (see Nam, 2006; Choi, 2008). In this context, the ability to take a test that is beyond the capability of most others of the same age is likely considered a noteworthy accomplishment and gives students a significant advantage in a very competitive educational marketplace. Students who have not demonstrated the necessary foundational skills in English (based on their performances on public school English tests or external publically available middle-school level tests) are generally not encouraged to take more advanced English tests. This creates an early separation of L2 students based on their L2 testing ability that can have profound and lasting effects of their L2 learning trajectory.

In terms of validity, the results of this study further vindicate the importance of investigating the use of test scores and the consequences of test use as an integral part of a continuing test validation process. While it is not possible for test designers to consider all possible uses (and misuses) of their test prior to its introduction, it is incumbent on test designers to consider all sources of evidence available in order to ensure their test is performing as intended within the social context. If the cognitive level required to understand and successfully complete a high-stakes test is beyond what a typical young ELL is able to achieve, it is at the very least worthwhile to investigate the consequences this may have on their short-term and long-term L2 attitudes, beliefs, and self-conceptions. If it can be shown that this has a positive impact on conceptions of proficiency and the learning trajectory of the young ELLs who are somehow able to do it,

while negatively impacting those who are not, it is an area where proponents of unified concept of validity should be gravely concerned as it is just these kinds of unfair and unjust consequences that modern language testing should be striving to avoid

At minimum, these results call for further investigation into the effects of high-stakes language testing on L2 motivation, especially for young learners. If one of our goals as parents, educators, curriculum developers, and hopefully as test designers and administrators, is to develop an environment conducive to learning, it is necessary to better understand not only what a test is getting from learners, but also what a test might be doing to them. To this end, the concept of ‘effect-driven test architecture’ advanced by Fulcher and Davidson (2009) holds considerable promise. This concept places social responsibility for test designs where it is needed most, at its source, by inspiring test designers to align “explicit statements of intended test-effect to test design decisions through explicit validity arguments and test architectures” (Fulcher, 2009, p. 13). In this way, tests that are used for purposes other than they are intended would be forced to ‘retrofit their validity argument’ (Fulcher & Davidson, 2009; Fulcher, 2009). This carries great potential to minimize negative test impact.

There is, and should be, a responsibility demanded of test designers to properly consider the extended effects of their tests, particularly when they are used for purposes other than intended. This becomes especially appropriate when there are financial rewards involved in the extended use. Responsibility should not evaporate across geo-cultural time and space. To wit, explicit statements should be encouraged beyond the ‘what and how’ of the test. These should include, amongst other statements: 1) who a test is designed to test, specifying the ages for which the test is designed; 2) any potential

socio-cognitive effects the test could have on young individuals, especially when the test is used other than intended; and 3) follow-up strategies to investigate the consequential effects of tests and to utilize this information to continually improve test design and deployment. If the value and ultimate power of consequential validity is to be taken as seriously as Messick intended, this needs to become an integral part of the test design and administration process itself. Just as we have come to expect a level of responsibility from companies producing products such as movies, computer games, or toys to consider the age appropriateness of their merchandise, hopefully we can expect an equal level of responsibility from language test designers, users, and administrators as well.

References

- American Psychological Association. (1954). Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin Supplement*, 51 (2), 1-38.
- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, 84 (3), 261-271.
- Anderson, N. (1991). Individual differences in strategy use in second language reading and testing. *The Modern Language Journal*, 75 (4), 460-472.
- Atkinson, J. (1964). *An introduction to motivation*. Princeton, NJ: Van Nostrand.
- Atkinson, J. & Raynor, J. (1974). *Motivation and achievement*. New York: Wiley.
- Au, S. (1988). A critical appraisal of Gardner's social-psychological theory of second language (L2) learning. *Language Learning*, 38 (1), 75-100.
- Bachman, L. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2 (1), 1-34.
- Bachman, L. & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman.
- Bandura, A. (2006). Adolescent development from an agentic perspective. In F. Pajares, & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 1-43). Greenwich, CT: Information Age Publishing, Inc.
- Canale, M. & Swain, M. (1980). The theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
- Carroll, J. (1961). Fundamental considerations in testing for English proficiency of foreign students. In *Testing the English proficiency of foreign students* (pp. 30-40). Washington, D.C.: Center for Applied Linguistics.
- Cattell, R. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276.
- Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. Cambridge: Cambridge University Press.

- Choi, I. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing, 25* (1), 39-61.
- Cohen, A. (1984). On taking language tests: What the test takers report. *Language Testing, 1* (1), 70-81.
- Cohen, A. (2007). The coming of age for research on test-taking strategies. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. Turner, & C. Doe (Eds.), *Language testing reconsidered* (pp. 89-111). Ottawa: University of Ottawa Press.
- Costello, A. & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10* (7), 1-9.
- Covington, M. (1992). *Making the grade: A self-worth perspective on motivation and school reform*. Cambridge: Cambridge University Press.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth, TX: Holt, Rinehart & Winston.
- Cronbach, L. (1988). Five perspectives on validity argument. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 3-17). Hillsdale, NJ: Lawrence Erlbaum.
- Cronbach, L. & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.
- Crookes, G. & Schmidt, R. (1991). Motivation: Reopening the research agenda. *Language Learning, 41* (4), 469-512.
- Deci, E. & Ryan, R. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.
- Deci, E. & Ryan, R. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11* (4), 227-168.
- DiStefano, C., Zhu, M. & Mindrilla, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation, 14* (20), 1-11.
- Dörnyei, Z. (2001). *Teaching and researching motivation*. Essex, England: Pearson Education Ltd.

- Dörnyei, Z. (2009). The L2 motivational self system. In Z. Dörnyei, & E. Ushioda (Eds.), *Motivation, language identity, and the L2 self* (pp. 9-43). Bristol: Multilingual Matters.
- Dörnyei, Z. & Kormos, J. (2000). The role of individual and social variables in oral task performance. *Language Teaching Research*, 4 (3), 275-300.
- Dörnyei, Z. & Taguchi, T. (2010). *Questionnaires in second language research: Construction, administration, and processing* (2nd Ed.). New York: Routledge.
- Fabrigar, L., Wegener, D., MacCallum, R. & Strahan, E. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4 (3), 272-299.
- Fairbairn, S. & Fox, J. (2009). Inclusive achievement testing for linguistically and culturally diverse test takers: Essential considerations for test developers and decision makers. *Educational Measurement: Issues and Practice*, 28 (1), 10-24.
- Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London: Sage.
- Firth, A. & Wagner, J. (1997). On discourse, communication, and (some) fundamental concepts in SLA research. *Modern Language Journal*, 81, 285-300.
- Fox, J. (2004). Test decisions over time: Tracking validity. *Language Testing*, 21 (4), 437-465.
- Fox, J. & Cheng, L. (2007). Did we take the same test? Differing accounts of the Ontario Secondary School Literacy Test by first and second language test-takers. *Assessment in Education*, 14 (1), 9-26.
- Fulcher, G. (2009). Test use and political philosophy. *Annual Review of Applied Linguistics*, 29, 3-20.
- Fulcher, G. & Davidson, F. (2009). Test architecture, test retrofit. *Language Testing*, 26 (1), 123-144.
- Gardner, R. (1985). *Social psychology and second language learning: The role of attitudes and motivations*. London: Edward Arnold.
- Gardner, R. & Lambert, W. (1972). *Attitudes and motivations in second language learning*. Rowley, MA: Newbury House.

- Gorsuch, G. (2009). Investigating second language learner self-efficacy and future expectancy of second language use for high-stakes program evaluation. *Foreign Language Annals*, 42 (3), 505-540.
- Gorsuch, R. (1983). *Factor Analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Gorsuch, R. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment*, 68 (3), 532-560.
- Haggerty, J. (in press). The polarizing effect of high-stakes testing on young ELLs: An analysis of test validity and language proficiency identity (LPID). *Asian EFL Journal*.
- Harter, S., Whitesell, N. & Kawalski, P. (1992). Individual differences in the effects of educational transitions on young adolescent's perception of confidence and motivational orientation. *American Educational Research Journal*, 29(4), 777-807.
- Henry, A. (2009). Gender differences in compulsory pupils' L2 self-concepts: A longitudinal study. *System*, 37, 177-193.
- Higgins, E. (1987). Self-discrepancy: A theory relating self and aspect. *Psychological Review*, 94 (3), 319-340.
- Hymes, D. (1971). *Towards communicative competence*. Philadelphia: University of Pennsylvania Press.
- Jang, E. (2009). Cognitive diagnostic assessment of L2 reading comprehension ability: Validity arguments for Fusion Model applications to LanguEdge assessment. *Language Testing*, 26 (1), 31-73.
- Kane, M. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112 (3), 527-535.
- Kane, M. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38 (4), 319-342.
- Kissau, S. (2006). Gender differences in second language motivation: An investigation of micro- and macro-level influences. *Canadian Journal of Applied Linguistics*, 9 (1), 73-96.

- Kormos, J. & Csizer, K. (2008). Age-related differences in the motivation of learning English as a foreign language: Attitudes, selves, and motivated learning behavior. *Language Learning, 58* (2), 327-355.
- Lado, R. (1961). *Language testing: The construction and use of foreign language tests*. London: Longman.
- Law, Y. (2009). The role of attribution beliefs, motivation and strategy use in Chinese fifth graders' reading comprehension. *Educational Research, 51* (1), 77-95.
- Li, D. (1988). "It's always more difficult than you plan or imagine": Teachers' perceived difficulties in introducing the communicative approach in South Korea. *TESOL Quarterly, 32* (4), 677-703.
- Locke, E. & Latham, G. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice Hall.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3*, 635-694.
- MacCallum, R., Widaman, K., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*, 84-89.
- MacIntyre, P., MacKinnon, S. & Clement, R. (2009). The baby, the bathwater, and the future of language learning motivation research. In Z. Dörnyei, & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 43-65). Bristol: Multilingual Matters.
- Markus, H. & Nurius, P. (1987). Possible selves: The interface between motivation and the self-concept. In K. Yardley, & T. Honess (Eds.), *Self and identity: Psychosocial perspectives* (pp. 157-172). Chichester: John Wiley and Sons.
- McNamara, T. (1996). *Measuring second language performance*. New York: Longman.
- McNamara, T. (2001). Language assessment as social practice: Challenges for research. *Language Testing, 18* (4), 333-349.
- McNamara, T. (2005). 21st Century Shibboleth: Language tests, identity and intergroup conflict. *Language Policy, 4*, 351-370.
- McNamara, T. (2006). Validity in language testing: The Challenge of Sam Messick's legacy. *Language Assessment Quarterly, 3* (1), 31-51.

- McNamara, T. & Roever, C. (2006). *The social dimension of language testing*. Malden, MA: Blackwell.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Education Researcher*, 18 (2), 5-11.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences of persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50 (9), 741-749.
- Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research*, 45, 35-44.
- Moss, P. (1998). The role of consequences in validity theory. *Educational Measurement: Issues and Practice*, 17 (2), 6-12.
- Nam, M. (2006). The changing role of English in Korea: From English as a tool for advancement to English for survival. *Journal of Pan-Pacific Association of Applied Linguistics*, 9 (2), 227-240.
- Noels, K., Pelletier, L., Clement, R. & Vallerand, R. (2000). Why are you learning a second language? Motivational orientations and self-determination theory. *Language Learning*, 50 (1), 57-85.
- Norton, B. (2000). *Identity and language learning: Gender, ethnicity and educational change*. Essex: Pearson Education Ltd.
- Popham, J. (1997). Consequential validity: Right concern - wrong concept. *Educational Measurement: Issues and Practice*, 16 (2), 9-13.
- Pychyl, T., Coplan, R. & Reid, P. (2002). Parenting and procrastination: Gender differences in the relations between procrastination, parenting style, and self-worth in early adolescence. *Personality and Individual Differences*, 33, 271-285.
- Ross, S. (2008). Language testing in Asia: Evolution, innovation and policy challenges. *Language Testing*, 25 (5), 5-13.
- Sakai, H. & Kikuchi, K. (2009). An analysis of demotivators in the EFL classroom. *System*, 37, 57-69.
- Schunk, D. & Meece, J. (2006). Self-efficacy development in adolescence. In F. Pajares, & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 71-96). Greenwich, CT: Information Age Publishing, Inc.

- Shepard, L. (1993). Evaluating test validity. *Review of Research in Education*, 19, 405-450.
- Shepard, L. (1997). The centrality of test use and consequences of test validity. *Educational Measurement: Issues and Practice*, 16 (2), 5-8, 13, 24.
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Harlow: Pearson Education Limited.
- Shohamy, E. (2008). Language policy and language assessment: The Relationship. *Current Issues in Language Planning*, 9 (3), 363-373.
- Skinner, B. (1965). *Science and human behavior*. New York: Free Press.
- Sparks, R., Patton, J., Ganschow, L. & Humbach, N. (2009). Long-term relationships among early first language skills, second language aptitude, second language affect, and later second language proficiency. *Applied Psycholinguistics*, 30, 725-755.
- Spolsky, B. (1995). *Measured words: The development of objective language testing*. Oxford: Oxford University Press.
- Spolsky, B. (2007). On second thoughts. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. Turner, & C. Doe (Eds.), *Language Testing Reconsidered* (pp. 9-18). Ottawa: University of Ottawa Press.
- Stipek, D. (1996). Motivation and instruction. In D. Berliner, & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 85-113). New York: Macmillan.
- Strong, M. (1984). Integrative motivation: Cause or result of successful second language acquisition. *Language Learning*, 34 (3), 1-14.
- Taguchi, T., Magid, M. & Papi, M. (2009). The L2 motivational self system among Japanese, Chinese and Iranian learners of English: A comparative study. In Z. Dörnyei, & E. Ushioda (Eds.), *Motivation, Language Identity and the L2 Self* (pp. 66-97). Bristol: Multilingual Matters.
- Volker, M. (2006). Reporting effect size estimates in school psychology research. *Psychology in the Schools*, 43 (6), 653-672.
- Warden, C. & Lin, H. (2000). Influence of integrative motivation in an Asian EFL setting. *Foreign Language Annals*, 33 (5), 535-545.

- Weiner, B. (1992). *Human motivation: Metaphors, theories, and research*. Newbury Park, CA: Sage.
- Weiner, B. (1994). Integrating social and personal theories of achievement striving. *Review of Educational Research*, 64 (4), 557-573.
- Wiley, D. (1991). Test validity and invalidity reconsidered. In R. Snow, & D. Wiley (Eds.), *Improving inquiry in the social sciences: A volume in honor of Lee J. Cronbach* (pp. 75-107). Hillside, NJ: Erlbaum.
- Williams, M. & Burden, R. (1999). Student's developing conceptions of themselves as language learners. *The Modern Language Journal*, 83 (2), 193-201.
- Woods, D. (1996). *Teacher cognition in language teaching: Beliefs, decision-making and classroom practice*. Cambridge: Cambridge University Press.
- Yashima, T. (2002). Willingness to communicate in a second language: The Japanese EFL context. *The Modern Language Journal*, 86 (i), 54-66.
- Zimmerman, B. & Cleary, T. (2006). Adolescent's development of personal agency: The role of self-efficacy beliefs and self-regulatory skill. In F. Pajares, & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 45-69). Greenwich, CT: Information Age Publishing, Inc.
- Zimmerman, B. & Schunk, D. (2008). Motivation: An essential dimension of self-regulated learning. In D. Schunk, & B. Zimmerman (Eds.), *Motivation and self-regulated learning* (pp. 1-30). Mahwah, NJ: Erlbaum.

Appendix A: Items Excluded from Taguchi Study

Construct: Ideal L2 Self (2/5 excluded)

- * I can imagine myself speaking English with international friends or colleagues
 - * Whenever I think of my future career, I imagine myself using English
-

Construct: Ought-to L2 Self (4/7 excluded)

- * I study English because close friends of mine think it's important
 - * I consider learning English important to me because the people I respect think that I should do it.
 - * It will have a negative impact on my life if I don't learn English
 - * Studying English is important to me because other people will respect me more if I have a knowledge of English.
-

Construct: Instrumentality –Prevention (1/5 excluded)

- * I have to learn English because I don't want to fail the English course.
-

Construct: Instrumentality – Promotion (4/8 excluded)

- * Studying English is important to me because English proficiency is necessary for promotion in the future
 - * Studying English can be important to me because I think I'll need it for further studies.
 - * The things I want to do in the future require me to use English
 - * Studying English is important to me in order to in order to achieve a special goal (e.g. to get a degree or a scholarship).
-

Construct: Attitudes to L2 Learning (0/4 excluded)

Construct: Attitudes to L2 Community (0/4 excluded)

Construct: Family Influence (2/5 excluded)

- * I must study English in order to avoid being punished by my parents/relatives.
 - * Studying English is important to me in order to bring honours to my family.
-

Construct: Intended Effort (3/6 excluded)

- * If an English course were offered in the future, I would like to take it
 - * I am prepared to expend a lot of effort in learning English
 - * I would like to concentrate on studying English more than any other topic
-

Appendix B: English Learner Questionnaire



이 설문지는 익명이므로, 가능한 한 정직하게 답해주세요. 모든 질문에 답해 주시기 바랍니다.
This survey is completely anonymous, so please answer as honestly as possible. Please try to answer all of the questions.

PART A:

다음 세부사항들을 적절한 박스에 체크를 하거나 빈칸을 채워 주세요.

Please provide the following details by checking [✓] the appropriate box or filling in the blanks.

1. 생년월일? / When were you born? _____ (년) Year _____ (월) Month

2. 성별 / Gender (남자) Male (여자) Female

3. 교육 수준 / Education Level

<input type="checkbox"/> 중학교 (1 학년)	<input type="checkbox"/> 중학교 (2 학년)	<input type="checkbox"/> 중학교 (3 학년)	<input type="checkbox"/> 기타
Middle School (1 st Year)	Middle School (2 nd Year)	Middle School (3 rd Year)	Other

4. 영어 능력 수준 / English Proficiency Level

당신의 현재 영어 능력 수준을 평가한다면? /
 How would you rate your current level of English proficiency?

<input type="checkbox"/> (초급)	<input type="checkbox"/> (초중급)	<input type="checkbox"/> (중급)	<input type="checkbox"/> (중고급)	<input type="checkbox"/> (고급)
beginner	low intermediate	intermediate	high intermediate	advanced

5. 영어 시험 경험 / English Testing Experience

아래 영어능력평가 시험이나 그 외에 가졌던 시험과 회수를 기입해 주세요.
 How many times have you taken the following tests? Include any other English tests you have taken.

TOEFL	=	_____ 회	TOEIC (Bridge)	=	_____ 회
TEPS	=	_____ 회	Jr. G-TELP	=	_____ 회
PELT	=	_____ 회	기타 (other)_____	=	_____ 회

6. 영어 시험 학습 / English Test Preparation

공인 영어능력평가 시험을 준비하기 위해 일주일에 대략 얼마나 많은 시간을 할애합니까?

Not including your school English tests, how many hours per week do you currently study for public English tests?

- 0 시간 (0 hours)
 1-3 시간 (1-3 hours)
 4-6 시간 (4-6 hours)
 7-9 시간 (7-9 hours)
 10+시간 (10+ hours)

PART B:

아래 진술에 대해 1~6 번까지 해당란에 표시 해주세요. 모든 질문에 답해주시기 바랍니다.

Please indicate how much you agree or disagree with the following statements by circling a number from 1 (strongly disagree) to 6 (strongly agree) Please try to answer all the questions

강한 반대 Strongly Disagree	반대 Disagree	약한 반대 Slightly Disagree	약한 동의 Slightly Agree	동의 Agree	강한 동의 Strongly Agree
1	2	3	4	5	6

1	나는 영어를 공부하는데 많은 시간을 할애하고 싶다. I would like to spend lots of time studying English	1	2	3	4	5	6
2	내가 지금까지 보았던 영어 시험은 공정하다고 생각한다. I think the English tests I have taken so far are fair	1	2	3	4	5	6
3	영어 공부를 하는 것은 미래 경력에 유용할 것이라고 생각하기 때문에 나에게 중요하다. Studying English is important to me because I think it will be useful in my future career	1	2	3	4	5	6
4	영어 시험은 나의 학습에 긍정적인 영향을 가져왔다. English tests have had a positive effect on my learning	1	2	3	4	5	6
5	높은 영어 능력 수준을 가지면 많은 돈을 벌 수 있기 때문에 영어 공부를 하는 것은 중요하다. Studying English is important because with a high level of English proficiency I can make a lot of m	1	2	3	4	5	6
6	학급 친구들과 비교해 볼 때, 나는 상대적으로 열심히 영어 공부를 한다고 생각한다. Compared to my classmates, I think I study English relatively hard	1	2	3	4	5	6
7	내 부모님/가족은 지적인 사람이 되기 위해 영어를 공부해야만 한다고 믿는다. My parents/family believe(s) that I must study English to be an intelligent person	1	2	3	4	5	6
8	나는 마치 원어민인 것처럼 영어로 말하는 것을 상상할 수 있다. I can imagine myself speaking English as if I were a native speaker of English	1	2	3	4	5	6
9	영어 시험은 내가 영어를 얼마나 잘 하고 있는지를 보여준다. English tests allow me to see how well I am doing in English	1	2	3	4	5	6

10	영어를 공부하는 것은 내 인생에 새로운 도전을 제공하기 때문에 나에게 중요하다. Studying English is important to me because it offers a new challenge in my life	1	2	3	4	5	6
11	나쁜 영어 점수를 받고 싶지 않기 때문에 영어 공부를 해야만 한다. I have to study English because I don't want to get bad marks in it	1	2	3	4	5	6
12	나는 영어 시험의 결과에 가치를 둔다. I value the results of my English tests	1	2	3	4	5	6
13	영어를 공부하는 것은 동료/선생님/가족들의 인정을 받기 위해 나에게 중요하다. Studying English is important to me in order to gain the approval of my peers/teachers/family	1	2	3	4	5	6
14	나는 영어 공부를 할 필요가 있다. 왜냐하면 영어능력평가 시험 (TOEFL, TEPS, etc) 에서 나쁜 점수를 받거나 낙제 점수를 받고 싶지 않기 때문이다. Studying English is necessary for me because I don't want to get a poor score or a fail mark in English proficiency tests (e.g. TOEFL, TEPS, etc)	1	2	3	4	5	6
15	나의 가족은 내가 영어공부를 하도록 많은 압력을 가한다. My family puts a lot of pressure on me to study English	1	2	3	4	5	6
16	나의 영어 시험 점수는 나의 영어 능력을 반영한다. My score on English tests reflects my English ability	1	2	3	4	5	6
17	내 주변 사람들은 내가 영어를 배우기를 바라기 때문에 나는 영어를 배울 필요가 있다. Learning English is necessary because people around me expect me to do so	1	2	3	4	5	6
18	나는 외국에서 살거나 영어로 토론하는 상상을 할 수 있다. I can imagine myself living abroad and having a discussion in English	1	2	3	4	5	6
19	만약 내가 영어에 대한 지식을 가지고 있지 않다면, 나는 열등생으로 생각 될 것이기 때문에 영어를 공부 하는 것은 나에게 중요하다. Studying English is important to me because, if I don't have knowledge of English, I'll be considered a weak student	1	2	3	4	5	6
20	영어를 공부하는 것은 높은 사회적 존경을 받기 위해 나에게 중요하다. Studying English is important to me in order to attain a higher social respect	1	2	3	4	5	6
21	영어를 배우기 위해 최선을 다하고 있다고 생각한다. I think that I am doing my best to learn English	1	2	3	4	5	6
22	나는 나 자신을 영어로 말할 수 있는 누군가로 상상 한다. I imagine myself as someone who is able to speak English	1	2	3	4	5	6
23	영어 시험 점수는 내가 더 열심히 영어 공부를 하고 싶도록 만든다. My English tests scores make me want to study English harder	1	2	3	4	5	6
24	나의 영어 시험 점수는 내가 영어를 배우는 것을 좌절시켜왔다 . My English test scores have discouraged me from learning English	1	2	3	4	5	6
25	나쁜 점수를 얻는다면 부끄럽기 때문에 영어 공부를 하는 것은 나에게 중요하다. Studying English is important to me, because I would feel ashamed if I got bad grades	1	2	3	4	5	6
26	나는 내 영어 시험 점수를 신뢰한다. I trust my English test scores	1	2	3	4	5	6
27	지성인이면 영어로 말할 수 있다고 생각되어 지기 때문에 영어 공부를 하는 나에게 중요하다 Studying English is important to me because an educated person is supposed to be able to speak English	1	2	3	4	5	6
28	영어로 성공하는 것은 부모님/친척들을 기쁘게 할 수 있기 때문에 나에게 중요하다 Being successful in English is important to me so that I can please my parents/relatives	1	2	3	4	5	6

PART C:

새로운 질문들이 있습니다. 위에서 했던 것과 같이 답해주세요.

These are new questions but please answer them the same was as you did before

전혀 아니다 Not at all	그다지 아니다 Not so much	그저 그렇다 So-so	약간 그렇다 A little	많이 그렇다 Quite a lot	매우 많이 그렇다 Very Much
1	2	3	4	5	6

29	영어를 배우는 것이 정말 재미있다고 생각합니까? <i>Do you find learning English really interesting?</i>	1	2	3	4	5	6
30	사람들이 영어를 배우지 않고도 성공할 수 있다고 생각합니까? <i>Do you think people can be successful without learning English?</i>	1	2	3	4	5	6
31	영어권 국가에서 여행하고 싶습니까? <i>Would you like to travel to English-speaking countries?</i>	1	2	3	4	5	6
32	당신은 항상 영어 수업을 학수고대 합니까? <i>Do you always look forward to English classes?</i>	1	2	3	4	5	6
33	모든 사람들이 영어를 배우는 것이 중요하다고 생각합니까? <i>Do you think it is important for everyone to learn English?</i>	1	2	3	4	5	6
34	영어권 국가의 사람들을 만나는 것을 좋아합니까? <i>Do you like meeting people from English-speaking countries?</i>	1	2	3	4	5	6
35	당신의 영어 수업 환경을 좋아합니까? <i>Do you like the atmosphere of your English classes?</i>	1	2	3	4	5	6
36	영어권 국가에 사는 사람들을 좋아합니까? <i>Do you like the people who live in English-speaking countries?</i>	1	2	3	4	5	6
37	모든 사람들이 성공하기 위해 영어 공부를 할 필요가 있다고 생각합니까? <i>Do you think everyone needs to study English in order to be successful?</i>	1	2	3	4	5	6
38	영어권 국가의 사람들에 대해 더 알고 싶습니까? <i>Would you like to know more about people from English-speaking countries?</i>	1	2	3	4	5	6
39	대학교와 회사는 영어 시험 점수에 가치를 두어야 한다고 생각합니까? <i>Do you think universities and companies should value English test scores</i>	1	2	3	4	5	6
40	당신은 영어 배우는 것을 정말 즐기니까? <i>Do you really enjoy learning English?</i>	1	2	3	4	5	6

이 설문 조사에 참여해 주셔서 감사합니다.

Thank you for participating in this survey

Appendix C: Ethics Clearance Form



Carleton University Research Office

5th Floor Tory Building
1125 Colonel By Drive
Ottawa ON K1S 5B6 Canada
Tel: 613-520-2516
Fax: 613-520-2521
www.carleton.ca/cu/research/curo/

Ethics Clearance Form

This is to certify that the Carleton University Research Ethics Board has examined the application for ethical clearance. The REB found the research project to meet appropriate ethical standards as outlined in the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans* and, the *Carleton University Policies and Procedures for the Ethical Conduct of Research*.

New clearance

Renewal of original clearance

Original date of clearance:

Date of clearance **8 June 2010**
 Researcher **John Haggerty**
 Status **M.A. student**
 Department **School for Linguistics and Language Studies**
 Supervisor **Professor Janna Fox**
 Title of project **High-stakes testing and young English language learners: Exploring motivation and test validity in South Korea**

Clearance expires: **31 May 2011**

All researchers are governed by the following conditions:

Annual Status Report: You are required to submit an Annual Status Report to either renew clearance or close the file. Failure to submit the Annual Status Report will result in the immediate suspension of the project. Funded projects will have accounts suspended until the report is submitted and approved.

Changes to the project: Any changes to the project must be submitted to the Carleton University Research Ethics Board for approval. All changes must be approved prior to the continuance of the research.

Adverse events: Should any participant suffer adversely from their participation in the project you are required to report the matter to the Carleton University Research Ethics Board. You must submit a written record of the event and indicate what steps you have taken to resolve the situation.

Suspension or termination of clearance: Failure to conduct the research in accordance with the principles of the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans* and the *Carleton University Policies and Procedures for the Ethical Conduct of Research* may result in the suspension or termination of the research project.

Leslie J. MacDonald-Hicks
 Research Ethics Board Coordinator
 For the Chair of the Carleton University Research Ethics Board
 Prof. Antonio Gaultieri

Appendix D: Descriptive Statistics for All Questionnaire Items

	N	Mean	Std Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Std Error	Statistic	Std Error
1 I would like to spend lots of time studying English.	340	4.10	1.301	-.642	.132	-.036	.264
2 I think the English tests I have taken so far are fair	340	3.97	1.243	-.500	.132	-.215	.264
3 Studying English is important to me because I think it will be useful in my future career	340	5.03	1.135	-1.428	.132	2.050	.264
4 English tests have had a positive effect on my learning	339	3.97	1.337	-.388	.132	-.444	.264
5 Studying English is important because with a high level of English proficiency I can make a lot of money	339	4.04	1.368	-.506	.132	-.347	.264
6 Compared to my classmates, I think I study English relatively hard	341	3.51	1.316	-.085	.132	-.551	.263
7 My parents/family believe(s) that I must study English to be an intelligent person	340	3.82	1.469	-.209	.132	-.803	.264
8 I can imagine myself speaking English as if I were a native speaker of English	341	4.01	1.436	-.518	.132	-.517	.263
9 English tests allow me to see how well I am doing in English	341	3.79	1.379	-.244	.132	-.719	.263
10 Studying English is important to me because it offers a new challenge in my life	341	4.22	1.322	-.474	.132	-.360	.263
11 I have to study English because I don't want to get bad marks in it	339	4.11	1.358	-.482	.132	-.474	.264
12 I value the results of my English tests	340	3.93	1.373	-.321	.132	-.647	.264
13 Studying English is important to me in order to gain the approval of my peers/teachers/family	341	3.70	1.342	-.172	.132	-.549	.263
14 Studying English is necessary for me because I don't want to get a poor score or a fail mark in English proficiency tests (e.g. TOEFL, TEPS, etc)	340	3.96	1.376	-.559	.132	-.287	.264
15 My family puts a lot of pressure on me to study English	341	3.33	1.511	.006	.132	-.865	.263
16 My score on English tests reflects my English ability	340	3.98	1.282	-.520	.132	-.226	.264
17 Learning English is necessary because people around me expect me to do so	341	3.51	1.303	-.215	.132	-.499	.263
18 I can imagine myself living abroad and having a discussion in English	341	4.08	1.484	-.506	.132	-.546	.263

19 Studying English is important to me because, If I don't have knowledge of English, I'll be considered a weak student	341	3 80	1 275	- 180	132	- 483	263
20 Studying English is important to me in order to attain a higher social respect	341	3 81	1 275	- 272	132	- 404	263
21 I think that I am doing my best to learn English	340	3 56	1 305	- 002	132	- 550	264
22 I imagine myself as someone who is able to speak English	340	4 11	1 314	- 427	132	- 279	264
23 My English tests scores make me want to study English harder	339	3 78	1 348	- 378	132	- 351	264
24 My English test scores have discouraged me from learning English	341	3 71	1 472	- 124	132	- 808	263
25 Studying English is important to me, because I would feel ashamed if I got bad grades	340	3 69	1 308	- 222	132	- 376	264
26 I trust my English test scores	340	3 46	1 297	- 125	132	- 516	264
27 Studying English is important to me because an educated person is supposed to be able to speak English	341	3 70	1 248	- 222	132	- 209	263
28 Being successful in English is important to me so that I can please my parents/relatives	340	3 57	1 378	- 176	132	- 582	264
29 Do you find learning English really interesting?	341	3 31	1 424	- 032	132	- 754	263
30 Do you think people can be successful without learning English?	341	3 63	1 569	- 048	132	-1 172	263
31 Would you like to travel to English-speaking countries?	341	4 64	1 417	- 953	132	124	263
32 Do you always look forward to English classes?	341	3 19	1 273	337	132	- 192	263
33 Do you think it is important for everyone to learn English?	340	3 70	1 520	- 172	132	- 922	264
34 Do you like meeting people from English-speaking countries?	341	3 90	1 505	- 262	132	- 819	263
35 Do you like the atmosphere of your English classes?	341	3 26	1 377	068	132	- 677	263
36 Do you like the people who live in English-speaking countries?	341	3 87	1 412	- 175	132	- 755	263
37 Do you think everyone needs to study English in order to be successful?	341	3 61	1 510	- 143	132	- 872	263
38 Would you like to know more about people from English-speaking countries?	341	3 99	1 446	- 285	132	- 704	263
39 Do you think universities and companies should value English test scores	340	3 19	1 488	085	132	- 845	264
40 Do you really enjoy learning English?	341	3 42	1 436	- 038	132	- 769	263
Valid N (listwise)	322						

Appendix E: Original Pattern Matrix before Systematic Item Elimination

	Factor								
	1	2	3	4	5	6	7	8	9
25 Studying English is important to me, because I would feel ashamed if I got bad grades	792								
19 Studying English is important to me because, If I don't have knowledge of English I'll be considered a weak student.	710								
11 I have to study English because I don't want to get bad marks in it.	647								
14 Studying English is necessary for me because I don't want to get a poor score or a fail mark in English proficiency tests (e.g. TOEFL, TEPS etc)	646								
12 I value the results of my English tests	599								
13 Studying English is important to me in order to gain the approval of my peers/teachers/family	574								
28 Being successful in English is important to me so that I can please my parents/relatives	566								
27 Studying English is important to me because an educated person is supposed to be able to speak English	561								
24 My English test scores have discouraged me from learning English	-529		319						
20 Studying English is important to me in order to attain a higher social respect	520								
32 Do you always look forward to English classes?		823							
29 Do you find learning English really interesting?		792							
40 Do you really enjoy learning English?		786							
1 I would like to spend lots of time studying English		608						327	
39 Do you think universities and companies should value English test scores		476				415			
10 Studying English is important to me because it offers a new challenge in my life		413							
35 Do you like the atmosphere of your English classes?		405	356						
4 English tests have had a positive effect on my learning		314	311						
16 My score on English tests reflects my English ability			698						
2 I think the English tests I have taken so far are fair			663						
9 English tests allow me to see how well I am doing in English			637						
26 I trust my English test scores			631						
23 My English test scores make me want to study English harder		354	405						
36 Do you like the people who live in English speaking countries?				940					
34 Do you like meeting people from English speaking countries?				782					
38 Would you like to know more about people from English speaking countries?				715					
31 Would you like to travel to English speaking countries?				376					
33 Do you think it is important for everyone to learn English?				342					
18 I can imagine myself living abroad and having a discussion in English					867				
8 I can imagine myself speaking English as if I were a native speaker of English					825				
22 I imagine myself as someone who is able to speak English					641				
37 Do you think everyone needs to study English in order to be successful?						654			
30 Do you think people can be successful without learning English?						608			
21 I think that I am doing my best to learn English							769		
6 Compared to my classmates I think I study English relatively hard							415		
5 Studying English is important because with a high level of English proficiency I can make a lot of money								654	
3 Studying English is important to me because I think it will be useful in my future career								550	
15 My family puts a lot of pressure on me to study English									628
17 Learning English is necessary because people around me expect me to do so	425								461
7 My parents/family believe(s) that I must study English to be an intelligent person									415

Extraction Method Principal Axis Factoring
 Rotation Method Promax with Kaiser Normalization
 Note Factor loading less than .30 are not shown