5-6-2019

Assessing English Writing in Multilingual Writers in Higher Education: A Longitudinal Study

Minkyung Kim

ASSESSING ENGLISH WRITING IN MULTILINGUAL WRITERS IN HIGHER EDUCATION: A LONGITUDINAL STUDY

by

MINKYUNG KIM

Under the Direction of Scott Crossley, Ph.D.

ABSTRACT

English writing skills are important components of multilingual students' successful academic performance in English-medium higher education. However, little research has been conducted on how multilingual writers develop their English writing skills over time in higher education. Thus, the purpose of the dissertation was to investigate the longitudinal development of English writing for multilingual students in higher education in relation to language skills and knowledge (vocabulary and reading), cognitive skills and knowledge (attention, working memory, and general knowledge), and language features (academic word use and language burst lengths [i.e., the number of characters produced between pauses]).

Seventy-seven multilingual undergraduates at a US university participated in two sessions with an at least five-month interval. They were from various countries including China, India, Mexico, and Zimbabwe. The students produced persuasive essays in English and took English reading and vocabulary tests on two occasions. They also completed an attention task, a working memory capacity task, and general knowledge test at the initial time of measurement. A writing process feature was captured by mean burst lengths. A written product feature was characterized by the production of academic words. Latent change score models were used.

Four main findings are reported. First, multilingual students' gains in English writing scores tended to rise as a function of lower initial levels of English writing scores, English reading scores, general knowledge scores, and academic words found in essays. This supports a "poor get richer" scenario rather than "rich get richer," such that initial lower levels may leave greater potential for gains in writing scores. Second, gains in English writing scores co-occurred with increases in academic words and gains in English reading scores. This indicates the positive longitudinal relationships of writing with reading and vocabulary use. Third, greater gains in writing scores were related to higher levels of working memory capacity, which suggests that working memory capacity is important in learning-to-write processes. Lastly, the presence of a latent variable of English literacy indicated by English writing, reading, and vocabulary was supported over time, providing a parsimonious understanding of English-literacy related variables. Theoretical and pedagogical implications are discussed.

INDEX WORDS: Multilingual Writer, English Writing, Cognitive Models of Writing, Assessing Writing, Latent Change Score Modeling

ASSESSING ENGLISH WRITING IN MULTILINGUAL WRITERS IN HIGHER EDUCATION: A LONGITUDINAL STUDY

by

MINKYUNG KIM

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2019

ASSESSING ENGLISH WRITING IN MULTILINGUAL WRITERS IN HIGHER EDUCATION: A LONGITUDINAL STUDY

by

MINKYUNG KIM

Committee Chair: Scott Crossley

Committee: Lee Branum-Martin

Sara Cushing

YouJin Kim

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2019

DEDICATION

To my dad, mom, and brother for their unwavering love and support

ACKNOWLEDGEMENTS

My journey of pursuing this degree would not have been possible without the support and encouragement from many individuals along this journey.

First of all, I would like to express the deepest appreciation to my advisor, Dr. Scott Crossley. Scott has inspired me to learn about quantitative methods and conveyed a spirit of excitement for research. Scott has offered hours and hours of his time to guide me, read hundreds of pages of my writing, and provided countless pieces of valuable feedback and advice over the past four years. Scott has also believed in me like nobody else. All of his persistent support has led me to grow personally, academically, and professionally. It was a true privilege to work with you and be your advisee, Scott, and I look forward to continuing to work with you.

I am also deeply indebted to my dissertation committee members, Drs. Lee Branum-Martin, Sara Cushing, and YouJin Kim, for their reading my drafts and providing constructive feedback. I am also grateful to Lee for guiding me with every step of statistical analysis in the dissertation project and answering dozens of my statistical questions. Beyond the dissertation, I am also thankful to Sara for inspiring me to be interested in language assessment and being willing to share her insights and ideas whenever I have questions related to assessment. I am also very thankful to YouJin for providing me with personal and academic guidance over the past five years and being my role model as a mentor and researcher.

I also very much appreciate all of my professors at Georgia State University. I would like to express my special thanks to Drs. Diane Belcher and Ute Römer, who have been very supportive since the first day of my PhD program, were inspiring during course work, and have always been cherished mentors. I am also very thankful to Ute for giving me countless

encouraging words, being available whenever I would like to chat with her, and greatly motivating me to stay healthy both mentally and physically.

Many thanks should also go to all of my friends and colleagues at Georgia State University for their friendship throughout my studies: Ali, Analynn, Cindy, Dave, Eunice, Hyejin, James, Jessica, John, Julie, Katia, Kris, Louise, Mackenzie, Meredith, Rurik, Ruth, Sally, Sanghee, Selahattin, Stephen, Susie, Yeon Joo, Yu Kyoung, and Ziyi. I am particularly grateful to Jessica for giving me wonderful memories and walking with me when I needed support.

Lastly, I gratefully acknowledge the participants who generously spent several hours for my dissertation. I also thank Jamila for scoring student essays and sharing insights about the essays. In addition, I would also like to extend my gratitude to those organizations that funded my dissertation research: The International Research Foundation for English Language Education (TIRF), Educational Teaching Service (ETS), and Georgia State University.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

College and university life can create considerable challenges for students, including academic adjustment (i.e., fitting in within an academic context; Ramsay, Barker, & Jones, 2006) and responsibility for one's physical, mental, and financial well-being (Clark, 2005). For students whose first language (L1) is not the language of instruction in higher education, college life can present additional challenges, including the use of the second language (L2; the language of instruction in higher education) and cultural barriers (Sherry, Thomas, & Hon, 2010). In addition, the population of multilingual students (i.e., students who are proficient in more than one language) in higher education has been becoming larger and more complex (Ferris, 2016). In the U.S.A., international students (i.e., individuals enrolled in higher education who are on temporary student visas; Andrade, 2006) have been increasingly enrolled in higher education (Institute of International Education, 2016). Furthermore, a greater number of US-educated multilingual students, including immigrants and Generation 1.5 students (i.e., the children of first-generation immigrants), have begun to attend US higher education institutions (Ferris, 2016; Roberge, Siegal, & Harklau, 2009). Throughout this dissertation, the term *multilingual student (or multilingual writer)* is used as a broad, neutral term that describes individuals who have started learning English as an additional language to include international students, immigrants, and Generation 1.5 students. The term *multilingual* was used to indicate that students' previous language experience is considered as a resource rather than a language deficit (Canagarajah, 2002; Kramsch, 2009).[1]

[1] The terms, *English as a second language (ESL)* or *second language (L2),* were not used to describe participants in this dissertation. This is because these terms tend to have negative connotations, such as lower proficiency, but participants in the dissertation included very proficient English speakers. Thus, ESL or L2 does not adequately describe the participants in this dissertation. However, when discussing previous studies, I followed authors' original wording and more commonly used terms (e.g., ESL and L2).

In higher education, multilingual students need a variety of skills in order to successfully adapt to academic contexts, such as understanding lectures, participating in discussions, communicating with advisors, reading academic materials, and producing academic writing (Andrade, 2006; Baird & Babb, 2014; Evans, Anderson, & Eggington, 2015; Ramsay et al., 2006; Zhang & Mi, 2010). Among various academic skills, English writing skills are crucial for successful academic performance in English-medium higher education (Baird & Babb, 2014; Evans et al., 2015; Tang, 2012). If multilingual students cannot express their ideas in written forms, they are less likely to successfully complete writing assignments and take written exams. Research has also indicated that higher-levels of writing skills predict better academic performance (i.e., grade point average; Andrade, 2006; Ramburuth, 2001), which in turn is predictive of student retention rates (Finnie, & Qiu, 2008). Furthermore, the development of writing skills in higher education is important because it forms a basis for enduring writing practices in a post-tertiary life (e.g., occupational achievement; Baird & Babb, 2014).

Despite the essential roles of writing skills in higher education and an increasing number of multilingual students enrolled in higher education, surprisingly little research has been conducted on how multilingual students develop their multilingual writing skills over the course of postsecondary education. In addition, while research has investigated the important role of language knowledge (e.g., vocabulary knowledge) in the longitudinal development of L2 writing (Schoonen, van Gelderen, Stoel Hulstijn, & de Glopper, 2011), less attention has been drawn to the roles of cognitive skills (e.g., attention) and language features (e.g., lexical sophistication) on the longitudinal development of multilingual writing. Furthermore, while early influential writing models (e.g., Hayes & Flower, 1980) have been tested in multilingual contexts (Weigle, 2002), more recent and sophisticated writing models (e.g., Hayes & Berninger, 2014) have not

been assessed in multilingual contexts. Additionally, the longitudinal development relationship among writing, reading, and vocabulary knowledge in English in multilingual writers is not clear. To address these research gaps, this dissertation will examine the longitudinal development of English writing in multilingual writers in relation to language knowledge, cognitive skills, and language features in higher education.

Thus, the main purpose of the dissertation is to examine English writing development in relation to a range of individuals' linguistic and cognitive resources in multilingual students in the U.S.A. In investigating writing development, the dissertation follows a cognitive definition of writing that involves "the use of products and principles of the writing system to get at the meaning of a written text" (Snow, Burns, & Griffin, 1998, p. 42). From this perspective, writing is considered a multifaceted cognitive process which involves a range of component skills that will lead to producing meaning in text. Component skills include both language processes (e.g., vocabulary knowledge) and cognitive processes (e.g., inferencing), which can help multilingual learners produce a coherent and elaborated text (Berninger et al., 2012; Hayes & Berninger, 2014).

This dissertation has three main purposes: (a) examine the relationship between general cognitive resources and English writing scores over time with intervals of at least five months by examining links between general cognitive resources that include attention, long-term memory, working memory, and reading and writing scores; (b) investigate the relationship between outcomes of the translator (i.e., turning verbal ideas into written text) and English writing scores over time; and (c) examine the longitudinal relationship among English writing, reading, and vocabulary.

The dissertation will provide important theoretical knowledge in three main aspects. First, the dissertation will be informed by a recent model of writing (i.e., Hayes & Berninger, 2014) that has not been examined in multilingual contexts. Second, beyond previous studies that have examined longitudinal writing development in young English learners (Schoonen et al., 2011), the dissertation investigates the longitudinal development of English writing in adult multilingual students in higher education. Lastly, the dissertation will shed light on important language, cognition, and language features that can predict the longitudinal development of English writing skills. In sum, the dissertation will contribute to expanding our understanding of English writing development that may involve a complex array of language and cognitive skills in adult multilingual writers.

The dissertation will also provide important pedagogical implications in two main ways. First, findings of the dissertation will present systematic assessment data on English writing development, which in turn can be used to make better informed decisions on English writing programs for multilingual writers in higher education. Second, findings of the dissertation will also help writing instructors offer international students appropriate support programming and services based on a clear understanding of their difficulties related to English writing skills (Zhang & Mi, 2010). In all, awareness of English writing issues and implementation of appropriate programming and services for multilingual students may help to improve not only academic performance for multilingual students but also retention rates at host institutions in higher education.

2 LITERATURE REVIEW

In this section, relevant previous research is presented according to five themes: (a) multilingual writers in higher education; (b) cognitive models of writing; (c) roles of English

linguistic knowledge on English writing development; (d) roles of cognitive skills on English writing development; and (e) language features and English writing quality.

2.1 Multilingual Writers in Higher Education

The population of multilingual writers in higher education settings has been becoming larger and more complex. In the U.S.A., international students have increasingly been present on US college campuses (Institute of International Education, 2016). US-educated multilingual residents, including the children of first-generation immigrants (i.e., Generation 1.5 students), have also been enrolled in greater numbers in US higher education since the 1790s (Harklau, Losey, & Seigal, 1999). These international and Generation 1.5 students comprise two main subgroups of multilingual writers in higher education, attending various degree programs across a range of academic disciplines (Ferris, 2009, 2016; Roberge, Siegal, & Harklau, 2009). These two groups of multilingual writers are considered to have distinct characteristics (Belcher, 2012; Ferris, 2009, 2011; Matsuda, 2008). Previous research on each of the two groups (i.e., international and Generation 1.5 students) followed by the comparison between the two is discussed below. Based on this line of discussion, the problems in defining L2 students in higher education in the U.S.A. is also discussed.

2.1.1 International students

International students are defined as individuals who are enrolled in higher education on temporary student visas and thus distinguished from non-native immigrants and citizens (Andrade, 2006). International students in higher education may study abroad for a short term (typically less than one year) to improve intercultural communication and/or study another language, or for a long term to earn a degree. The largest population of international students study abroad in the U.S.A. (Institute of International Education, 2018). Furthermore, in the

U.S.A. an increasing number of international students have been enrolled in higher education in the last few decades (Wu, Garza, & Guzman, 2015). Many US higher education institutions have also actively recruited international students (Bartlett & Fischer, 2011). In the 2017/18 academic year, 1,094,792 international students (an increase of 1.5% over the prior year) were enrolled in US colleges and universities, which made up 5.5% of the total enrollment in U.S. higher education (Institute of International Education, 2018). In the 2017/18 academic year, around 60% of international students were from China (33.2%), India (17.9%), South Korea (5%), and Saudi Arabia (4.1%).

For international students to be successful in English-medium higher education settings, English proficiency is vital (Andrade 2006; Evans et al., 2015; Sherry et al., 2010). To address English proficiency-related issues, host universities often set an appropriate English proficiency level for admission, such as a minimum score of the Test of English as a Foreign Language (TOEFL). However, achieving the minimum score of English proficiency for college entrance does not necessarily guarantee that international students are competent with the English language in academic contexts (Andrade, 2006; Wan, 2001). Although international students pass an English proficiency examination for admission, they may confront many cultural, language, and social challenges (Andrade, 2006; Evans & Andrade, 2015; Ferris, 2009). Specifically, international students may experience struggles in negotiating "a new range of sociocultural situations such as faculty office hours, team work, public presentations, and frequently, independent living" (TESOL, 2010, p. 1), which may substantially differ from those of their countries. In a review of factors that influence international students' adjustment to US higher education, Zang and Goodson (2011) found that the frequently reported factors included

stress, social support, English language proficiency, acculturation, self-efficacy, personality, country of origin, and social interaction with native English speakers.

Among many challenges that international students confront in higher education, language proficiency is considered one of the serious academic issues which prevent adjustment for international students (Galloway & Jenkins, 2009). A lack of proficient English skills has negative impacts on international students' academic achievement, class participation, social interaction with classmates and professors (Andrade, 2006; Yeh & Inose, 2003). For example, Stoynoff (1997) reported modest correlations between international undergraduates' TOEFL scores and their academic achievement (as measured by GPA, credits completed, and number of withdrawals). Terui (2012) found six international students in an US university tended to pretend to understand contents of interactions with native speakers (including their professors and classmates) to compensate for their limited English proficiency and to overcome difficulties in conversing with native speakers.

With respect to international students' adjustment challenges in higher education, different perspectives between international students and professors have also been reported. Robertson, Line, Jones, and Thomas (2000) reported that in an Australian university, international students considered their difficulty in class participation due to their language proficiency issues, while professors attributed this difficulty to be cultural. Robertson et al. (2000) also reported that international students had difficulty in understanding professors' use of colloquial English and considered their professors uninterested in their learning. On the contrary, professors felt that international students had insufficient critical thinking abilities and weak writing abilities and did not take responsibility for their own learning, although students showed a willingness to employ self-help strategies and to improve English language skills. Thus, while

both professors and international students in higher education likely agree that one of the main difficulties experienced by international students is English language proficiency (Galloway & Jenkins, 2009, Robertson et al., 2000), they may criticize each other of not taking their own responsibilities as teachers and students, respectively.

In addition to language-related issues, English written composition is crucial for the success of international students (Evans et al., 2015; Tang, 2012). However, English writing performance can be more challenging to international students than English oral performance. This is partly because English listening and speaking skills can often develop naturally through repeated exposure to the English language in English-speaking environments, whereas for English writing skills, opportunities to practice and to receive feedback from experts are considered more important than exposure to the English (Storch & Hill, 2008). Indeed, research has found that many international students have difficulties in drafting writing assignments in English academic contexts (Andrade, 2006, 2009; Ramsay et al., 2006; Sherry et al., 2010).

To aid international students' English writing development, host institutions have adopted various approaches, including providing English writing classes and writing centers for international students (Andrade, 2006; Sherry et al., 2010). In turn, these English writing classes and writing centers provided by host universities were perceived helpful by international students for their development in English writing skills (Andrade, 2009; Lawrick & Esseili, 2015; Ramsay et al., 2006; Zhang & Mi, 2010). Despite host institutions' approaches, English writing difficulties experienced by international students may also be linked to cultural differences and institutional responsibilities. Specifically, Fox (1994) found that international students' written assignments were considered inadequate by their professors because the professors often did not recognize different cultural communication styles. Holmes (2004) reported that international

Chinese students in a New Zealand university tended to be accustomed to employing indirect writing styles and unaccustomed to producing critical analyses of arguments due to their cultures in which directness and criticisms are considered as unacceptable communication practices. In addition, Lee (2018) analyzed ten narratives of Chinese international students in an US university who failed an ESL writing course, and suggested that their failure may not simply due to students' lack of responsibility or persistence, but also due to other systematic factors, such as instructors' failure to communicate to the students, a campus climate in which international students are not welcome, and the lack of institutional support systems that cater for international students' needs.

In short, international students' cultural, language, and social challenges, including English writing difficulties, have been well reported in past research. However, to my knowledge, assessment of longitudinal development in English writing in relation to cognitive and language skills in international students has not been systematically conducted.

2.1.2 Generation 1.5 students

In general, Generation 1.5 students are US-educated children of immigrants who began learning English at their early ages and have attended all or part of their formal education in schools in the U.S.A. (Harklau, Losey, & Siegal, 1999). These students began to appear in U.S. colleges and universities in the 1970s (Ferris, 2016). One of the most recent definitions narrowly describes Generation 1.5 students as those who speak a language that is not English with their family, have received five or more years of education in the U.S.A., are less than 22 years old, and graduated from a US high school (or passed a high school equivalence test; Doolan, 2017, p. 2). While the term *Generation 1.5 students* is a commonly used term in L2 research based on work by Rumbaut and Ima (1988), other terms also include *early-arriving resident students*

(Ferris, 2009, 2011), *US-educated multilingual writers* (Nakamaru, 2010), and *resident nonnative speakers of English* (Levi, 2004). Additionally, among Generation 1.5 students, there may be differences by length of residence in the USA: US-born citizens, early arrived residents, recently arrived residents (Ferris, 2009; Roberge, Losey, & Siegal, 2009).

Generation 1.5 students in higher education are considered as a type of language learners distinct from native students (Ferris, 2009; Roberge, 2009). While Generation 1.5 students tend to learn English naturally by being exposed to it in immersion settings rather than receiving formal instruction (Reid, 1998), their language learner status has an influence on their educational experiences, such as taking ESL courses rather than mainstream courses (Ferris et al., 2011). In higher education, unlike native students who are learners of university-level academic writing only, Generation 1.5 students need to undertake both learning of the English language and that of academic writing (Ferris 2009). Generation 1.5 students may also have struggles in transitions between high school and college (Allison, 2009; Harklau, 2000), identity negotiations resulting from co-existing multiple cultural and language identities (Chiang & Schmida,1999), and difficulties in achieving academic success in college (Muchinsky & Tangren, 1999).

In past research, as compared to international students enrolled in higher education whom L2 researchers have widely drawn attention to, Generation 1.5 students enrolled in higher education have not been fully recognized in either L1 or L2 studies. L2 researchers tend not to focus on Generation 1.5 students due to their absence in college ESL courses. Generation 1.5 undergraduate students who have long-term U.S. residence and education likely resist being placed in ESL writing classes with international students (Ortmeier-Hooper, 2008) because they are reluctant to be labeled as ESL (Thonus, 2003) and may carry over “the stigma associated

with K-12 ESL" to college ESL courses (Lawrick & Esseili, 2015, p. 86). Likewise, L1 writing researchers have not been much interested in Generation 1.5 students due to their English language learner status (Thonus, 2003).

Generation 1.5 students also receive minimal attention from higher education institutions. For example, while the population of international students in higher education has been well surveyed by the Institute of International Education (e.g., around 5% of the total enrollment in U.S. higher education in the 2017/2018 academic year), the number of Generation 1.5 students in U.S. higher education is unknown (Andrade et al., 2015). In addition, because Generation 1.5 students in higher education have varying level of English language proficiency, some of them may not have sufficient English language proficiency necessary in academic contexts (Ferris, 2009). However, colleges and universities typically do not require Generation 1.5 students to take English language proficiency tests for admission or to establish their English language proficiency prior to admission. Furthermore, ESL (or multilingual) composition courses in higher education mainly consider international students, and barely address Generation 1.5 students' linguistic needs (Evans & Andrade, 2015).

Empirical research on Generation 1.5 writing has only recently began to emerge (di Gennaro, 2013; Doolan, 2017; Levi, 2004). This line of studies has compared writing performances, such as holistic writing quality and error patterns, produced by Generation 1.5 students with those produced by L1 or/and international students. Generally, researchers agree that Generation 1.5 writers are a type of language learner due to error patterns found in their writing (di Gennaro, 2009; Doolan & Miller, 2012; Eckstein & Ferris, 2018; Levi, 2004) and their "varied and inconsistent" grammar (Holten, 2009, p. 179) as in L2 students' grammar. For example, Eckstein and Ferris (2018) analyzed writings of L1 and L2 students (mostly Generation

1.5 with long-term residence in the U.S.A.), and found that L2 writers received lower holistic scores of language use and made more errors than L1 students. On the contrary, a few researchers have found that Generation 1.5 writing are similar to L1 writing. For example, Doolan (2017) compared writing of early arrival Generation 1.5, L1, and L2 students, and reported no significant differences in the number of total errors in student writing between early arrival Generation 1.5 and L1 writing, suggesting that early arrival Generation 1.5 students may need to be described as L1 writers or bilinguals (when students have strong L1 and L2 language skills).

In sum, while Generation 1.5 students are increasingly present in U.S. colleges and universities, they have relatively been underrepresented in the L1 and the L2 literature and by higher education institutions in general. In addition, despite recent research on Generation 1.5 writing, assessment of longitudinal English writing development in Generation 1.5 students in higher education has not been conducted yet.

2.1.3 Comparing international and Generation 1.5 students

Both international and Generation 1.5 students are generally considered L2 learners (Eckstein & Ferris, 2018; Evans et al., 2015; Ferris, 2009; Harklau et al., 1999). While sharing characteristics of language learner status, Generation 1.5 and international students bring different educational and cultural backgrounds to higher education in at least four aspects. First, in terms of academic and socio-economic status, Generation 1.5 students may have been behind academically in their elementary and secondary school years due to their limited English language proficiency, whereas international students may have been high-performing and socioeconomically advantaged students in their countries of origin (Collier, 1987; Thonus, 2003). Second, while Generation 1.5 writers may partially or barely develop L1 literacy skills,

international students tend to have higher levels of L1 literacy skills (Harklau et al., 1999). Third, Generation 1.5 students tend to be "ear learners" with being fluent in oral and aural communication, whereas international students tend to be "eye learners" relying on grammar rules and written communication (Reid, 2005). Lastly, Generation 1.5 students are familiar with US educational systems, cultures, conversational language, and slang (Harklau et al., 1999), but international students may not.

In higher education, higher levels of writing skills are linked to higher academic results (Kobrin, Patterson, Shaw, Mattern, & Barbuti, 2008; Ramburuth, 2001). As compared to L1 students, both international and Generation 1.5 students tend to receive lower writing scores and produce more errors in their writing (Doolan & Miller, 2012; Eckstein & Ferris, 2018. Levi, 2004; Ramburuth, 2001). When comparing writing of international and Generation 1.5 students, mixed findings are reported. In holistic quality, Doolan (2017) found no difference in holistic scores of timed essays written by international students (most of them had spent less than two years in U.S. schools) and Generation 1.5 students (most of them had spent more than 10 years in U.S. schools), indicating that these two groups produced essays of similar quality. In contrast, di Gennaro (2013) found that international students (most of them had spent less two years in the U.S.A.) received higher scores of timed essays on average than Generation 1.5 students (most of them had spent more than five years in the U.S.A.), indicating that international students produced better essays than Generation 1.5 students. In terms of error patterns, research generally found that errors produced by international students are different in types from those by Generation 1.5 students (di Gennaro, 2013; Levi, 2004). In terms of error quantity, international students tend to produce greater errors than Generation 1.5 students (Doolan, 2013, 2014, 2017; Levi, 2004).

2.1.4 Problems in defining L2 students in higher education in the U.S.A.

In higher education in the U.S.A., international and Generation 1.5 students have often been considered as main groups of L2 learners. For example, a recent study by Eckstein and Ferris (2018) categorized both international students and Generation 1.5 students in the U.S.A. into a single L2 group. However, using these terms of international and Generation 1.5 students along with L2 students likely results in at least three problems in defining the inclusion and exclusion criteria of the respective populations.

First, the term Generation 1.5 involves individuals who vary substantially in their English language proficiency levels from very high proficiency English speakers (i.e., early arrivals) to low proficiency English speakers (i.e., late arrivals). While it seems appropriate to consider late arrival Generation 1.5 students as L2 learners, it may not be accurate to describe early arrival Generation 1.5 students who are English dominant throughout childhood and formal schooling as L2 learners. Instead, it may be most appropriate to describe early arrival Generation 1.5 students as L1 speakers, or bilinguals/multilinguals if they are skillful in English and another language (Doolan, 2017).

Second, as in Generation 1.5 students, although there are considerable variations in English proficiency levels, international students are generally considered as English language learners. However, not all international students are learners of English. For example, it may not be accurate to describe those who are from counties in which English was spread as a colonial language through imperial expansion in Asia and Africa, such as India, the Philippines, Nigeria, and Tanzania (Kachru, 1992) as English learners. In these countries, English is an official language in administration and education. Thus, international students who have been educated

in these countries are likely to be English dominant throughout formal schooling and have high proficiency, and may be best described as bilinguals/multilinguals, rather than L2 learners.

Lastly, as international and Generation 1.5 students are defined with different criteria, there are students who can be ambiguously categorized into both. For examples, according to Doolan's (2017) definition, students who have received five or more years of education in the U.S.A., are less than 22 years old, and graduated from a US high school are considered Generation 1.5 students. However, if these students have stayed in the U.S.A. for more than five years on a student visa only, they can also be described as international students. Thus, describing L2 learners with these two terms (i.e., international and Generation 1.5) does not seem to be accurate enough to reflect the complex nature of these populations.

Taken together, although international and Generation 1.5 students are commonly used to describe L2 students who are present on U.S. campuses, each of these two terms inadvertently includes a broad range of individuals who vary substantially in English proficiency and English learning backgrounds. Thus, alternative notions and terms may be needed. Indeed, in L2 literature, to avoid a simplified, dichotomous understanding between native/L1 and non-native/L2 speakers, alternative terms have been suggested, such as *language expertise* (Rampton, 1990), *more or less accomplished* (Edge, 1988), and *proficient users of English* (Paikeday, 1985).

While these alternative terms consider language proficiency or expertise as a core criterion to distinguish English speakers, it seems to fail to take into account the existence of variations in language proficiency, particularly in the areas of written language, which exist even in monolingual L1 speakers. To explain this point, Hulstijn's (2015) distinction between Basic Language Cognition (BLC) versus Higher Language Cognition (HLC) is helpful. BLC is the

language, mainly involving the spoken modes, pertaining to simple every-day matters, and consisting of high-frequency vocabulary items and syntactic structures, which all native speakers have commonly acquired. On the other hand, HLC is the language involving both written and spoken modes, pertaining to more topics addressed in schools and work places, and reflecting educational and social profiles (e.g., level of education). L1 learners likely show much larger individual variations in HLC discourse than BLC discourse, such that all L1 speakers are competent in BLC domains, while they may have different levels of competence in HLC domains.

In the context of this dissertation (i.e., English-medium higher education in the U.S.A.), international and Generation 1.5 students can be reconceptualized in terms of their acquisition of BLC and their educational backgrounds. It is likely that BLC is acquired by early arrival Generation 1.5 students as well as international students who had been educated via the English language in their countries (e.g., India). These students may perform at ceiling in BLC discourse. In contrast, international students who had been educated via languages other than English in their home countries (e.g., South Korea) may not fully acquire BLC discourse, showing individual variations. However, in terms of HLC, it seems that both international and Generation 1.5 students will show individual variations in HLC discourse as a function of their educational backgrounds.

Based these conceptualizations, the current dissertation suggests two distinctive groups for all individuals in US higher education (not only international and generation 1.5 students but also monolingual English speakers and bilinguals) based on their educational backgrounds. One group includes students who have been educated via the English language at least six years (i.e., equivalent to around a half of the 12 years of elementary and secondary education) and thus are

expected to have acquired BLC discourse. The other group includes students who have been educated via the English language less than six years and are expected to have individual variations in BLC discourse. Importantly, both groups are expected to have individual variations in HLC discourse. Table 2.1 shows the proposed characteristics of these students.

Table 2.1 Two Main Groups of Individuals in US Higher Education

Characteristics	Students educated via the English language	Students not educated via the English language
Acquisition of BLC domain	Acquired most or all of the BLC features	Acquire some or many features of BLC
Acquisition of HLC domain	Acquire some or many features of HLC	Acquire some or many features of HLC
Examples	Monolingual English speakers, bilinguals of English and another language, early arrival Generation 1.5 students, and international students educated via the English language	Late arrival Generation 1.5 students, and international students educated via languages other than English

While this distinction is not of primary interest in the dissertation and may inadvertently lead to overgeneralization, it can be meaningful in that international and Generation 1.5 students can be described based mainly on educational backgrounds (i.e., length of English immersion years) rather than nationality, ethnicity, or residency status.

2.2 Cognitive Models of Writing

From a cognitive perspective, writing is a process through which a writer creates meaning in written form (Murray, 1980). Many scholars have proposed writing-specific, cognitive models that consider the unique processes and operations of writing (Berninger, 2000; Berninger et al., 2002; Hayes & Berninger, 2014; Hayes & Flower, 1980). These models indicate that writing involves various cognitive processes, such as retrieving linguistic knowledge, generating ideas, and evaluating. Three influential models of writing are discussed below: an early model of Hayes and Flower (1980), the Simple View of Writing (Berninger et al., 2002), and a more recent

model of Hayes and Berninger (2014).

2.2.1 Hayes-Flower model (1980)

One of the most influential models of writing is the Hayes-Flower cognitive model of adult skilled writing (Hayes & Flower, 1980) as shown in Figure 2.1. In this model, three major elements are identified: the writer's cognitive writing processes, the writer's long-term memory, and the writer's task environment. Arrows indicate the information transfer. The cognitive writing processes include four main processes involved in writing: planning, translating, revising, and monitoring. Planning is to decide what to say ideas. The input of planning includes the writing assignment and the writer's long-term memory, while its output is a bunch of conceptually generated ideas. The sub-processes of planning are generating, organizing, and goal setting. Translating (i.e., text generation) is to turn ideas and plans into written text (i.e., the text produced so far). Revising takes the text produced so far as input and modifies it for improvement. The monitor appearing as a box parallel in status to the three writing process boxes is viewed as a process that coordinates planning, translating, and revising. The writer's long-term memory includes knowledge of topic, audience, and stored writing plans. The task environment includes all of the factors that influence the writing process beyond the writer, such as social factors (e.g., the writing topic and audience) and physical factors (e.g., the text the writer has produced so far). During text production, the writer participates in the recursive process of writing (i.e., planning, translating, reviewing, and monitoring), simultaneously utilizing his or her long-term memory (e.g., knowledge of topic and knowledge of audience) and considering the task environment (e.g., writing assignment and the text produced so far). Importantly, these writing sub-processes are considered interactive and non-linear.

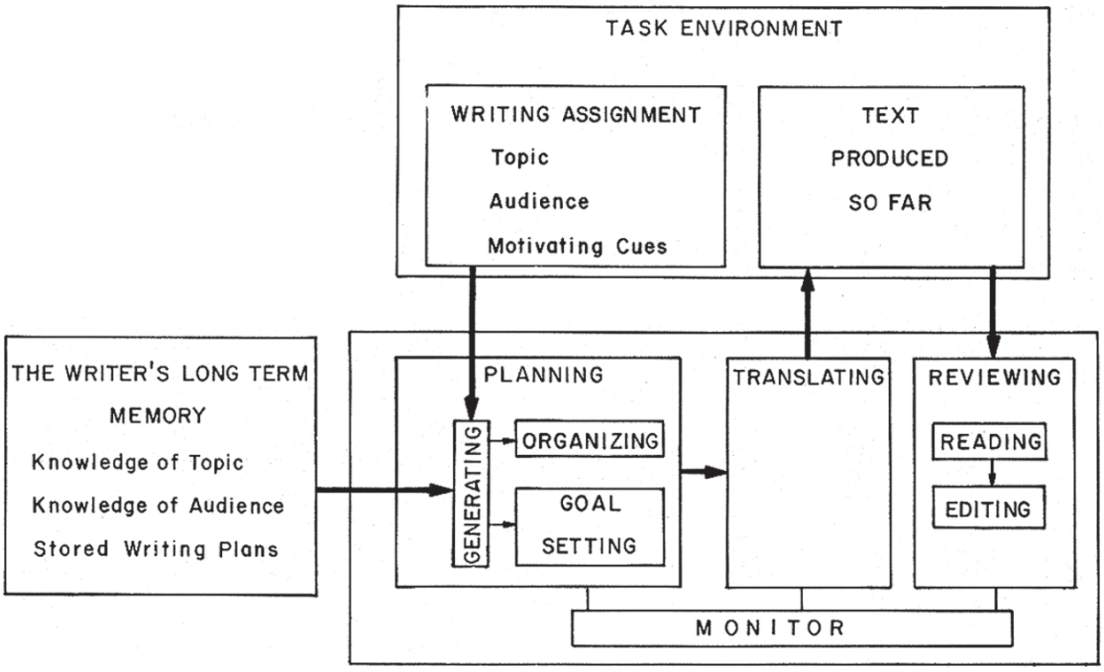


Figure 2.1 The Hayes-Flower model (Hayes & Flower, 1980).

2.2.2 Simple View of Writing

The Simple View of Writing was proposed by Berninger and colleagues (Berninger, 2000; Berninger et al., 2002). It speculates that transcription and text generation are crucial in writing. Transcription is a low-level, basic writing process which needs to be mastered from early on. Via transcription, writers externalize language in written form. Transcription involves two main processes: spelling (i.e., retrieving and selecting orthographic symbols from memory) and handwriting/typing (i.e., executing motor movements with a writing tool to produce orthographic symbols; Abbott & Berninger, 1993). Spelling is typically measured by accuracy, such as counting correctly spelled words in isolation or in text. Handwriting or typing fluency is often measured by both accuracy and speed, such as calculating letters produced accurately and quickly within a given time. Text generation (i.e., ideation by producing words, sentences, and discourse) is a high-level writing process which involves the translation of verbal ideas into written language representations. Text generation also involves linguistic processes, such as

retrieving appropriate words and encoding syntactic structures. While transcription processes are fast and do not require many attentional resources after being automatized, text generation processes are more cognitively demanding and require conscious effort. Furthermore, the Simple View of Writing emphasizes the crucial role of executive functions (e.g., attention, planning, translating, revising, and monitoring) in a working-memory environment in which the writer creates a text via transcription and idea generation because these functions occur within the constraints of working memory (Berninger, 2000). In writing, mastering transcription skills are important because automatic low-level transcription lessens cognitive demands on limited-capacity working memory, enabling executive function resources to be more directed towards higher-level idea and text generation.

2.2.3 Hayes-Berninger model (2014)

Since the first model proposed by Hayes and Flower (1980), Hayes and colleagues have revised and expanded the writing models by adding other elements, such as working memory and transcription, and referring to a number of empirical studies and modeling (e.g., Hayes, 1996, 2012; Hayes & Berninger, 2014; Hayes & Chenoweth, 2006). A recent model was proposed in Hayes and Berninger (2014; see Figure 2.2). The Hayes-Berninger model (2014) includes three major levels of cognitive processing: the resource level, the process level, and the control level.

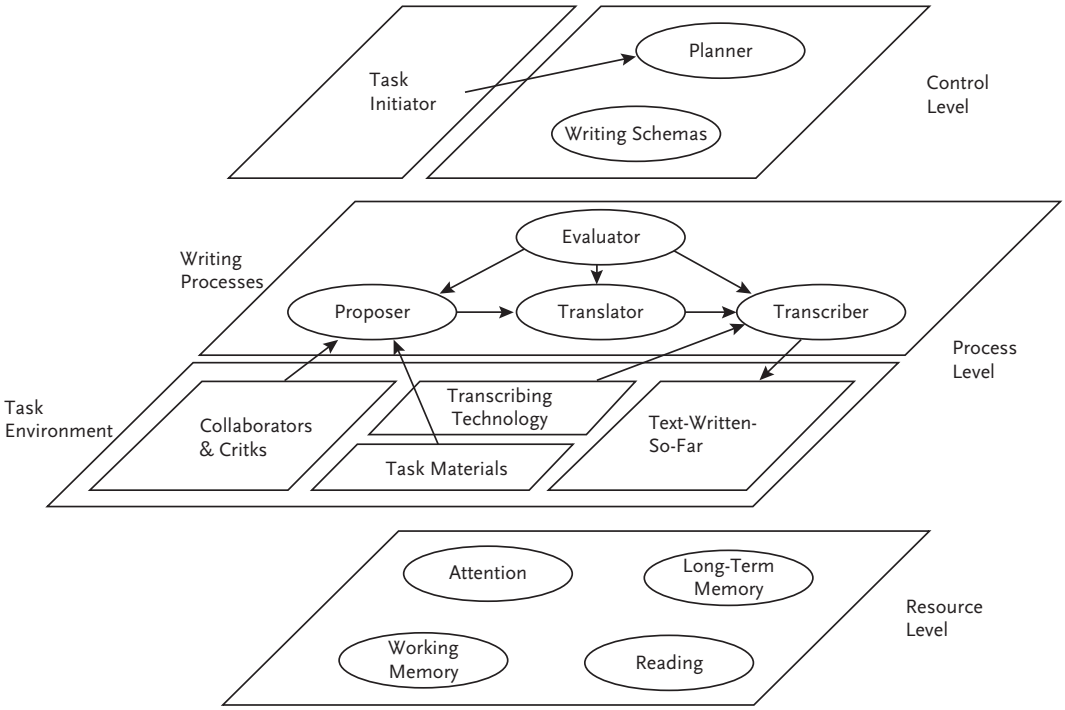


Figure 2.2 The Hayes-Berninger model (Hayes & Berninger, 2014).

The bottom, or resource, level represents general cognitive resources that writers draw on while composing, including attention, working memory, reading, and long-term memory. Attention refers to "the ability to maintain focus on a task in the face of distraction" (Hayes & Berninger, 2014, p. 4). Working memory is a memory system that stores and processes information while carrying out a task. Working memory is important in retrieving relevant knowledge from long-term memory, holding the relevant information and, simultaneously, turning ideas into written forms. Long-term memory is a complex resource which contains not only knowledge of language, including orthography, vocabulary, grammar and discourse schema, but also knowledge of the world, facts, episodes, and experiences. Long-term memory also stores the social and cultural factors, such as knowledge of the genre, the audience (e.g., how the audience would respond to text), and the social and cultural context of writing. The effective use of knowledge resources stored in long-term memory is important for proficient

writing (Chenoweth & Hayes, 2001). For example, writers with rich knowledge about the topic of the writing assignment stored in long-term memory are more likely to produce higher-quality essays. Reading is also an important resource during writing because writers typically read the text they have produced and reread it for revision and edit, which may in turn help construct cohesive text.

The middle, or process, level represents operation of, and interaction between, cognitive processes, including writing processes and the task environment. Writing processes include four main processes: a proposer, a translator, an evaluator, and transcriber. First, the proposer suggests a package of ideas that can be included in the text. The input of the proposer comes from various sources and resources, such as the task environment, long-term memory, and the text produced so far. Second, the translator transforms ideas taken from the proposal into language strings of verbal forms. It may also transform visual or auditory language strings stored in long-term memory into language strings (Hayes & Chenoweth, 2007). It is likely that the translator operates more fluently for writers with greater linguistic experience and stronger verbal working memory capacity (Chenoweth & Hayes, 2001). Third, the transcriber turns the language strings produced by the translator into written text. The less automatic process of transcription (i.e., spelling and handwriting/typing) may put demands on memory resources (Hayes & Chenoweth, 2006). Lastly, the evaluator judges the adequacy of all of the writing processes. In these writing processes, a revision process is not included because revision is considered as a specialized writing task with the aim of replacing an earlier text or idea at any levels of writing processes, including proposing, translating, transcribing, and evaluating.

Within the process level, another level is the task environment which represents the immediate social and physical factors influencing the writing processes. The task environment

includes four main factors: collaborators and critics, task materials, transcribing technology, and text-written-so-far. The collaborators and critics represent concurrent social input as writers produce text. The task materials might include an assignment sheet, a source text, and/or graphics. The transcription technology, such as handwriting and typing, can influence writing processes. For example, some writers may write substantially faster by keyboard than by hand. The text-written-so-far is also a physical factor which writers may read and reread frequently. Reading the text-written-so-far may help writers keep textual features (e.g., tense and cohesion) consistent across phrases, sentences, and paragraphs and maintain text coherence.

The top, or control, level represents factors that direct operations at the process level. It contains three factors: the task initiator, the planner, and writing schemas. The task initiator might be an instructor who gives students a written assignment in class or a writer himself or herself who wants to write a letter to friends. The planner sets writing goals. It can be as simple as the single goal of writing about an episode. It can also be specific with a sequence of topics and subtopics and tone. Writing schemas represent both genre knowledge and strategic knowledge (i.e., how the writer advances the text), though these types of knowledge might also be stored in long-term memory. Based on Berninger, Fuller, and Whitaker's modeling of children's writing (1996), and Bereiter and Scardamalia's knowledge-telling model (1987), three different strategies of writing are identified: the flexible-focus strategy, the fixed-focus strategy, and the topic-elaboration strategy. The flexible-focus strategy, the simplest one, is considered as stream-of-consciousness writing. The fixed-focus strategy is to connect every idea proposed in a single topic. The topic-elaboration strategy, the most sophisticated one, is to focus on a single main topic, but also include subtopics that elaborate the main topic. These three strategies are selected by the writing schemata, which in turn impacts the quality of writing (Hayes, 2011).

Last but not least, in the Hayes-Berninger model (2014), one of the important facets, but not explicitly represented in Figure 2.2, are language bursts. A language burst is generally defined as a chunk of letters or words produced between two consecutive pauses (Kaufer, Hayes, & Flower, 1986). Pauses reflect periods of graphomotor inactivity typically longer than two seconds (Kaufer et al., 1986; Limpo & Alves, 2017). Adults tend to write texts with an average language burst length of six to 12 words (Chenoweth & Hayes, 2001). In the Hayes-Berninger model (2014), it is claimed that language bursts are "produced through the interaction of four cognitive processes: a proposer, a translator, an evaluator, and transcriber" (p. 6), from prelinguistic ideas to language strings of verbal forms to written text. Among these cognitive processes, the translator is considered as a key source of language bursts, such that bursts are associated with the capacity of the translator for searching for appropriate linguistic forms to encode ideas (Hayes, 2009). In addition, the translator capacity is limited depending on the writer's linguistic experience and working memory resources (Chenoweth & Hayes, 2001, 2003). That is, the writer composes a text in a choppy fashion by repeating the production of one language burst followed by a pause. Importantly, each of these language burst produced occurs within the demands that the translator puts on available working memory capacity. When the limit of the translator capacity is reached, it is likely that the translation stops and then the language strings are externalized by the transcriber. Thus, it seems that translation is "the bottleneck limiting fluency" (Hayes & Berninger, 2014, p. 6).

2.2.4 Applications of cognitive writing models in L2 contexts

The cognitive models of writing proposed by Hayes and his colleagues have been dominant in North America and Europe since the early 1980s, providing insights into how individuals think and perform while composing texts. These cognitive models of writing have

also been widely applied in L2 contexts by L2 writing researchers with their own research agendas (Silva, 1993). Initial studies described general L2 writing processes (Raimes, 1987) and planning processes (Jones & Tetroe, 1987). In subsequent L2 studies directly or indirectly inspired by L1 writing process models, three main areas have been examined: L2 writers' cognitive processes, L1 use during L2 writing, and comparisons of writing processes across L1s and L2s. Each area is briefly discussed below.

First, L2 research inspired by L1 writing process models has found that L2 writers produce texts through planning, translation, revision, and monitoring processes in a similar manner as L1 writers (Raimes, 1987; Wang, 2003; Zimmermann, 2000). However, as compared to L1 writers, L2 learners tend to retrieve and process linguistic resources for encoding complex ideas less automatically and produce texts less fluently (Chenoweth & Hayes, 2001). In addition, L2 writers' limited L2 linguistic resources likely impose constraints on various writing processes, such as the difficulty in maintaining translation processes (i.e., turning ideas into language strings; Sasaki & Hirose, 1996), the attention drawn to language concerns at the expense of idea generation and text organization (Whalen & Ménard, 1995), and the difficulty in making strong arguments in academic writing (Flowerdew & Li, 2009). As L2 writers become proficient writers, they tend to be free of these language constraints, having greater flexibility in coordinating writing processes, finding appropriate lexical and syntactic forms more automatically, and easily refining their language use (Manchón et al., 2009; Sasaki, 2004).

Second, research on L2 writing processes has also found that the use of L1 during L2 writing occurs at various writing processes, such as planning, formulating, revising, and monitoring (Manchón et al. 2007). The use of L1 can be useful in generating ideas and organizing information, checking whether linguistic expressions correspond to the intended

meaning, controlling writing processes, and revising texts written so far (Cumming, 1989; Wang, 2003; Wolfersberger, 2003; Woodall, 2002). L2 writers at the beginning level may also use the L1 to compensate their limited L2 resources and obtain cognitive stability during L2 writing (Manchón et al., 2007). On the other hand, advanced L2 writers may use the L1 when carrying out cognitively demanding tasks to engage in deeper processing (van Weijen, 2009).

Lastly, when writers produce both L1 and L2 texts, similarities in writing processes across L1s and L2s have been reported, suggesting that cognitive processes during writing may be common across L1s and L2s, while differences lie in how ideas are linguistically encoded. Similar writing processes across L1s and L2s likely occur in various writing stages, such as planning and revising, and in learners from different L2 proficiency levels from bilinguals to less proficient L2 writers (Armengol-Castells, 2001; Beare & Bourdages, 2007; Stevenson, Schoonen, & De Glopper, 2006; Whalen & Ménard, 1995).

Taken together, cognitively oriented writing models represent a framework of cognition specific to writing processes, providing useful information about how writers generally create written texts and use their cognitive resources in a specific task environment. While various L1-based cognitive models of writing have been applied in multilingual contexts, the recent Hayes and Berninger writing model (2014) has not been fully tested in multilingual contexts yet. Based on the Hayes-Berninger model, the dissertation will focus on the resource-level cognitive processing (i.e., attention, working memory, long-term memory, and reading) and the role of the translator (i.e., length of language bursts).

2.3 Roles of Language Knowledge in L2 Writing

In L1-based cognitive models of writing, language knowledge and experience presumably stored in long-term memory are considered important (Chenoweth & Hayes, 2001;

Hayes, 1980; Hayes & Berninger, 2014; Hayes & Flower, 1980). Two types of L2 language knowledge have been considered important in writing processes: vocabulary knowledge and grammar/syntactical knowledge. Below, the importance of each of vocabulary and grammar knowledge in L2 research in general and L2 writing in particular is discussed, followed by discussion of comparing the two.

2.3.1 Vocabulary knowledge and L2 writing

The importance of vocabulary knowledge in L2 research has also been widely studied (e.g., Dallar, Milton, & Treffers-Daller, 2007; Milton, 2009; Nation, 2001). In general, the more words L2 learners know, the better they will be able to carry out L2 tasks. In the L2 vocabulary research, vocabulary knowledge has traditionally divided into two domains: receptive and productive knowledge (Milton, 2009; Nation, 2001). Receptive knowledge is related to words that are recognized in spoken or written input, while productive knowledge is related to words that are retrieved and expressed in speech or writing. In general, an L2 learner has receptive knowledge greater than productive knowledge. Nation (2001) further divides vocabulary knowledge into three areas: knowledge of form (i.e., knowing the orthographic, morphological and phonological form of a word), meaning (i.e., linking the word form to its meaning, and knowing concepts, referents, and associations) and use (i.e., knowing grammatical functions, collocations, and constraints on use). An additional term related to productive knowledge is fluency, which concerns how easily and quickly learners can access and produce the words in context (Dallar et al., 2007).

In writing, vocabulary knowledge is important because writing is a process of meaning making through the use of a range of words. That is, written words in the text are the outcomes of translating nonverbal ideas into language strings (Berninger, 2000; Hayes, 1980; Hayes &

Berninger, 2014). In context of L2 writing, L2 writers with greater vocabulary may be more fluent in expressing ideas without hesitations, whereas L2 writers with lower levels of vocabulary knowledge may be featured by frequent pauses and hesitations due to constraints of lexical decisions during writing processes. Indeed, research has found that rich vocabulary knowledge is an important element in successful L2 writing (Lu, 2010; Milton, Wade & Hopkins, 2010; Stæhr, 2008; Schoonen et al., 2003, 2011). For instance, Stæhr (2008) found that for Danish (L1) adolescents, English (L2) receptive vocabulary knowledge size was strongly correlated with L2 writing scores (r = .73), and around 52% of the variance in writing scores was explained by vocabulary knowledge. Similarly, Milton, Wade and Hopkins (2010) reported that for L2 adult learners of English from various language backgrounds in U.K., vocabulary size scores were strongly correlated with writing scores (r = .76), explaining around 60% of the variance in writing scores. Schoonen et al. (2003) also found a strong correlation of .63 between vocabulary knowledge and writing skills for Dutch (L1) adolescents learning English (L2). These results confirm that vocabulary knowledge is a major contributor to success in L2 writing performances.

2.3.2 Grammar knowledge and L2 writing

In addition to vocabulary knowledge, the importance of grammatical knowledge has also been extensively studied in L2 research (Cumming, 2001; Grabe & Kaplan, 1996; Purpura, 2004; Schoonen et al., 2003, 2011). Grammatical knowledge generally refers to a set of internalized informational structures of grammatical form and meaning available for language use (Purpura, 2004). Grammatical knowledge involves understanding utterances or sentences and producing those which are grammatically correct and contextually meaningful. One of the well-known conceptualizations of grammar is Larsen-Freeman's (1991) framework, which includes three

dimensions: linguistic form (i.e., morphology and syntactic patterns), semantic meaning (i.e., meaningfulness of an utterance), and pragmatic use (i.e., appropriateness of a message in context). In assessment, grammatical knowledge is often measured by morphosyntactic features (e.g., verb tense and subject-verb number agreement) without the need to find the intended word to minimize the test-taker's involvement in vocabulary knowledge (Schoonen et al., 2011).

In writing, grammatical knowledge is considered important in not only translating ideas into syntactic strings of expressions, but also monitoring structures that have already produced and are currently produced in the process of reviewing the text produced so far (Hayes, 1980; Hayes & Berninger, 2014). In addition, writers' ideas are expressed by grammatical structures beyond the use of single words. Consequently, writers need to have knowledge of how to put words into longer units, such as clauses and sentences. As limited vocabulary resources may lead to difficulties in expressing ideas, limited knowledge of, and access to, grammatical structures in long-term memory may also lower the quality of written text (Cumming, 2001; Grabe & Kaplan, 1996). The notion that grammatical knowledge is an important element in successful L2 writing is also supported by empirical L2 writing research (Y. Lu, 2010; T. McNamara, 1996; Schoonen et al., 2003, 2011). For example, for Dutch (L1) adolescents learning English (L2), Schoonen et al. (2003) found that writing skills and grammatical knowledge were strongly correlated (r = .84), and grammatical knowledge test scores significantly predicted L2 writing scores. Similarly, in the context of an occupational English test, grammar scores were found to explain around 60% of the variance in writing scores (McNamara, 1996). On the other hand, in a study for Chinese (L1) university students learning English (L2), Lu (2010) found that while L2 grammatical knowledge was moderately correlated with L2 writing scores (related to both content and

language), it yielded a negligible effect in predicting writing scores when vocabulary knowledge was considered together.

2.3.3 Comparing vocabulary knowledge and grammar knowledge

Both vocabulary and grammar knowledge have been widely researched as important components in L2 research under the notion that vocabulary and grammar are the basic building blocks of second language acquisition (SLA; Ortega 2009). In comparing vocabulary and grammar knowledge, it is generally argued that vocabulary learning precedes grammar learning. In SLA research, Pienemann's (1998) processability model assumes that grammar acquisition is driven by vocabulary acquisition. Similarly, Lexical Learning Hypothesis (N. Ellis, 1997) presumes that vocabulary knowledge is necessary to grammar learning. In addition, Wilkins (1972) states that "without grammar very little can be conveyed, without vocabulary nothing can be conveyed" (p. 111), emphasizing the greater importance of vocabulary knowledge in conveying meanings. Empirical research has also compared the relative contributions of vocabulary and grammar knowledge to L2 performances. Results have reported mixed findings. For example, in predicting L2 reading comprehension, Shiotsu and Weir (2007) found the relatively greater contribution of syntactic knowledge over vocabulary knowledge, while Zhang (2012) reported the superiority of vocabulary knowledge over syntactic knowledge.

However, this comparison of the relative importance of vocabulary or grammar in L2 performance treats the two constructs as independent, whereas an increasing body of the literature in L2 research argues against the clear-cut boundaries between vocabulary and grammar (e.g., Alderson & Kremmel, 2013; Hulstijn, 2015; Römer, 2009). For example, from a usage-based perspective, language patterns can be defined at varying levels of abstraction and complexity rather than as a dichotomy between grammar and vocabulary that are inseparable

from each other. That is, grammar and lexis are conceptualized along a continuum (Goldberg, 1995), such that "language consists of grammaticalised lexis" (Lewis, 1993, p. vi). From this perspective, vocabulary is mostly acquired through exposure to utterances or sentences that provide grammatical context for word use, rather than words in isolation. Thus, the learning of vocabulary and grammatical knowledge likely goes hand in hand (Hulstijn, 2015).

Empirical research has also supported the close relationship between vocabulary and grammar components, reporting moderate-to-strong correlations between the two (Shiotsu & Weir, 2007; Yamashita & Shiotsu, 2017; Zhang, 2012). Furthermore, because vocabulary and grammar knowledge significantly correlate with each other and thus show multicollinearity, when predicting language performances, suppression effects (i.e., when two predictors are strongly correlated, a unique contribution of a less strong predictor may disappear in the presence of a stronger predictor; Tabachnick, & Fidell, 2012, p. 155) may occur. For instance, in an L2 writing study of Schoonen et al. (2011), the effect of L2 vocabulary knowledge on L2 writing disappeared in regression analysis due to the presence of other variables that were more strongly correlated with L2 writing, such as L2 grammar knowledge. Thus, it appears that an attempt to answer the question of whether vocabulary knowledge is more important than grammar knowledge in predicting L2 performance may not be fruitful (Alderson & Kremmel, 2013; Hulstijn, 2015).

Following the ideas that the relationship between vocabulary and grammar is a continuum and that it is difficult to identify the relative contribution of grammar or vocabulary knowledge in L2 performance, this dissertation uses vocabulary knowledge as a proxy measure of language knowledge stored in long-term memory for three main reasons. First, from an SLA theoretical perspective, vocabulary is considered key to learning grammar, rather the other way

around (N. Ellis, 1997; Milton, 2009; Pienemann, 1998). Second, in writing, vocabulary might be more important in generating ideas because it is more likely that writers first search for lexical items that match their intended ideas and then cast these lexicalized ideas into grammatical structures, or simultaneously look for lexical items and grammatical structures, rather than think about grammatical structures first. Lastly, for practical reasons, measuring vocabulary knowledge is easily standardized and convenient, while measuring grammatical knowledge often is not (Milton, 2009).

2.4 Roles of Cognitive Skills on L2 Writing

Beyond language demands, cognitive knowledge and skills (e.g., attention and working memory) are also important components in cognitive models of writing (Chenoweth & Hayes, 2001; Hayes, 1980; Hayes & Berninger, 2014; Hayes & Flower, 1980). Three important types of cognitive skills and knowledge in writing processes are attention, working memory, and general knowledge stored in long-term memory. Below, the importance of each of attention and working memory in L2 research in general and L2 writing in particular is discussed. The importance of general knowledge in L2 writing is then discussed.

2.4.1 Attention and L2 writing

Attention is considered as an important cognitive element in L2 research (e.g., Gass & Lee, 2011; Kormos, 1999; Robinson, 2003; Schmidt, 1995). From a broader information-processing perspective, attention is described in two main ways: attention as selection of information for further processing, and attention as effort to sustain performance on a task (Sanders, 1998). First, from a view of attention as selection of information, attention enables L2 learners to select L2 input, keep it active in working memory, and link it to long-term memory (Robinson, 2003). Specifically, attention is considered important as being selective of input, such

that linguistic information detected and selected from L2 input via (focal) attention can be "noticed", and then via this notice, input can become "intake" for learning and being stored in long-term memory (Schmidt, 1995). In addition, attention can also function as an inhibitory mechanism of not perceiving all of the much larger set of the detected information from incoming L2 input. Second, from a view of attention as effort, attention is described as energy or activity devoted to maintaining performance on a task. Failure to sustain attention to a task likely has a negatively influence on the task performance. For example, failure to maintain attention to L2 spoken communication may have a negative impact on self-repairing and monitoring of output (Kormos, 1999; Robinson, 2003).

Conceptualizing attention as both selection and effort is also relevant to L2 writing. First, attention as selection likely helps L2 writers selectively activate appropriate language among various competing L2 expressions for intended meanings to be conveyed. In addition, via focal attention, L2 writers may also help inhibit competing L1 expressions (when needed) during writing. Thus, this selective attention helps coordinate attentional resources to what needs to be activated and inhibits what needs to be suppressed. Second, attention as effort may help L2 writers sustain their energy to writing tasks that typically require high cognitive and language demands. Furthermore, when producing timed essays, attention likely helps writers direct more resources towards composing the text and stay focused on writing in order to express their ideas and messages as fluently as possible within the limited time. In this aspect, attention can be conceptualized "the ability to maintain focus on a task in the face of distraction" as described in the Hayes-Berninger model (2014, p. 4).

Attention is often measured using the Stroop task (Gass & Lee, 2011; MacLeod, 1992; Stroop, 1935). During the Stroop task, participants are asked to speak aloud the ink color of a

color word or that of a symbol. In congruent, or neutral, situations, participants are asked to name the ink color of a symbol (e.g., naming the color 'blue' of @ printed in blue ink). In incongruent situations, participants need to say the ink color of a color word printed in a different ink color (e.g., saying green for the word 'red' written in green ink, ignoring the word 'red'). Incongruent conditions are more difficult and result in longer response times than neutral conditions because participants need to suppress the distracting stimulus of the word name to correctly name the color of the word. This effect of response differences (i.e., taking longer times in incongruent conditions as compared to congruent conditions) is called the Stroop effect. It is generally interpreted that the smaller the difference is (i.e., smaller Stroop effects), the higher levels of attention an individual has (MacLeod, 1992).

In L2 research, studies have examined attention in relation to L2 proficiency (Chen & Ho, 1986; Gass, Behney, & Uzum, 2013; Gass & Lee, 2011; Sumiya & Healy, 2004). For example, Gass et al. (2013) reported that for English-speaking adult learners of Italian, learning gains from feedback during oral interactions were related to Stroop test scores, such that high gainers had better attentional control than low gainers. However, to my knowledge, little research has been conducted on the role of attention in L2 writing performances.

2.4.2 Working-memory and L2 writing

In addition to attention, another important cognitive construct in L2 research is working memory (Kormos, 2012; Kormos & Sáfár, 2008; Robinson, 2003; Service & Kohonen, 1995; Williams, 2011). Working memory is generally defined as a central executive processor that stores, retrieves, manages, and manipulates information in an active state within a limited capacity system (Engle, 2002). Baddeley's seminal model of working memory suggests two main systems: a central executive system that manages information between short-term memory

stores and long-term memory resources, and storage-based, slave systems that consist of *phonological loop* and *visuospatial sketchpad* (Baddeley, 1986). The phonological loop controls verbal information, while the visuospatial sketchpad manages imagery and spatial domains. An additional system, *episodic buffer*, which temporarily stores information and communicates with long-term memory, has been added in a more recent model (Baddeley, 2000).

Working memory capacity has been extensively researched in L2 research (e.g., Linck, Osthus, Koeth, & Bunting, 2014; Robinson, 2003; Williams, 2011). Individuals with greater working memory resources tend to have larger L2 vocabulary size (Service & Kohonen, 1995), perform better on learning foreign vocabulary in laboratory settings (Williams & Lovatt, 2003), and have better L2 reading comprehension abilities (Jeon & Yamashita, 2014). In addition, in a meta-analysis of working memory studies using data from 79 samples with 3,707 participants, Linck et al. (2011) support that working memory is positively related to L2 processing and proficiency variables.

In measuring working memory capacity, two task types are generally used: simple span tasks (i.e., measuring an ability to store information only) and complex span tasks (i.e., measuring an ability to store information and handle additional processing tasks; Linck et al., 2011). In addition, based on the content domain of stimuli, working memory tasks can be either verbal (i.e., processing language information, such as words and letters) or nonverbal (i.e., processing non-language information, such as numeric digits and math equations). For example, letter span tasks in which participants are asked to recall a series of letters are simple and verbal, while digit span tasks are simple and nonverbal. In addition, reading span tasks in which participants are asked to recall a series of items after reading written sentences are complex and verbal, while operation span tasks in which participants are asked to recall a series of items after

solving math equations are complex and nonverbal. In general, while complex span tasks are better predictors of L2 performance (Linck et al., 2013) and L1 performance (Daneman & Merikle, 1996) than are simple span tasks, both complex and simple tasks are found to relate to L2 performance.

In cognitive models of writing, working memory is an important element of writing (e.g., Hayes & Berninger, 2014). Working memory is related to processing the current contents available to the writer. Also, as working memory is responsible for allocating attentional resources and utilizing long-term memory, it has influences on writing across various processes, including generating ideas, translating ideas into written text, and evaluating writing processes. In addition, considering the limited-capacity working memory, automatic language processing during producing the text likely lessens language demands, enabling higher-level cognitive processing, such as using general knowledge stored in long-term memory. Thus, L2 writing may require more cognitive and language demands than L1 writing due to less automatized L2 knowledge (DeKeyser, 2007).

Although working memory is an important component of cognitive models of writing, there has been scarce research on the role of working memory in L2 writing performance. In addition, among the few L2 writing studies that have examined working memory, findings reported mixed results partly due to various operationalizations and measures of working memory. For instance, Kormos and Sáfár (2008) found that when working memory was measured by a phonological non-word span task (in which participants listen to non-words and asked to recall them), it was moderately correlated with L2 writing scores. However, when measured by a backward digit span task, working memory was not correlated with L2 writing scores. Similarly, in an L2 writing study for Chinese (L1) university students, Lu (2010) reported

that English (L2) writing scores were not correlated with working memory capacity as measured by a verbal, operational span task.

2.4.3 General knowledge and L2 writing

The use of general knowledge (i.e., retrieving relevant information from long-term memory) is also considered crucial in producing L2 texts (Berninger et al., 2002; Hayes & Berninger, 2014; Hayes & Flower, 1980). Higher-levels of general knowledge likely facilitate flexible access to context-relevant concepts and ideas, which may in turn enhance writing processes, particularly during planning stages when writers generate ideas and develop texts. This retrieval of knowledge from long-term memory during writing is likely to be automatic, such that information related to the topic might be automatically probed, which would be followed by the exploration of additional information for elaboration from working memory.

In L1 writing contexts, writers with richer general knowledge on the writing topic in long-term memory tend to produce higher-rated essays and with less effort (Dansac & Alamargot, 1999; Hayes & Berninger, 2014). In L2 writing contexts, topical knowledge (i.e., prior knowledge about the topic a writer is writing about) has been considered important in independent writing in which writers need to write about a topic based on their prior knowledge and experience (Bachman & Palmer, 1996; He & Shi, 2012; Lee & Anderson, 2007). For instance, He and Shi (2012) found that ESL college learners from various proficiency levels produced essays of higher quality on a general topic (i.e., university students) than they did on a specific topic (i.e., federal politics).

Taken together, while research has examined the roles of working memory capacity and the effects of topics (i.e., specific vs. general) on L2 writing performance, relatively little

research has been conducted on the effects of writers' general knowledge and other cognitive skills (e.g., attention) on multilingual writing performances.

2.5 Language Features and L2 Writing Quality

The L2 writing research mentioned above measured language knowledge using various types of tests (e.g., receptive vocabulary tests; Schoonen et al., 2011) in relation to L2 writing performance. Beyond language knowledge, a substantial body of L2 writing literature has also examined linguistic features as found in student writing, such as fluency and linguistic complexity (i.e., lexical and syntactic complexity), in relation to writing scores (e.g., Crossley & McNamara, 2012; Leki, Cumming, & Silva, 2008). Generally, L2 writers' linguistic features can impact judgments of writing quality such that higher rated L2 essays tend to contain more sophisticated vocabulary and more complex structure (Kyle & Crossley, 2016; Ortega, 2015). L2 writers' more fluent writing (i.e., greater number of words) can also result in higher writing quality (van Waes & Leijten, 2015). Below, two important linguistic features, linguistic complexity and fluency, in relation to L2 writing quality are discussed.

2.5.1 Linguistic complexity

Complexity has been traditionally defined as "[t]he extent to which the language produced in performing a task is elaborate and varied" (Ellis, 2003, p. 340). Researchers generally agree that measures of linguistic complexity include both lexical and syntactic domains (Bulté & Housen, 2012). With respect to lexical complexity, two lexical measures that have been widely used in L2 writing research include lexical diversity and lexical sophistication (Read, 2000). Lexical sophistication is related to the use of sophisticated and advanced words while lexical diversity is associated with the use of unique words in a text. Lexical diversity has been conceptualized as a multidimensional construct that consists of various subcomponents such as

size (number of tokens), richness (number of types), and effective number of types (Jarvis, 2013). Likewise, lexical sophistication has also been considered as composed of various lexical features such as frequency, concreteness, familiarity, and hypernymy (Kim, Crossley, & Kyle, 2018; Kyle & Crossley, 2015). Research has found that more proficient L2 writers tend to produce texts with greater lexical diversity (Crossley & McNamara, 2012; Jarvis, 2002), and use lower-frequency words, less imageable words, less familiar words, and more specific words (Crossley et al., 2014; Guo, Crossley, & McNamara, 2013; Kyle & Crossley, 2016; Laufer & Nation, 1995).

With respect to syntactic complexity, a comprehensive view has been proposed by Norris and Ortega (2009), such that syntactic complexity is conceptualized as composed of four sub-constructs: overall complexity (i.e., length-based measures, such as mean length of sentence), subordination complexity (e.g., dependent clauses per clause), sub-clausal complexity via phrasal elaboration (e.g., mean length of clauses), and coordination. They further suggest that global complexity likely captures syntactic complexity across different levels of L2 proficiency. Research has found that more proficient L2 writers tend to use longer clauses and sentences with phrasal elaboration (Biber, Gray, & Poonpon, 2011; Crossley & McNamara, 2012; X. Lu, 2010; Yang, Lu, & Weigle, 2015).

2.5.2 Fluency

In the field of SLA, fluency has been traditionally defined as "the extent to which the language produced in performing a task manifests pausing, hesitation, or reformulation" (Ellis, 2003, p. 342). While complexity is related to learners' representations of L2 knowledge (e.g., vocabulary and grammar knowledge), fluency is related to learners' access to, and control over, their L2 knowledge in language use (Housen, Kuiken, & Vedder, 2012). In written production,

fluency is often measured as length of text produced in timed writing, defined as "a measure of the sheer number of words or structural units a writer is able to include in their writing within a particular period of time" (Wolfe-Quintero, Inagaki, & Kim, 1998, p. 14). Generally, longer essays tend to receive higher scores (e.g., Grant & Ginther, 2000; Kobrin, Deng, & Shaw, 2007). However, it should be noted that fluency in written production itself does not entail lexical complexity or accuracy, and thus longer texts do not necessarily entail better writing (Hoswell, 2000). That is, more fluent writing may be either good or bad.

In cognitive models of writing, fluency has been conceptualized in terms of writing processes, such that fluent writing processes are featured with short pause time and higher production rate (MacArthur, Graham, & Fitzgerald, 2008; van Waes & Leijten, 2015). Furthermore, a recent investigation with keystroking data for L1 and L2 writing (van Waes & Leijten, 2015) suggested a multi-dimensional nature of writing fluency as composed of four main subcomponents: production (i.e., mean number of characters), process variance (i.e., standard deviation of characters), revision (i.e., mean number of characters), and pausing behavior. Importantly, writing fluency is related to burst length, such that longer bursts can lead to greater fluency (Chenoweth & Hayes, 2001; Hayes & Chenoweth, 2007; van Waes & Leijten, 2015). Burst length further reflects "the capacity of the translator to handle complex language structures" (Chenoweth & Hayes, 2001, p. 94). Thus, an increase in burst length (i.e., producing larger chunk of language between pauses) may indicate not only an increase in writing fluency, but also an improvement in the ability of the translator.

2.6 Current Study

Prior research has examined the crucial role that language and cognitive skills play in predicting L2 writing skills, and the importance of linguistic features in predicting L2 writing

scores. However, there remains a paucity of research in examining the longitudinal development of English writing skills in relation to language and cognitive skills for multilingual students in higher education. Thus, the purpose of this dissertation was to investigate the longitudinal development of English writing skills for multilingual undergraduate students in US higher education. Grounded on the Hayes and Berninger (2014) writing model (see Figure 2.2), the dissertation addressed three main research questions. First, it examined the relationships of English writing scores and writing score changes with general cognitive/language resources (i.e., attention, long-term memory, working memory, and reading, and writing success). Also, an additional variable of years of English immersion instruction was included because years of English immersion instruction indicate different degrees of exposure to English through formal schooling. The second research question investigated the longitudinal relationship between roles of the translator (i.e., turning ideas into language strings) and L2 writing scores over time. Lastly, the dissertation examined the longitudinal relationship among English writing, reading, and vocabulary. The first two questions tested theoretical hypotheses based on Hayes and Berninger (2014) in multilingual contexts and thus were confirmatory in nature, while the last research question did not hold an a-prior hypothesis and was thus exploratory in nature. The research questions that guided this dissertation are elaborated below.

2.6.1 Research Question 1: Relationship between general resources, years of English immersion instruction, and English writing scores

To answer the first research question (i.e., the relationships of English writing scores and writing score changes with general cognitive/language resources along with years of English immersion instruction), the dissertation was guided by the following research question:

1. How do initial levels of general cognitive/language resources and years of English immersion instruction predict the initial level of English writing scores and changes in English writing scores in multilingual undergraduate students?

English writing scores were based on holistic scores of writing quality using prompts from the Scholastic Aptitude Test (SAT). Higher English writing scores represented higher English writing ability. English writing ability was measured on two occasions: Time 1, and Time 2 which occurred in at least five months (approximately equivalent to one academic semester) after Time 1. Five cognitive/language resources based on the Hayes-Berninger model (2014) were included as predictors of both English writing scores and writing score changes: attention, working memory, long-term memory related to general knowledge, long-term memory related to language (i.e., vocabulary knowledge), and reading skills. Each of the general cognitive resources was measured at Time 1 only.[2] As suggested by Hayes-Berninger (2014), attention was measured by a Stroop task (Stroop, 1935). Working memory capacity was measured using a verbal, simple running span task (Kim, Payant, & Pearson, 2015). Long-term memory related to general knowledge was measured by a general knowledge test (Roscoe, Crossley, Snow, Varner, & McNamara, 2014). Long-term memory related to language was measured by a standardized receptive English vocabulary knowledge test (Gates-MacGinitie Reading Test [GMRT]; MacGinitie, MacGinitie, Maria, Dreyer, & Hughes, 2000a). Reading skills were measured using a standardized English reading comprehension test (GMRT; MacGinitie et al., 2000a). Years in English immersion instruction were collected using a background survey item.

---

[2] While English vocabulary knowledge and English reading skills likely change over time in English-speaking environments and academic contexts in higher education, the first question focused on initial levels of general cognitive/language resources only. The longitudinal relationship among vocabulary, reading, and writing in the English language was addressed in the third research question.

2.6.2 Research question 2: Relationship between the translator and English writing scores

The second research question addressed the longitudinal relationships between roles of the translator and English writing scores. Roles of the translator were measured in terms of both processes features and product features. A writing processes features was measured as the mean length of language bursts (i.e., the mean number of characters produced between pauses longer than two seconds) using key-stroke data. A product feature was measured using linguistic features (i.e., lexical sophistication) as found in the essays. Lexical sophistication was measured by the use of academic words. The dissertation was guided by the following research question:

> 2. What are the longitudinal relationships among English writing score, burst length, and academic word use in multilingual undergraduate students?

2.6.3 Research question 3: Longitudinal relationship among English writing, reading, and vocabulary

The third research question addressed the longitudinal relationship among English writing ability, reading ability, and vocabulary knowledge, each of which was measured on two occasions. The dissertation addressed two sub-research questions, 3a and 3b. The research question 3a examined the longitudinal relationships among English writing, reading, and vocabulary. Considering that English writing, reading, and vocabulary pertain to English literacy skills, the research question 3b tested whether a common latent variable of English literacy that was informed by English writing, reading, and vocabulary could be constructed across the two times of measurement. The dissertation was guided by the two following exploratory research questions:

> 3a. What is the longitudinal relationship among English writing, reading, and vocabulary in multilingual undergraduate students?

3b. Can a common latent variable of English literacy that is informed by English writing, reading, and vocabulary in multilingual undergraduate students be constructed across two occasions?

3 METHOD

3.1 Participants

A total of 101 undergraduate students from a research-oriented university located in the southern United States were recruited. In recruiting participants, three criteria were used: undergraduate students who (a) were enrolled in the university where the research was conducted; (b) began to learn English other than a mother tongue; and (c) regularly speak/spoke a language other than English at home.[3] As such, both international and Generation 1.5 students (i.e., the children of first-generation immigrants) were eligible to participate. Participants were from various countries around the world, such as Colombia, Ethiopia, India, Iran, Mexico, Pakistan, Venezuela, Viet Nam, and Zimbabwe. In the second session, 79 students returned (a 21.78% attrition rate). Among them, one student's performances in the second session were not recorded due to technical errors and excluded from analyses. In addition, another student was an international student from Bermuda, but based on the background survey data, it was evident that the student was a monolingual English speaker. Thus, this student was also excluded from analyses. Therefore, in this dissertation, data from 77 students were analyzed. Table 3.1 shows demographic characteristics of the participants. Fifty students were female, and 27 were male. On average, students were 20.53 years old (SD = 2.82). Forty-six students were international students, while 31 students were non-international students (i.e., citizens and residents).

[3] At the very beginning of the dissertation data collection, the target participants were international students. However, due to difficulties in recruiting international students, the recruitment criteria were expanded such that diverse multilingual students were eligible to participate.

Table 3.1 Demographic Characteristics of the Participants (N = 77)

Characteristics	
Gender	Female (50); Male (27)
Academic year	Freshman (27); Sophomore (22); Junior (18); Senior (10)
Colleges/majors	Arts and Humanities (4); Business (20); Health and Medicine (8); Science, Math and Technology (31); Social Science (14)
Age	18–19 (33); 20–21 (28); 22+ (16)
International (visa) student	Yes (46); No (31)
L1	Amharic (1); Arabic (2); Chinese (12); Creole (1); Croatian (2); Farsi (1); French (10); German (1); Gujarati (1); Hindi (5); Korean (7); Malayalam (1); Marathi (1); Oriya (1); Portuguese (1); Pulaar (1); Russian (3); Shona (2); Spanish (11); Tigrinya (1); Urdu (4); Vietnamese (8)
Citizenship	Brazil (1); Burkina Faso (1); Cameroon (1); China (10); Colombia (1); Croatia (2); Democratic Republic of Congo (4); El Salvador (2); Ethiopia (2); France (3); Germany (1); India (7); Iran (1); Mexico (1); Pakistan (1); Peru (1); Russia (3); South Korea (3); Syria (1); U.S.A. (21); Panama (1); Venezuela (2); Vietnam (5); Zimbabwe (2)
Age of initial English learning	0–5 (32); 6–10 (29); 11–15 (12); 16+ (4)
Years of learning English (both immersion and classroom settings)	0–5 (4); 6–10 (13); 11–15 (39); 16+ (21)
Years in English immersion instruction settings	0–5 (36); 6–10 (10); 11–15 (15); 16+ (16)

3.2 Measures

3.2.1 Background survey

The background survey was administrated through Qualtrics Research Suite software (Qualtrics, Provo, UT). It asked questions about demographic information (i.e., L1s, age, major, academic year, gender, and citizenship). It also included questions about English learning backgrounds: the age of initial English learning, years of learning English, and years living in an English-speaking country. The background survey is provided in Appendix A.

3.2.2 Attention

Attention is important in maintaining focus and avoiding distraction during writing (Hayes & Berninger, 2014). Attention was measured using a Stroop test. The Stroop test was followed the experimental design of Stroop (1935) but was administrated using E-Prime 2.0 software (Schneider, Eschman, & Zuccolotto, 2012). Participants completed the test by naming font colors as quickly as possible within a given time limit in two conditions. Each condition had 20 items. Before real trials, an instruction followed by four practice items was provided. Among the two conditions, one was a congruent one, and the other was an incongruent one. In the congruent condition, participants named font colors that appeared with the symbols "@@@@" repeatedly. The symbol strings were displayed in red, green, brown, blue or purple ink. In the incongruent condition (i.e., interference condition), participants were asked to name the font colors of words that did not match the word meaning. These words were shown repeatedly in red, green, brown, blue or purple ink. For example, participants saw the word 'blue' in a green font color and were instructed to say 'green' (color) instead of 'blue' (word). The order of presenting the conditions was random. Within each condition, items were randomly presented. In the item sequence, each item was preceded by a 1000 ms fixation point (+). Each item lasted two seconds. This test took approximately 5 minutes to complete.[4]

During testing, response latencies (i.e., the time between the stimulus onset and the participant's response onset) were recorded by E-Prime. For each condition, each participant's

---

[4] Following the original Stroop's work (1935), there were two additional conditions. One was a congruent condition in which participants were asked to read color words (e.g., 'red') printed in a black font color, while the other was an incongruent condition in which participants were asked to read color words printed in a different font color from the word, ignoring the color (e.g., read the word 'red' printed in a green font color, ignoring the green color). However, these two conditions were used for filler items only because Stroop (1935) found that the interference of conflicting color stimuli was not reliable. In addition, the Stroop effect is often measured by differences between response times to naming font colors different from the color words and those to naming colors printed in neutral symbols (Ludwig, Borella, Tettamanti, & de Ribaupierre, 2010; Roy et al., 2016).

average response times were calculated in milliseconds. When errors occurred or no response was recorded, their response times were not calculated toward the average response times. Following Ludwig et al. (2010), a normed naming interference index was calculated and used as follows: (response times in the incongruence condition – response times in the congruence condition)/response times in the congruence condition. Lower scores on the normed naming interference index means better attentional capacity.

3.2.3 Working memory capacity

Working memory capacity is an important component of writing (Hayes & Berninger, 2014) in order to retrieve relevant knowledge from long-term memory and process cognitive and language demands. To measure working memory capacity, a verbal, simple running memory span task (Kim et al., 2015) was used. The running span task was administered on a computer, with automated, built-in instructions and practice sessions, using E-Prime 2.0 software (Schneider et al., 2012). During the task, participants were presented with a random series of letters (e.g., FGHJKQ), and then asked to report the last *n* letters in the same order as presented (target length = 3, 4, 5, 6, or 7 letters). For example, participants were presented with five letters (e.g., HJQRP), and asked to recall the final three letters in correct order (i.e., QRP). When participants forgot one or more letters, they were asked to click 'blank' to leave a spot for each missing letter. After each trial, feedback that informed the number of correctly remembered letters in the presented order was provided. This test took approximately 10 minutes to complete. The total number of correctly remembered letters regardless of whether the letters were remembered in the presented order was scored. Although complex span tasks, such as an operation span task, are generally better predictors of language performance (e.g., writing) than

are simple running span tasks, simple tasks, such as a running span task used in this dissertation, are also found to closely relate to language performance (Linck et al., 2013).

3.2.4 General knowledge

Having strong general knowledge and the ability to retrieve it from long-term memory during writing likely helps generate ideas (Berninger et al., 2002; Hayes & Berninger, 2014; Hayes & Flower, 1980), which are important elements of writing quality. Students' general knowledge was assessed using a 30-item test that asked about students' knowledge about science, literature, and history. Each item had four choices with one correct answer. The use of this test has been validated in previous research related to L1 reading comprehension (Roscoe, Crossley, Snow, Varner, & McNamara, 2014) and L1 writing (Allen, Snow, & McNamara, 2016) with high reliability ranging from .72 to .81. An example item is "*Who is the author of the mystery fiction Sherlock Holmes?*" along with one correct answer (i.e., *Arthur Conan Doyle*) and three distractors (i.e., *Agatha Christie, Edgar Allan Poe*, and *James Joyce*). Questions were presented in a random order on a computer screen, using Qualtrics Research Suite software (Qualtrics, Provo, UT). One caveat for using this general knowledge test for multilingual students is that it might be biased in favor of native speakers of English because some of the question items are specifically related to English and American literature and history. Cronbach's alpha for the general knowledge test (k = 30) was .667.

3.2.5 English vocabulary knowledge

Vocabulary knowledge is vital in creating meanings during writing (Berninger, 2000; Hayes, 1980; Hayes & Berninger, 2014). Vocabulary knowledge was assessed on two occasions using two different vocabulary sections of the Gates-MacGinitie Reading Skill tests (Fourth Edition; MacGinitie et al., 2000a): Level 10/12, Forms S and T. The Gates-MacGinitie Reading

Skill tests were chosen because they are standardized tests that have two different, comparable forms (which were needed for the purpose of the current study) and have been widely used in both L1 and L2 contexts (e.g., Crossley, Yang, & McNamara, 2014; Ozuru, Rowe, O'Reilly, & McNamara, 2008; Uccelli, Galloway, Barr, Meneses, & Dobbs, 2015). Each of the two Gates-MacGinitie vocabulary tests had 45 multiple-choice questions. An item was presented in a short phrase with the target word underlined (e.g., a big garage). Participants were asked to select the most closely related word or phrase from a list of five options. To compare scores from the two different test forms, a normed scale for extended scale scores (ESSs) developed based on all test items from Level 3 through 10/12 and Adult Reading (MacGinitie, MacGinitie, Maria, Dreyer, & Hughes, 2000b) was used. The tests were administrated using Qualtrics Research Suite software (Qualtrics, Provo, UT). Participants was given 20 minutes to complete each vocabulary test. The tests were counterbalanced across participants, such that 44 participants took the Form S at Time 1 and the Form T at Time 2, while 33 did the opposite. The Cronbach's alpha values for the Form S and the Form T were .870 and .859, respectively.

3.2.6 English reading comprehension skills

English reading comprehension were measured on two occasions using Qualtrics Research Suite software (Qualtrics, Provo, UT). For assessing English reading comprehension skills, two forms of the Gates-MacGinitie Reading Skill tests (Fourth Edition; MacGinitie et al., 2002a) were used: Level 10/12, Forms T and S. Each of the two Gates-MacGinitie reading tests comprised 48 multiple-choice questions with passages from various domains, including narratives, autobiographies, and academic texts. Participants were given 40 minutes to complete each reading comprehension test. The comprehension questions included literal and inferential questions. To compare scores from the two different test forms, a normed scale for ESSs

(MacGinitie et al., 2002b) was used. The tests were counterbalanced across participants, such that 44 participants took the Form S at Time 1 and the Form T at Time 2, while 33 did the opposite. The Cronbach's alpha values for the Form S and the Form T were .891 and .888, respectively.

3.2.7 English writing ability

English writing abilities were also measured on two occasions. Two essays were written in response to SAT-based prompts. SAT prompts were chosen because although SAT writing is a type of pseudo-academic five-paragraph essays (MacDonald, 1994), it is a type of persuasive "essays" (rather than critiques or narrative recounts), whose purpose are to demonstrate the ability to make a coherent argument to persuade the reader by typically accompanying an introduction, series of arguments, a conclusion (Gardner & Nesi, 2013). Indeed, these types of essays are the most frequent occurring genres in a corpus of texts produced by undergraduate and postgraduate students in an English-speaking country (i.e., England) for assessment purposes (Gardner & Nesi, 2013), accounting for around 40% of the assignments collected. Thus, the ability to produce a well argument essay is crucial in academic contexts in higher education. In addition, producing an essay is important as it forms a basis for more advanced academic writing involving discipline-specific manner of thinking and arguing (MacDonald, 1994). SAT-based prompts were chosen over writing prompts that are designed for ESL and English-as-a-foreign-language (EFL) students (e.g., TOEFL prompts) because some participants (e.g., those from India and those who came to the U.S.A at their early age) were very proficient English speakers, and thus ESL- or EFL-based prompts would not be relevant to those students.

Two SAT-based prompts were used to control for the prompt effects. One was about competition and the other was about appearance. Each prompt is presented in Table 3.2.

Instructions were “Think carefully about the issue presented in the following excerpt and the assignment below.” Students were given 25 minutes to complete each writing task. The writing test forms were counter-balanced, such that 44 participants wrote essays about competition at Time 1 and about appearance at Time 2, while 33 did the opposite. Essays were collected using Qualtrics (Qualtrics, Provo, UT). During writing on Qualtrics, spelling and grammar checks were not included. Participants were not allowed to use any reference materials. The participants saw their written texts during their entire writing processes. During each writing test, participants’ keyboard strokes were also recorded using the keystroke logging program Inputlog (Leijten & Van Waes, 2013).

Table 3.2 Two SAT-based Prompts

Topic	Prompt
Competition	While some people promote competition as the only way to achieve success, others emphasize the power of cooperation. Intense rivalry at work or play or engaging in competition involving ideas or skills may indeed drive people either to avoid failure or to achieve important victories. In a complex world, however, cooperation is much more likely to produce significant, lasting accomplishments. Do people achieve more success by cooperation or by competition?
Appearance	All around us appearances are mistaken for reality. Clever advertisements create favorable impressions but say little or nothing about the products they promote. In stores, colorful packages are often better than their contents. In the media, how certain entertainers, politicians, and other public figures appear is more important than their abilities. All too often, what we think we see becomes far more important than what really is. Do images and impressions have too much of an effect on people?

Student writings were assessed by two trained raters using a six-point holistic rating scale developed for the SAT (see Appendix B). The rating scale holistically assessed essay quality with a minimum score of one and a maximum score of six. The scale focuses on development of a point of view on the topic, critical thinking, use of appropriate examples, accurate and adapt use of language, use of variety of sentence structure, and errors in grammar and mechanics as well as text organization and coherence. The two raters were PhD students in applied linguistics

who were familiar with university writing. Raters were first trained to use the rubric with 20 essays previously written in a similar condition. To measure inter-rater reliability, square-weighted Cohen's Kappa coefficients (i.e., disagreements were weighted based on their squared distance from exact agreement) were calculated. After an interrater reliability of at least Cohen's Kappa = .70 was reached in the training set, the raters scored the essays collected for this dissertation independently. During the rating process, the raters used randomly ordered essays without information about test takers. If two ratings differed by more than one point, the raters adjudicated the ratings so that the disagreement between the raters was one point or less. Inter-rater reliability was acceptable with square-weighted kappa values of .774 (for scores at Time 1) and .726 (for scores at Time 2). Average scores between the raters were calculated for each essay.

3.2.8 Language features in student writing

In relation to the translator, language features in student writing were measured in terms of both processes and products. A process feature of the translator was measured by the mean length of language bursts (i.e., the mean number of characters produced between pauses longer than two seconds; Limpo & Alves, 2017) because burst lengths are considered to be associated with the capacity of the translator that searches for appropriate language forms to encode ideas (Hayes, 2009). That is, longer burst lengths may indicate higher capacity of the translator in terms of its processes. Pauses longer than two seconds were chosen to calculate mean burst lengths because two seconds can represent a period when at least one word is formulated (Alves & Limpo, 2015; Chenoweth & Hayes, 2001; Connelly, Dockrell, Walter, & Critten, 2012; Kaufer et al., 1986; Limpo & Alves, 2017). This means that the two seconds can reflect a minimal period that represents the process of transforming ideas into language forms. Less than

two seconds is not likely to sufficiently reflect the translating process of turning ideas into meaningful language strings. In addition, burst lengths included all of the characters produced between two seconds. Thus, when writers made edits, the number of characters prior to editing were also included in burst lengths. The mean length of bursts for each participant was recorded by using Inputlog (Leijten & Van Waes, 2013).

A product feature of the translator was measured using the lexical features found in student essays. Specifically, lexical features were measured by the production of academic words (e.g., *eliminate*, *regulate*) that are frequently found in an academic corpus as compared to other corpora (e.g., a spoken corpus).[5] This was because producing persuasive essays is a type of academic writing (Gardner & Nesi, 2013), and to produce academic prose, the role of the translator may involve retrieving academic words that suit the writer's persuasive purposes in academic contexts. Thus, the greater use of academic words may indicate higher capacity of the translator in terms of its production in academic texts. In addition, the use of academic words is considered important because the greater use of academic words indicates higher quality academic writing (Douglas, 2013). To measure the use of academic words, the academic word list (AWL; Coxhead, 2000) was used, which includes 570-word families which are widely used in many academic disciplines. Generally, 10 words in every 100 in academic texts (10%) are

[5] For the statistical analysis, I needed to select one lexical sophistication index because examining all possible lexical features would be not possible. To select a lexical index, word frequency indices were first considered to indicate sophisticated words because less frequently used words have been generally considered as sophisticated or advanced words for more than two decades (Laufer & Nation, 1995). Different word frequency scores were calculated using various corpora, such as Corpus of Contemporary American English (COCA) corpora, and British National Corpus (BNC). However, the relationships between word frequency scores and writing scores were different depending on reference corpora used for calculating frequency, which made it difficult to choose a single index that can best represent frequency scores. Instead of using frequency indices, different indices were examined including academic words, age-of-acquisition, word concreteness, word familiarity, contextual diversity. Academic word counts were chosen for the analysis not only because academic words are considered important in academic writing but also because the academic word index showed the stronger correlation with writing scores at Time 1 than age-of-acquisition, word concreteness, word familiarity, contextual diversity indices.

likely to be found in the AWL (Coxhead, 2000). The number of academic words included in the AWL which were also found in the student's essay was calculated, normed by text length, and multiplied by 100, so that the number indicates the percentage of academic words in the text.

3.3 Data Collection Procedure

Participants were tested longitudinally on two occasions over at least a five-month interval. There were two recruiting periods. One group participated in their first session in September or October, 2017 and their second session in April or May, 2018. The second group participated in their first session in March, 2018 and their second session in September, 2018. Table 3.3 shows an overview of the data collection procedure. Each participant was seated in a quiet, small laboratory room which was equipped with one desktop computer. During the experiment, unless the participant called the researcher for help, no one was allowed to enter the room where the participant was seated. In Time 1, participants attended one session that lasted around two hours. In this session, participants first signed an informed consent form, and then provided demographic information and English learning backgrounds. Participants then completed a set of six test batteries: an English reading comprehension test, an English writing test, an English vocabulary test, an attention test, a working memory capacity test, and a general knowledge test. The order of these six test batteries was counter-balanced across participants. In Time 2, participants attended one session that lasted around 1.5 hours. In this session, they took a set of three batteries (i.e., English vocabulary, reading, and writing tests) in a counterbalanced order. While carrying out all of the tasks, participants were not allowed to use any reference tools, such as dictionaries. Participants received a $20 gift card for each session attended plus a $10 gift card after completion of the second session.

Table 3.3 Data Collection Procedure

Time 1 (2 hours)	Time 2 (1.5 hours)
a. English reading comprehension 1	a. English reading comprehension 2
b. English writing 1	b. English writing 2
c. English vocabulary 1	c. English vocabulary 2
d. Background survey	
e. General knowledge	
f. Working memory	
g. Attention	

3.4 Statistical Analysis

3.4.1 Basics of latent change score modeling

The primary statistical approach in this dissertation was latent change score modeling (Ghisletta & McArdle, 2012; McArdle, 2009). Latent change score modeling is a subtype of longitudinal structural equation modeling, which constructs change at the latent level. That is, as is true for structural equation models in which unobserved (latent) variables are created and tested by using observed (manifest) variables, latent change score models create and test latent change scores using observed scores repeatedly measured over time. Latent changes are defined as the score at a time point (t) that is not explained by the previous time point (t -1). Latent change score models are characterized by the "the direct inclusion of latent change scores to express specific developmental hypotheses about individuals and groups" (McArdle, 2009, p. 581). The change (development) period is captured by discrete time intervals with a minimum of two time points. In addition, latent change score models include explicit mean and covariance structures to compare means and covariances using longitudinal data. A detailed overview of latent change score modeling with technical descriptions can be found in Ghisletta and McArdle (2012), Grimm, McArdle, Zonderman, and Resnick (2012), and McArdle (2009).

There are at least four advantages of using latent change score models. First, the estimation of latent change scores allows for the test of not only intra-individual changes, but also variances in intra-individual changes that represent individual differences in changes, and a covariance between the score at the previous time and the score change (e.g., the relationship between the initial level and the change). Second, latent change score models allow researchers to explore the longitudinal relationships among different constructs as they develop over time. For example, it can be tested whether performance on one construct is related to subsequent change in performance on another construct. It can also be tested whether change in performance on one construct is associated with change in performance on another construct. Third, unlike repeated measures analysis of variance (ANOVA) which assumes compound symmetry (i.e., an assumption of an equal variance and an equal correlation over time), latent change score models do not hold such assumption. Lastly, missing data can be handled using full information maximum likelihood (FIML) estimation.

Latent change score models are drawn using representations of structural equation models. Based on McArdle (2009), Figure 3.1 represents a univariate latent change model using two time points. The two repeated scores are labeled as *Y*[1] and *Y*[2] with [1] indicating Time 1 and [2] indicating Time 2. The observed variables *Y*[1] and *Y*[2] are drawn as squares, while the latent change factor (Δ) is drawn as a circle. μ indicates a mean, α indicates a variance, and φ indicates a covariance. The implied constant of *1* is drawn in a triangle along with two one-headed arrows to represent two group effects, i.e., one that represents the mean of the initial scores (μ_1), and the other that represents the mean of the changes (μ_Δ). One-headed arrows represent directed relationships such as fixed effects and factor loadings, while two-headed arrows represent undirected relationships such as random effects and (co)variances. To calculate

Δ, two fixed values (= 1) are added on two one-headed arrows, i.e., one from *Y*[1] to *Y*[2] and the other from Δ to *Y*[2], so that an equation (*Y*[2] = 1***Y*[1] + 1***Δ) can be created. Thus, the change score (1***Δ = *Y*[2] – 1***Y*[1]) is defined as the score of *Y*[2] that is not explained by *Y*[1] (McArdle, 2009). In addition, the variance of the initial scores (σ_I^2), the variance of the changes (σ_Δ^2) and the covariance between the initial scores and the changes ($\varphi_{I\Delta}$) are estimated.

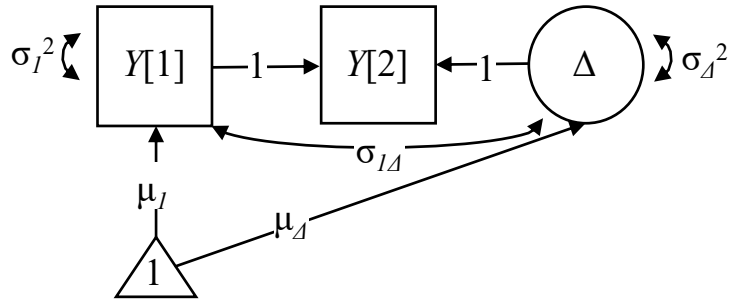


Figure 3.1 Univariate Latent Change Score Model for Two-Occasion Data
Note. [1] = Time 1, [2] = Time 2, Δ = "change in"

3.4.2 Statistical analysis for the research question 1

The first research question addressed the relationship of English writing scores and writing score changes with general cognitive/language resources and years of English immersion instruction. Prior to considering predictors of writing scores and score changes, a univariate, unconditional latent change score model without predictors was tested to examine change statistics for two-occasion L2 writing scores (see Figure 3.2). The two repeated scores of writing quality, labeled as *Writing*[1] (measured at Time 1) and *Writing*[2] (measured at Time 2), are drawn as observed variables in rectangles. A change score between the two writing scores, labeled as Δ*Writing*, is added as a latent variable drawn in a circle. To calculate Δ, fixed values (= 1) are added as two one-headed arrows, i.e., one from *Writing*[1] to *Writing*[2] and the other from Δ to *Writing*[2], so that an equation (*Writing*[2] = 1*** *Writing*[1] + 1***Δ) can be created. In this way, the change score (Δ) is defined as the part of the score of *Writing*[2] that is not identical

to *Writing*[1]. Finally, three two-headed arrows are drawn to represent the variance of writing scores at Time 1 (σ_1^2), the variance of the changes (σ_Δ^2), and the covariance between writing scores at Time 1 and the changes ($\sigma_{1\Delta}$).

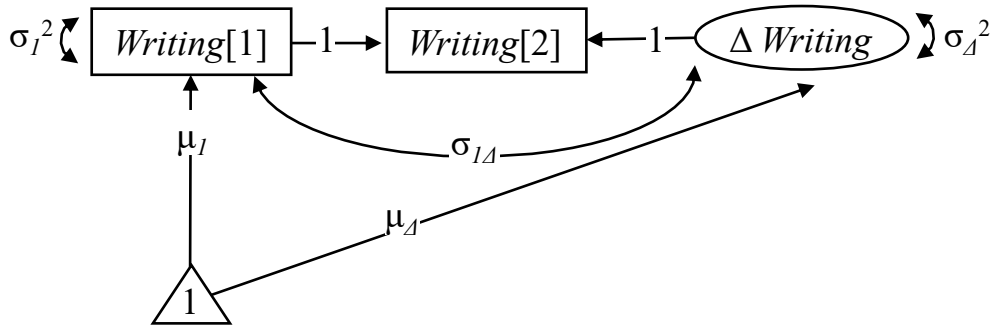
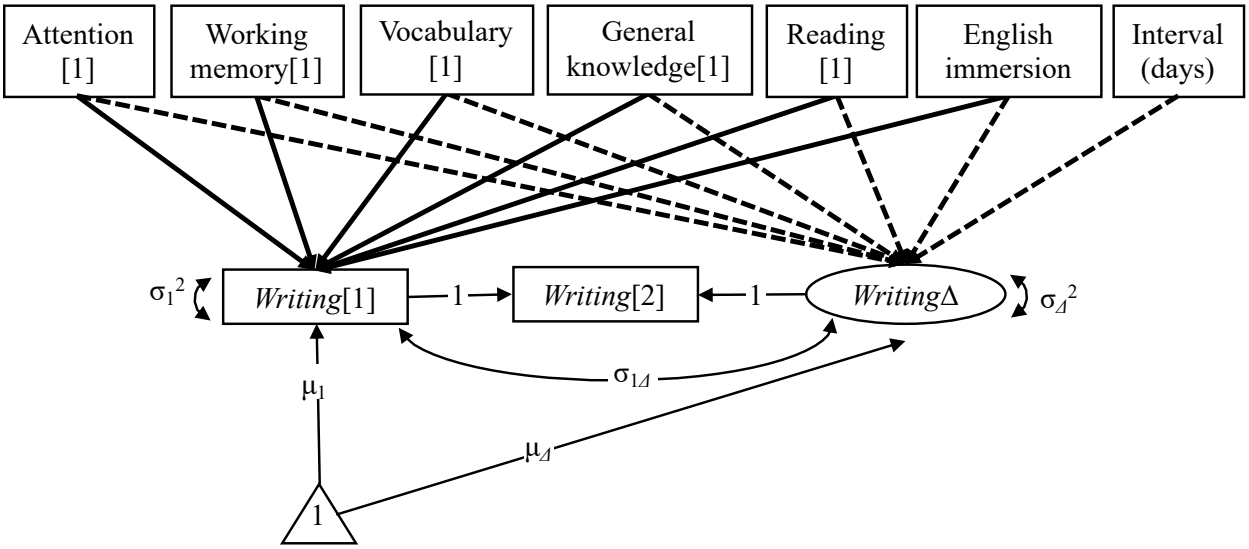


Figure 3.2 Univariate, Unconditional Latent Change Score Model for English Writing Scores
Note. [1] = Time 1, [2] = Time 2, Δ = "change in"

The first research question specifically addressed how the initial levels of general cognitive/language resources and years of English immersion instruction would predict the initial level of English writing scores and writing score changes in multilingual undergraduate students. The five cognitive/language resources were included as predictor variables of initial writing scores and wring score changes. These predictors were attention, working memory capacity, English vocabulary knowledge, general knowledge, and English reading skills. In addition, interval days between the two writing occasions were added as a predictor of writing score changes because interval days, which varied from around five months to around one year, might influence the degree of changes. The univariate, conditional latent change score model that includes predictors is drawn in Figure 3.3. In this figure, paths that predict the initial level of English writing scores are shown in bold lines, while paths that predict changes in English writing scores are shown in dashed lines.



v

Figure 3.3 Univariate, Conditional Latent Change Score Model with Predictors of the Initial Level of, and Changes in, L2 Writing Scores

Note. [1] = Time 1, [2] = Time 2, Δ = "change in"; Parameters that predict the initial level of English writing scores are shown in bold lines, while parameters that predict the change in English writing scores are shown in dashed lines. For clarity, covariances among the predictors are not drawn.

3.4.3 Statistical analysis for the research question 2

The second research question addressed the relationship between the roles of the translator and English writing scores. The roles of the translator were measured in terms of both process and product features. A process feature of the translator was measured by the mean length of language bursts (i.e., the mean number of characters produced between pauses longer than two seconds). A product feature of the translator was measured by the lexical features found in written products. Lexical features were measured by the production of academic words based on the academic word list (AWL; Coxhead, 2000). The number of academic words included in the AWL which were also found in students' essays was calculated, normed by text length. And multiplied by 100. Thus, the second research question examined the relationships among three constructs, i.e., English writing scores, burst lengths, and the percentages of academic words.

Figure 3.4 presents the trivariate, unconditional latent change score model. In this figure, first, three separate latent change scores models were created for English writing scores, burst length, and the use of academic words, respectively. For each of the three variables, mean initial scores (μ_1), mean changes (μ_Δ), variances in initial scores, (σ_1^2), and variances in changes (σ_Δ^2) were estimated (see solid lines in Figure 3.4). In addition, covariances between initial scores and score changes of each variable ($\sigma_{1\Delta}$) were estimated (see double lines in Figure 3.4). Then, three types of relationships among the three variableswere estimated. First, covariances among initial levels of the three variableswere estimated (see solid, bold lines in Figure 3.4). Second, covariances among changes in the three variableswere estimated (see dashed lines in Figure 3.4). Lastly, cross-lagged covariances (i.e., relationships between different variables across different time points) between initial levels and changes across different variables (e.g., a covariance between the initial level of English writing scores and the change in the use of academic words) were estimated (see solid, grey lines in Figure 3.4).

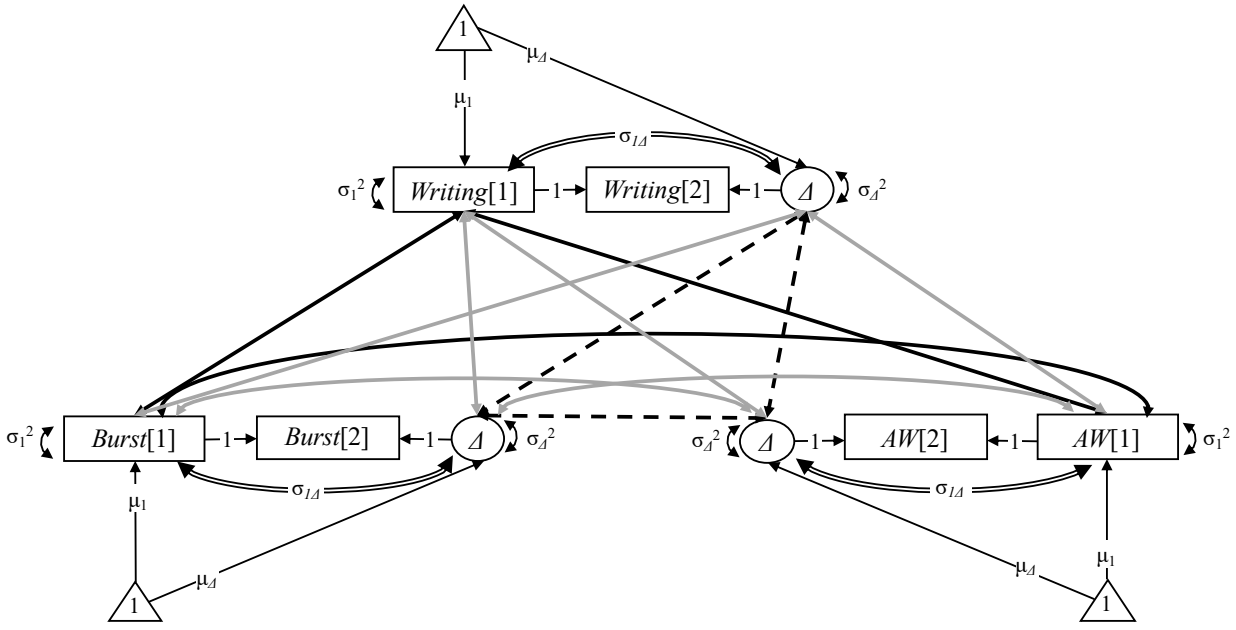


Figure 3.4 Trivariate, Unconditional Latent Change Score Model for Relationships among Initial Levels of, and Changes in, English Writing Score, Burst Length, and Academic word use

Note. [1] = Time 1, [2] = Time 2, Δ = "change in", AW = academic word; Covariances between initial levels and changes for each variable are shown in double lines.; Covariances among initial levels are shown in solid, bold lines.; Covariances among changes are shown in dashed lines.; Cross-lagged covariances are shown in grey lines.

Four hypotheses related to the second research question were that (a) there would be changes in English writing quality, length of bursts, and the number of academic words, respectively, over time, such that writing scores would increase, the mean length of bursts would increase, and academic word counts would increase; (b) initial levels of English writing scores, length of bursts, and academic word percentages would be correlated with each other; (c) changes in English writing scores, length of bursts, and academic word percentages would co-vary with each other; and (d) there would be cross-lagged covariances between initial levels and changes across different variables.

3.4.4 Statistical analysis for the research question 3

The third research question addressed the longitudinal relationship among three literacy-related variables. These are English writing ability, English reading ability, and English vocabulary knowledge, each of which was measured on two occasions. Specifically, the dissertation addressed two sub-research questions, 3a and 3b. The research question 3a examined the longitudinal relationships among English writing, reading, and vocabulary. The research question 3b tested whether a common latent variable of English as a second language literacy that was informed by English writing, reading, and vocabulary could be constructed across the two times of measurement. Each analysis for the two sub-research questions is elaborated below.

3.4.4.1 Examining longitudinal relationships among English writing, reading, and vocabulary

As in research question 2, research question 3a addressed all of the possible cross-sectional and longitudinal relationships among the three variables (i.e., English writing, reading, and vocabulary): covariances among initial levels, covariances among changes, covariances

between initial levels and changes of each variable, and cross-lagged covariances between initial levels and changes across different variables (e.g., covariance between the initial level of English vocabulary and the change in English writing). Figure 3.5 shows the hypothetical model related to the research question 3a.

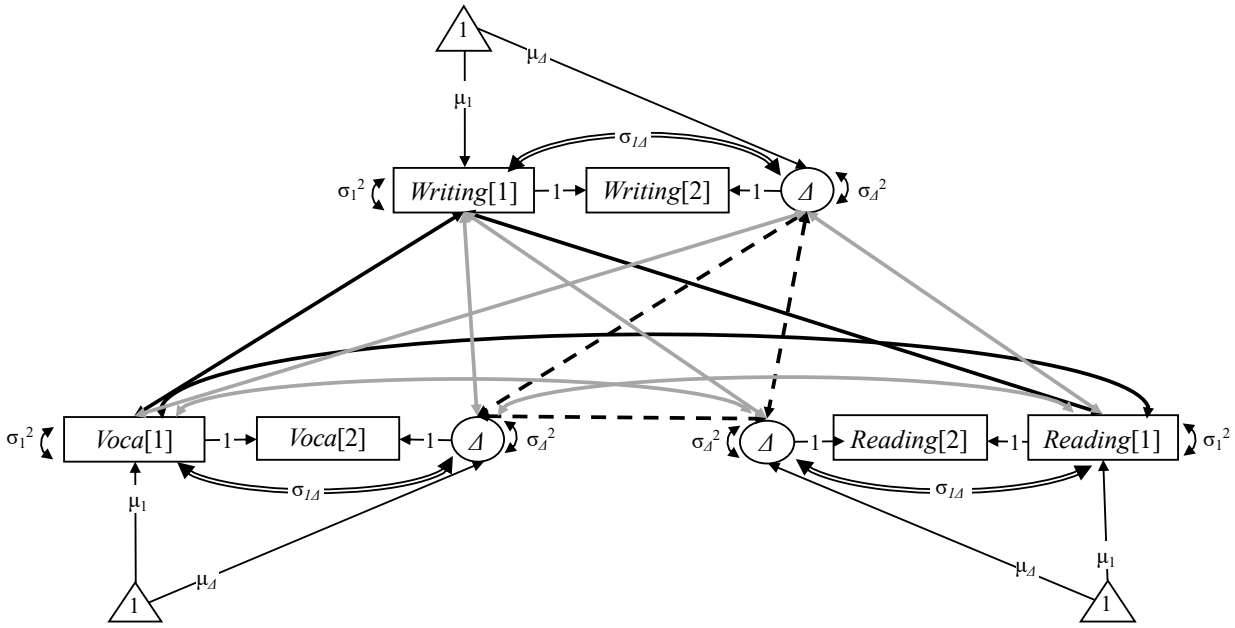


Figure 3.5 Trivariate, Unconditional Latent Change Score Model for Relationships among English Writing, Vocabulary, and Reading
Note. [1] = Time 1, [2] = Time 2, Δ = "change in", Voca = Vocabulary; Covariances between initial levels and changes for each construct are shown in double lines.; Covariances among initial levels are shown in solid, bold lines.; Covariances among changes are shown in dashed lines.; Cross-lagged covariances are shown in grey lines.

Regarding the 3a research question, four hypotheses were that (a) there would be gains in English writing scores, vocabulary scores, and reading scores; (b) initial levels of English writing scores, vocabulary scores, and reading scores would be correlated with each other; (c) changes in English writing scores, vocabulary scores, and reading scores would co-vary with each other; and (d) there would be cross-lagged covariances between initial levels and changes across different variables.

3.4.4.2 Testing a common latent variable of English literacy across time

For the research question 3a, a latent change score model for each of English writing, reading, and vocabulary was separately constructed. Instead of creating the three separate models, given that the three variables are related to English literacy, in the research question 3b, a common latent variable (or factor) informed by English writing, reading, and vocabulary scores was created and tested. Specifically, the research question 3b tested whether a common latent variable of English literacy could be constructed across the two times of measurement, and, then, whether there would be changes in mean scores on the latent variable across the two occasions. Creating a common latent variable is of interest because it can provide a parsimonious understanding of related observed variables in an explicit measurement model.

In longitudinal measurement, the presence of a latent variable that represents multiple observed variables measured at each time of measurement can be examined via longitudinal measurement invariance. Essentially, longitudinal measurement invariance tests the equality of factor structure across time (i.e., whether the same latent factor is measured in the same manner across different measurement times; Little, 2013; Vandenberg & Lance, 2000). Measurement invariance tests have benefits of removing measurement errors of observed variables. Also, supporting measurement invariance ensures a solid comparison of mean scores of a latent variable. In short, longitudinal measurement invariance tests whether the relationship between the latent variables and the observed variables is equal across separate times of assessment, regardless of whether average latent variable scores increase, decrease, or remain the same over time.

To examine longitudinal measurement invariance, confirmatory factor analysis was used. Measurement invariance tests involve a series of comparisons of two nested models in which one

model is more constrained than the other (Little, 2013; Vandenberg & Lance, 2000; Widaman & Reise, 1997). Testing measurement invariance is generally conducted with four sequential stages: Configural invariance and three stages of measurement invariance (metric or weak, scalar or strong, and strict).

The first step is to create a baseline model for testing configural invariance across time. To test configural invariance, the latent variable scores across time are standardized by fixing their latent means to zero and their variances to one, so that they can have standardized metric. Invariance at the configural level indicates that the latent variable is formed by the same number of observed variables across time.

If configural invariance is met, the next step is to test metric or weak invariance by constraining factor loadings (i.e., the loading of the observed variables on the latent variable) to be equal across time. Given the configural invariance, while the latent variables measured at the initial time remain standardized ($M = 0$, $SD = 1$), the means of latent variables measured at subsequent times remain zero, but their variances are freely estimated. Metric measurement invariance indicates that each observed variable contributes to the latent variable to a similar degree across time.

If metric invariance supported, scalar or strong invariance is tested by constraining intercepts of observed variables to be equivalent across time. Given the scalar invariance, while the latent variables measured at the initial time remain standardized, the means and covariances of the latent variables measured at subsequent times are freely estimated. Scalar invariance indicates that mean differences in the latent variable capture all of the mean differences in the shared variance of the observed variables. Scalar invariance is a prerequisite for testing mean differences in the latent variables across time.

If scalar invariance is met, the final step is to test for strict invariance by constraining residuals of the observed variables to be equivalent. This step is not necessary for testing latent mean differences because residuals are not related to the latent variables. Thus, because the dissertation intended to examine the equality of the latent structure and test latent mean differences, strict invariance was not further considered.

In summary, configural invariance concerns equivalence of the latent structure organization, metric invariance concerns equivalence of factor loadings, and scalar invariance concerns equivalence of intercepts of observed variables. The summary of longitudinal measurement invariance test steps from configural invariance to scalar invariance is presented in Table 3.4 in which repeated measurements occurred from time 1 to *T*.

Table 3.4 Summary of Longitudinal Measurement Invariance

Invariance	Factor loading	Intercept of observed variable	α_1	φ_1	$\alpha_{2 \ldots T}$	$\varphi_{2 \ldots T}$
Configural	Freely estimated	Freely estimated	0	1	0	1
Metric or weak	Equivalent across time	Freely estimated	0	1	0	Freely estimated
Scalar or strong	Equivalent across time	Equivalent across time	0	1	Freely estimated	Freely estimated

Note. α = latent mean; φ = latent variance; T = total measurement times

When configural invariance across groups was met, measurement invariance tests were conducted for two nested models: Metric vs. Configural and Scalar vs. Metric. Invariance was assessed in terms of chi-square difference ($\Delta\chi^2$), and the CFI difference (ΔCFI = CFI_1 – CFI2). When $\Delta\chi^2$ is insignificant and a ΔCFI value is greater than −.01, measurement invariance is warranted (Dimitrov, 2010).

Figure 3.6 shows a common latent factor model across time to test scalar measurement invariance. English writing (W), English reading (R), and English vocabulary (V) were repeatedly measured at Time 1 [1] and Time 2 [2]. Each set of these three observed variables

reflects the underlying, latent *English Literacy* variable of interest on both occasions. In addition, residual covariances for each observed variable across two time points are represented as double-headed arrows. For the test of metric invariance, the factor loadings of W, R, V on the latent variables are set to be equivalent across time: $\lambda_{W,}$ λ_R, and $\lambda_{V.}$ Next, for the test of scalar invariance, the intercepts of W, R, V are constrained to be equivalent across time: τ_W $\tau_{R,}$ and τ_V. Given scalar invariance with the mean (α_1) and variance (φ_1^2) of the latent variable at Time 1 being zero and one, respectively, the mean (α_2) and variance (φ_2^2) of the latent variable at Time 2 are estimated to compare latent factor means across time. The covariance of the latent *L2 Literacy* variable between Time 1 and Time 2 (φ_{12}) is also estimated.

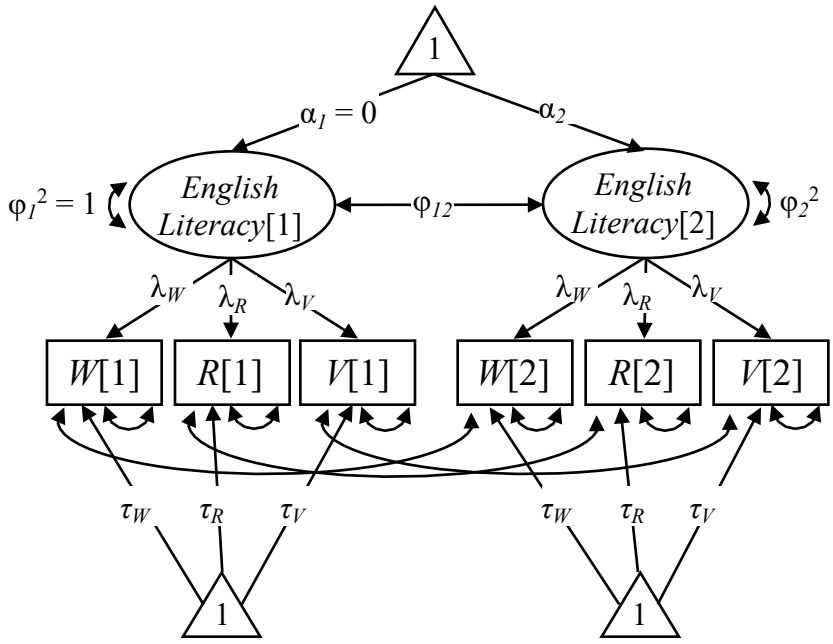


Figure 3.6 Common Factor Model of English Literacy Across Two Time Points
Note. W = writing; R = reading; V = vocabulary

To generate and test latent variable models including latent change score models and common factor models, *R* (R Core Team, 2018) and *lavaan* packages (Rosseel, 2012) were used. To handle the missing data, a full information maximum likelihood (FIML) approach was used, which allows participants with missing data to be retained and provides the least biased estimates

of missing data, under the assumption that data are missing at random (Buhi, Goodson, & Neilands, 2008).

Model fit statistics were computed using maximum-likelihood estimations with robust standard errors (MLR). The MLR parameter estimation can handle non-normality and missing data, yielding a robust chi-square test (Kaplan, Kim, & Kim, 2009). To evaluate model fit, three goodness-of-fit measures were used: the robust χ^2 (Chi-square), comparative fit index (CFI), and standardized root mean square residual (SRMR). The χ^2 measures absolute fit of the model to the data. Indicators of good model fit included robust CFI statistics greater than .95 and SRMR less than .08, while indicators of acceptable model fit included CFI statistics greater than .90 (Hu & Bentler, 1999). Unless otherwise noted, χ^2 and CFI values are robust statistics. Root mean square error of approximation (RMSEA) was not included because it is not found to be most adequate for sample sizes smaller than 250 (Hu & Bentler, 1999).

4 LONGITUDINAL RELATIONSHIP BETWEEN RESOURCES, YEARS OF ENGLISH IMMERSION INSTRUCTION, AND ENGLISH WRITING SCORES

The first research question focused on the relationship among general cognitive and language resources, years of English immersion instruction, and English writing scores over time. Specifically, the research question examines how initial levels of general cognitive/language resources and years of English immersion instruction predict the initial level of English writing scores and changes in English writing scores over time in multilingual undergraduate students. English writing ability was measured two times (Time 1 and Time 2) with an interval of at least five months. General cognitive and language resources measured at the initial time point (Time 1) included attention, working memory, English vocabulary knowledge, general knowledge, and

English reading skills. A latent change score modeling approach was used. Results are provided in Section. 4.1, followed by discussion provided in Section 4.2.

4.1 Research Question 1 Results

4.1.1 Descriptive statistics

For the first research question, 77 participants' data were analyzed. Among them, the Stroop results of one student were not recorded due to technical errors. To handle the missing data, a full information maximum likelihood (FIML) approach was used. Table 4.1 presents the descriptive statistics of writing scores at two time points, writing score changes, English immersion years, intervals in days between the two writing tasks, Stroop test scores, working memory test scores, general knowledge test scores, English vocabulary test scores, and English reading comprehension test scores. Writing score changes were calculated by subtraction (i.e., writing scores at Time 2 minus writing scores at Time 1). Figure 4.1 presents these variables' scatter plots and histograms. On average, participants were educated via the English language for 8.44 years (*SD* = 6.32). The intervals in days between the two writing tasks ranged from 141 to 366 with a mean of 211.79 (*SD* = 50.80). The skewness values ranged from .02 to 1.70, and the kurtosis values ranged from –1.35 to 3.50.

Table 4.1 Descriptive Statistics for Variables Related to Research Question 1

Variable	*N*	Mean	*SD*	Min.	Max.	Skewness	Kurtosis
English writing at Time 1	77	2.92	1.16	1	5.5	.39	–.35
English writing at Time 2	77	4.04	1.00	1.50	6	.08	–.36
Change in writing score	77	1.12	1.19	–1	4.50	.68	.34
English immersion years	77	8.44	6.32	1	22	.36	–1.35
Interval in days	77	211.83	50.80	141	366	1.70	2.18
Stroop score	76	.22	.17	.01	.96	1.49	3.50
Working memory	77	40.69	11.60	13	63	.07	–.81
General knowledge	77	16.62	4.39	7	25	.04	–.78
English vocabulary	77	555.21	27.93	499	612	.02	–.88
English reading	77	564.71	26.98	513	653	.49	.68

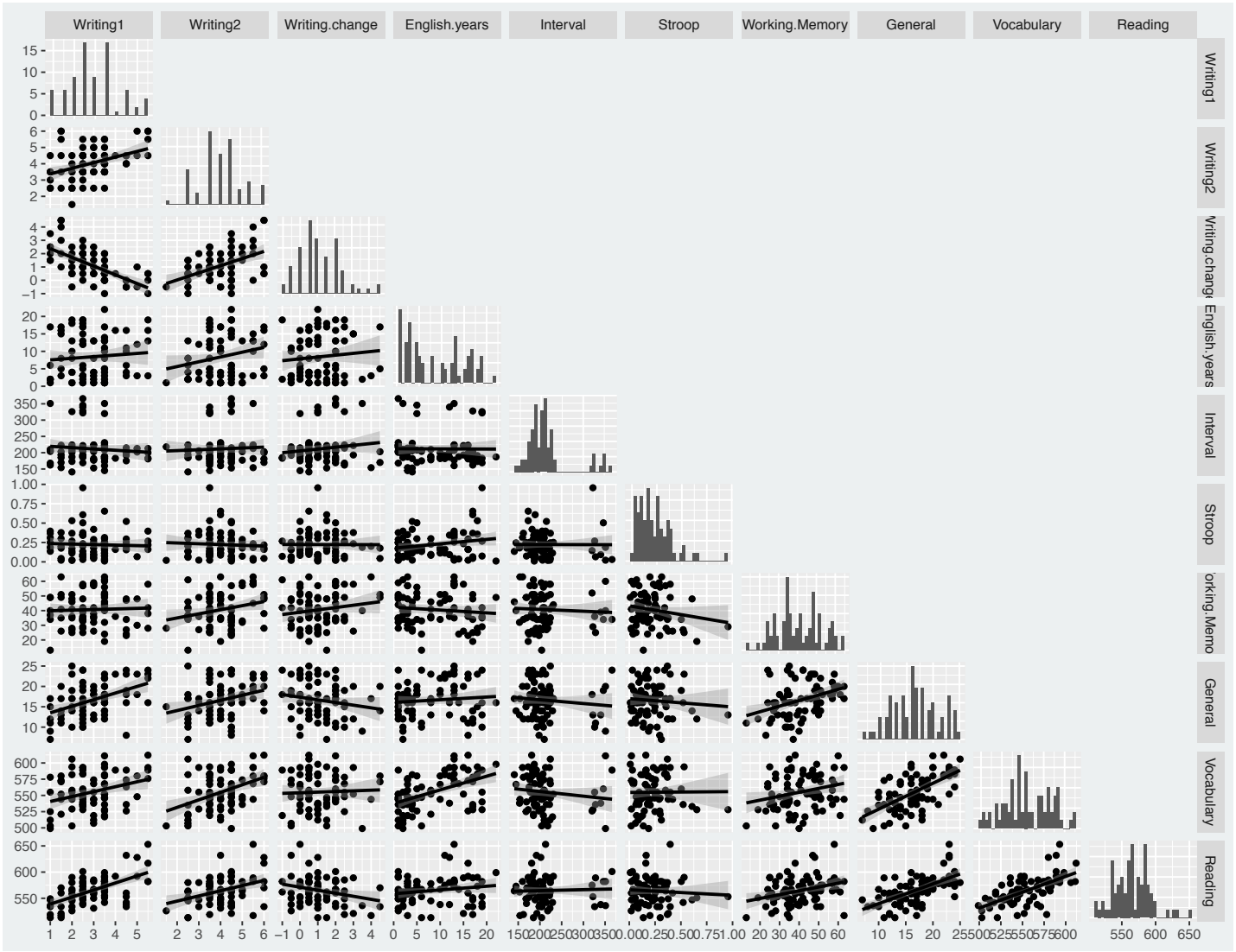


Figure 4.1 Scatter Plots and Histograms of Variables Related to Research Question 1

Note. 1 = Time 1, 2 = Time 2, English.years = English immersion years; General = general knowledge.; Diagonal graphs show histograms of each variable. On scatter plots, lines indicate linear predictions from a linear regression model of the two variables.

Overall, average writing scores increased from 2.92 to 4.04 over time, while standard deviations decreased from 1.16 to 1.00. Figure 4.2 graphically shows changes in writing scores over time. Higher scorers at Time 1 who received scores four or higher tended to receive similar scores at Time 2 within score ranges from four to six. In contrast, lower scorers at Time 1 who received scores 3.5 or lower tended to receive higher scores at Time 2 than at Time 1.

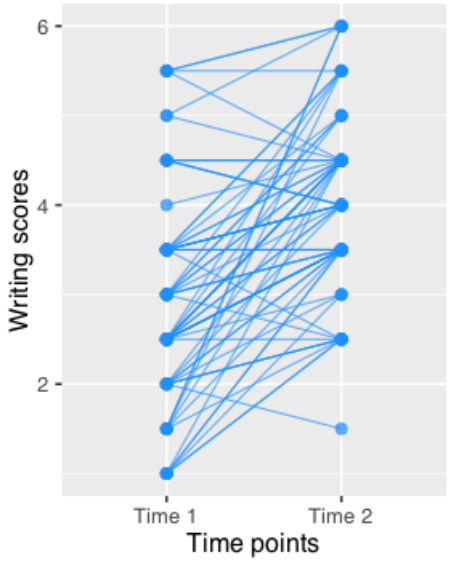


Figure 4.2 Writing Score Changes over Time (N = 77)

4.1.2 Correlation analysis

Table 4.2 shows correlations among the variables related to the research question 1. Writing scores across the two times were moderately correlated ($r = .40$, $p < .01$). Changes in writing scores were negatively correlated with initial writing scores ($r = –.64$, $p < .01$) and initial reading scores ($r = –.26$, $p < .05$). English immersion years, intervals in days, and Stroop test results were not significantly correlated with writing scores across the two times. Working memory capacity scores were moderately correlated with writing scores at Time 2 only ($r = .23$, $p < .05$), and moderately correlated with general knowledge, English vocabulary, and English reading scores ($.25 < r < .37$). General knowledge, English vocabulary, and English reading

scores were significantly correlated with both of the writing scores, ranging from $r = .29$ ($p < .05$) to $r = .58$ ($p < .01$). Additionally, years of English immersion years were significantly correlated only with English vocabulary knowledge ($r = .48$, $p < .01$). Intervals were not related with any other variables. Finally, general knowledge, English vocabulary, and English reading comprehension scores were strongly correlated with each other, ranging from $r = .61$ ($p < .01$) to $r = .63$ ($p < .01$).

Table 4.2 Correlations among Variables Related to Research Question 1

Variable	1	2	3	4	5	6	7	8	9	10
1 English writing at Time 1	1									
2 English writing at Time 2	.40**	1								
3 Change in writing score	–.64**	.45	1							
4 English immersion years	.09	.22	.10	1						
5 Interval in days	–.09	.05	.13	–.01	1					
6 Stroop score	–.05	–.06	.00	.22	–.00	1				
7 Working memory	.04	.24*	.16	–.11	–.05	–.17	1			
8 General knowledge	.43**	.29*	–.18	.10	–.11	–.08	.37**	1		
9 English vocabulary	.32**	.42**	.04	.48**	–.13	.01	.25*	.62**	1	
10 English reading	.58**	.37**	–.26*	.17	.04	–.07	.31**	.61**	.63**	1

Note. Except for Stroop scores ($N = 76$), *N*s for scores in all variables are 77; ** indicates $p < .01$, and * indicates $p < .05$.

4.1.3 Results of latent change score modeling

As a baseline model, a univariate, unconditional latent change score model (i.e., model without predictors of writing scores and score changes) was tested. The results of the baseline model are provided in Figure 4.3 and Table 4.3. Because this model was identified without degree of freedom, model fit indices were not calculated. Changes in writing scores were significant. In addition, there were significant variances in changes, which indicates that there

was variability in the students' writing score changes. These changes in writing scores had a negative relationship with the initial level of writing scores.

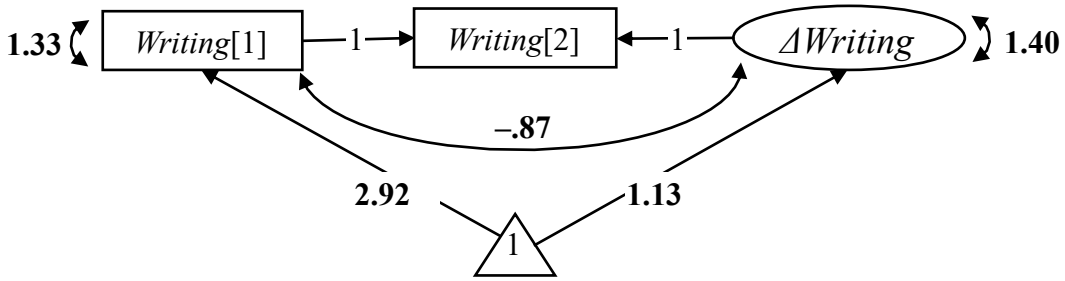


Figure 4.3 Unconditional Latent Change Score Model (Baseline Model)
Note. Δ = "change in", [1] = Time 1, [2] = Time 2. Arrows beginning from △1 indicate estimates for mean scores. Estimates are unstandardized coefficients. All estimates were significant at $p < .01$.

Table 4.3 Results of the Unconditional Latent Change Score Model

Estimate in writing model	Estimate	*SE*	*z*	*p*	Standardized Estimate
Intercept of writing[1]	2.92	.13	2.28	< .01	2.54
Intercept of Δwriting	1.13	.16	8.30	< .01	.95
Variance of writing[1]	1.33	.20	6.69	< .01	1.00
Variance of Δwriting	1.40	.29	5.64	< .01	1.00
Covariance between writing[1] and Δwriting	–.87	.16	–5.46	< .01	–.64

Note. Δ = "change in"; [1] = Time 1.

Given the baseline model, the conditional latent change score model as presented in Figure 3.3 was tested by adding predictors of writing scores and score changes. These predictors were attention, working memory capacity, English vocabulary knowledge, general knowledge, years of English immersion instruction, English reading skills, and intervals between the two writing tests. When the model was tested, because general knowledge, vocabulary, and reading scores were multicollinear (i.e., strongly correlated with each other; $r > .60$), suppression effects (i.e., when two or more predictors are strongly correlated, unique contributions of less strong predictors disappear in the presence of the strongest predictor; Tabachnick & Fidell, 2012, p.

155) occurred in predicting initial writing scores. That is, because correlations with initial writing scores were stronger for reading scores (r = .58) than for vocabulary scores (r = .32) and general knowledge scores (r = .43), the role of vocabulary and general knowledge scores on predicting initial writing scores disappeared in the presence of reading scores.

To resolve the multicollinearity issue, a latent factor consisting of vocabulary, general knowledge, and reading scores was constructed and tested in the conditional latent change score model. The model fit badly (χ^2 = 63.69, df = 17, p < .01, CFI = .738, SRMR = .124). Thus, the latent variable was not further considered.

Instead of creating the latent variable, to handle multicollinearity, three separate latent change score models were tested, such that while all of the three models included attention, working memory, immersion years, and intervals as predictors, one model included vocabulary scores (henceforth, 'vocabulary model'), another included general knowledge scores (henceforth, 'general knowledge model'), and the other included reading scores (henceforth, 'reading model').

First, the vocabulary model was tested. It fit perfectly (robust χ^2 = .37, df = 1, p = .54, robust CFI = 1.000, SRMR = .008). The results of the latent change score modeling are presented in Figure 4.4 and Table 4.4. For clarity purposes, the results of baseline latent change modeling were not drawn in Figure 4.4 (but presented under the heading 'parameters of writing scores' in Table 4.4) because all estimates in the vocabulary model were similar to those in the baseline model, and the focus of the vocabular model was on the predictions of initial writing scores and writing score changes. In terms of predicting initial writing scores and writing score changes, results revealed one significant path only, such that vocabulary scores significantly predicted initial writing scores. Attention, working memory, and English immersion years were

not related to either initial writing scores or changes in writing score. Intervals were also not related to changes in writing scores. Overall, 11% and 6% of the variance in initial writing scores and changes in writing scores were explained, respectively.

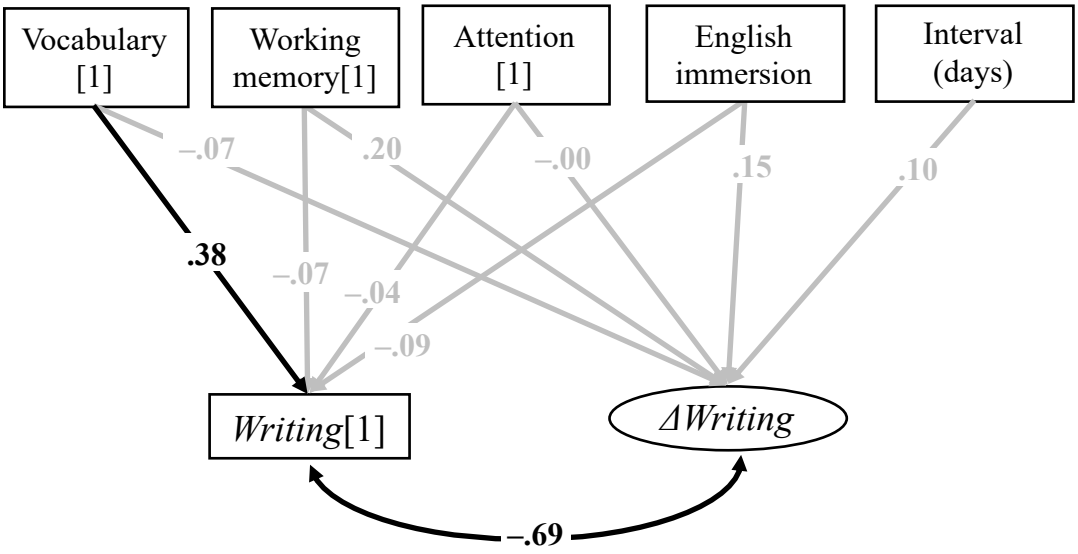


Figure 4.4 *Latent Change Score Model with English Vocabulary Scores*
Notes. Δ = "change in"; [1] = Time 1. Estimates are standardized coefficients. Significant coefficients are shown in black color, while nonsignificant coefficients are shown in grey color. Covariances among predictors are shown in Table 4.4.

Table 4.4 Results of Latent Change Score Model with English Vocabulary Scores

Predicted variable	Predictors	Estimate	*SE*	*z*	*p*	Standardized Estimate
Writing[1]	Vocabulary[1]	.02	.01	2.66	< .01	.38
Writing[1]	Working memory[1]	–.01	.01	–.61	.54	–.07
Writing[1]	Attention[1]	–.29	.68	–.43	.67	–.04
Writing[1]	English immersion	–.02	.02	–.72	.47	–.09
ΔWriting	Vocabulary[1]	–.00	.01	–.53	.60	–.07
Δwriting	Working memory[1]	.02	.01	1.58	.11	.20
Δwriting	Attention[1]	–.00	.61	–.01	1.00	–.00
Δwriting	English immersion	.03	.02	1.22	.22	.15
ΔWriting	Interval (days)	.00	.00	1.54	.12	.10
Parameters of writing scores		Estimate	*SE*	*z*	*p*	Standardized Estimate
Intercept of writing[1]		2.92	.12	23.59	< .01	2.54
Intercept of Δwriting		1.12	.13	8.57	< .01	.95
Variance of writing[1]		1.18	.16	7.62	< .01	1.00
Variance of Δwriting		1.31	.24	5.54	< .01	1.00
Covariance between writing[1] and Δwriting		–.85	.15	–5.66	< .01	–.69
Covariance among predictors		Estimate	*SE*	*z*	*p*	Standardized Estimate
Vocabulary[1] ↔ Working memory[1]		80.39	37.74	2.13	< .05	.25
Vocabulary[1] ↔ Attention[1]		.03	.53	.05	.96	.00
Vocabulary[1] ↔ English immersion		83.18	18.52	4.49	< .01	.48
Vocabulary[1] ↔ Interval (days)		–183.30	186.82	–.98	.33	–.13
Working memory[1] ↔ Attention[1]		–.33	.24	–1.37	.17	–.17
Working memory[1] ↔ English immersion		7.60	7.90	–.96	.34	–.11
Working memory[1] ↔ Interval (days)		31.81	62.43	–.51	.61	–.06
Attention[1] ↔ English immersion		.22	.15	1.47	.14	.21
Attention[1] ↔ Interval (days)		–.03	1.38	–.02	.98	–.00
English immersion ↔ Interval (days)		–2.07	38.43	–.05	.96	–.01

Note. Δ = "change in", [1] = Time 1, ↔ = "covariance with"

Second, the general knowledge model was tested. It fit the data perfectly (χ^2 = .40, *df* = 1, *p* = .53, CFI = 1.000, SRMR = .009). The results of the latent change score modeling are presented in Figure 4.5 and Table 4.5. The estimates of the relationship between initial writing scores and writing changes were similar to those in the baseline model, and thus they were not drawn in Figure 4.5. In predicting initial writing scores, higher levels of initial writing scores were predicted by higher levels of initial general knowledge scores. In predicting changes in wring scores, greater gains in writing scores were predicted by lower levels of initial general knowledge scores and higher levels of working memory capacity. Attention and English immersion years were not related to either initial writing scores or changes in writing score. Intervals were also not related to writing score changes. Overall, 21% and 13% of the variance in initial writing scores and changes in writing scores were explained, respectively.

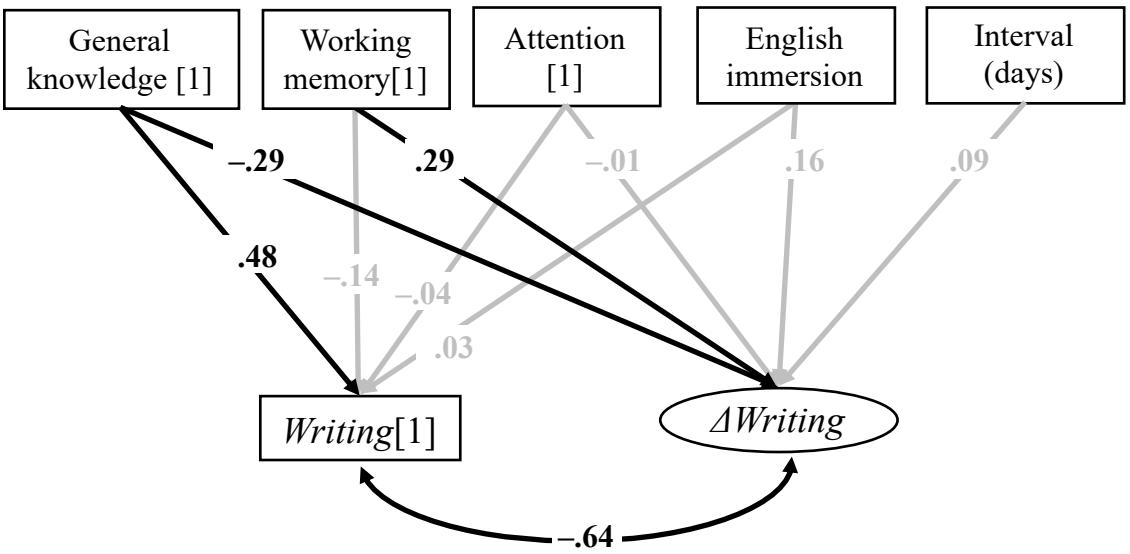


Figure 4.5 *Latent Change Score Model with General Knowledge Scores*
Notes. Δ = "change in"; [1] = Time 1. Estimates are standardized coefficients. Significant coefficients are shown in black color, while nonsignificant coefficients are shown in grey color. Covariances among predictors are shown in Table 4.5.

Table 4.5 *Results of Latent Change Score Model with General Knowledge Scores*

Predicted variable	Predictors	Estimate	*SE*	*z*	*p*	Standardized Estimate
Writing[1]	General knowledge[1]	.13	.04	3.63	< .01	.48
Writing[1]	Working memory[1]	–.01	.01	–1.17	.24	–.14
Writing[1]	Attention[1]	–.28	.63	–.44	.66	.04
Writing[1]	English immersion	.01	.02	.33	.74	.03
Δwriting	General knowledge[1]	–.08	.03	–2.59	< .01	–.29
Δwriting	Working memory[1]	.03	.01	2.36	< .05	.29
Δwriting	Attention[1]	–.08	.60	–.13	.90	–.01
Δwriting	English immersion	.03	.02	1.48	.14	.16
Δwriting	Interval (days)	.00	.00	1.18	.24	.09
Parameters of writing scores		Estimate	*SE*	*z*	*p*	Standardized Estimate
Intercept of writing[1]		2.92	.12	24.96	< .01	2.54
Intercept of Δwriting		1.12	.13	8.90	< .01	.95
Variance of writing[1]		1.05	.15	6.85	< .01	1.00
Variance of Δwriting		1.22	.22	5.50	< .01	1.00
Covariance between writing[1] and Δwriting		–.72	.13	–5.58	< .01	–.64
Covariance among predictors		Estimate	*SE*	*z*	*p*	Standardized Estimate
General knowledge[1] ↔ Working memory[1]		18.26	4.97	3.68	< .01	.37
General knowledge[1] ↔ Attention[1]		–.06	.08	–.77	.44	–.08
General knowledge[1] ↔ English immersion		2.63	3.12	.85	.40	.10
General knowledge[1] ↔ Interval (days)		–23.78	28.35	–.84	.40	–.11
Working memory[1] ↔ Attention[1]		–.33	.24	–1.37	.17	–.17
Working memory[1] ↔ English immersion		–7.60	7.90	–.96	.34	–.11
Working memory[1] ↔ Interval (days)		–31.81	62.43	–.51	.61	–.06
Attention[1] ↔ English immersion		.22	.15	1.47	.14	.21
Attention[1] ↔ Interval (days)		–.03	1.38	–.02	.98	–.00
English immersion ↔ Interval (days)		–2.07	38.43	–.05	.96	–.01

Note. Δ = "change in", [1] = Time 1, ↔ = "covariance with"

Lastly, the reading model was tested. It fit adequately (robust $\chi^2 = 2.84$, $df = 1$, $p = .09$, CFI = .975, SRMR = .022). The results of the latent change score modeling are presented in Figure 4.6 and Table 4.6. The estimates of the relationship between initial writing scores and writing changes were similar to those in the baseline model, and thus they are not drawn in Figure 4.6. In predicting initial writing scores, higher levels of initial writing scores were predicted by higher levels of initial reading scores. Greater gains in writing scores were predicted by lower levels of initial reading scores and higher levels of working memory capacity. Attention and English immersion years were not related to either initial writing scores or changes in writing score. Intervals were also not related to changes in writing scores. Overall, 36% and 18% of the variance in initial writing scores and changes in writing scores were explained, respectively.

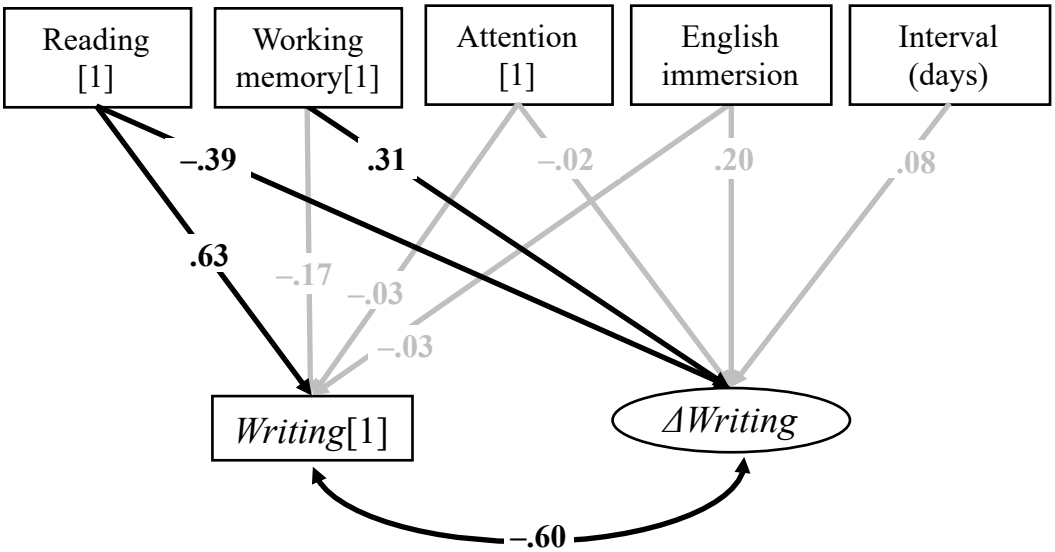


Figure 4.6 Latent Change Score Model with English Reading Scores
Note. Δ = "change in"; [1] = Time 1. Estimates are standardized coefficients. Significant coefficients are shown in black color, while nonsignificant coefficients are shown in grey color. Covariances among predictors are shown in Table 4.6.

Table 4.6 Results of Latent Change Score Model with Reading Scores

Predicted variable	Predictors	Estimate	*SE*	*z*	*p*	Standardized Estimate
Writing[1]	Reading[1]	.03	.00	6.94	< .01	.63
Writing[1]	Working memory[1]	–.02	.01	–1.70	.09	–.17
Writing[1]	Attention[1]	–.18	.57	–.31	.76	–.03
Writing[1]	English immersion	–.01	.02	–.34	.73	–.03
Δwriting	Reading[1]	–.02	.01	–3.63	< .01	–.39
Δwriting	Working memory[1]	.03	.01	2.68	< .01	.31
Δwriting	Attention[1]	–.14	.57	–.25	.80	–.02
Δwriting	English immersion	.04	.02	1.94	.05	.20
Δwriting	Interval (days)	.00	.00	1.09	.28	.08
Parameters of writing scores		Estimate	*SE*	*z*	*p*	Standardized Estimate
Intercept of writing[1]		2.92	.11	27.78	< .01	2.54
Intercept of Δwriting		1.12	.12	9.21	< .01	.95
Variance of writing[1]		.85	.14	6.16	< .01	1.00
Variance of Δwriting		1.13	.21	5.49	< .01	1.00
Covariance between writing[1] and Δwriting		–.59	.13	–4.57	< .01	–.60
Covariance among predictors		Estimate	*SE*	*z*	*p*	Standardized Estimate
Reading[1] ↔ Working memory[1]		97.07	37.83	2.57	< .05	.31
Reading[1] ↔ Attention[1]		–.32	.46	–.69	.49	–.07
Reading[1] ↔ English immersion		28.39	18.32	1.55	.12	.17
Reading[1] ↔ Interval (days)		48.82	143.48	.34	.73	.04
Working memory[1] ↔ Attention[1]		–.33	.24	–1.37	.17	–.17
Working memory[1] ↔ English immersion		–7.60	7.90	–.96	.34	–.11
Working memory[1] ↔ Interval (days)		–31.81	62.43	–.51	.61	–.06
Attention[1] ↔ English immersion		.22	.15	1.47	.14	.21
Attention[1] ↔ Interval (days)		–.03	1.38	–.02	.98	–.00
English immersion ↔ Interval (days)		–2.04	38.43	–.05	.96	–.00

Note. Δ = "change in", [1] = Time 1, ↔ = "covariance with"

4.2 Research Question 1 Discussion

4.2.1 Summary of results

The first research question examined the extent to which initial levels of general cognitive/language resources and years of English immersion instruction predicted the initial level of English writing scores and changes in English writing scores in multilingual undergraduate students using latent change score modeling approaches. In predicting initial writing scores and writing score changes, three different latent change score models were tested. First, in the vocabulary model, higher initial vocabulary scores predicted higher initial writing scores. Second, in the general knowledge model, higher initial general knowledge scores predicted both higher initial writing scores and lower writing score gains, and higher initial working memory scores predicted higher writing score gains. These results in the general knowledge model were similar in the reading model. However, the reading model explained more variance in writing scores and score changes than the general knowledge model. The summary of these three models is presented in Table 4.7. Additionally, it is worth mentioning that writing score gains had the negative relationship with initial writing scores in all of the three models. One important caveat when considering the results of these models is that each model was piecemeal because the three predictors originally proposed (i.e., reading, vocabulary, and general knowledge) could not be included in a single model due to multicollinearity. Thus, the results have the limitation of not capturing the whole picture of the related variables.

Table 4.7 Summary of the Three Latent Change Score Models

Model	Predicting initial writing scores		Predicting writing score changes	
	Significant predictor	R^2	Significant predictor	R^2
Vocabulary model	Initial vocabulary (+)	.11	None	.06
General knowledge model	Initial general knowledge (+)	.21	Initial general knowledge (–) Working memory (+)	.13
Reading model	Initial reading (+)	.36	Initial reading (–) Working memory (+)	.18

Note. Predicting directions (positive vs. negative) are shown in parentheses.

Below, the relationship between initial writing scores and writing score changes is first discussed. Then, the role of each predictor (i.e., attention, working memory, vocabulary, general knowledge, reading, years of English immersion instruction, and intervals) is discussed.

4.2.2 Relationship between initial writing scores and writing score changes

The average change in English writing scores was 1.12 (*SD* = 1.19). In addition, English writing score changes showed a negative relation with initial L2 writing scores, which indicates that greater gains were more closely linked to multilingual students who received lower initial writing scores than those who received higher initial writing scores. To illustrate, three groups with different levels of writing score gains were compared: one group of students without gains (ranging from –.1 to zero; henceforth, 'no-gain group', *n* = 18), another of students with lower gains (ranging from .5 to 1.5; henceforth, 'lower-gain group'; *n* = 36), and another group of students with higher gains (ranging from two to 4.5; henceforth, 'higher-gain group'; *n* = 23). Both higher- and lower-gain groups tended to receive lower initial writing scores with means of 2.13 (*SD* = .80) and 2.89 (*SD* = 1.03), respectively, though the lower-gain group had a higher mean than the higher-gain group. On the other hand, the mean initial writing score of the no-gain group was 3.97 (*SD* = .96), indicating that this group was composed mainly of students who received higher initial writing scores. Figure 4.7 shows a boxplot for the three gain groups'

initial writing scores. In the figure, a horizonal line and a number within each group' box indicates the group's median and mean of initial writing scores, respectively. Overall, it appears that L2 students who received lower scores at Time 1 were more likely to achieve greater gains in writing scores at Time 2.

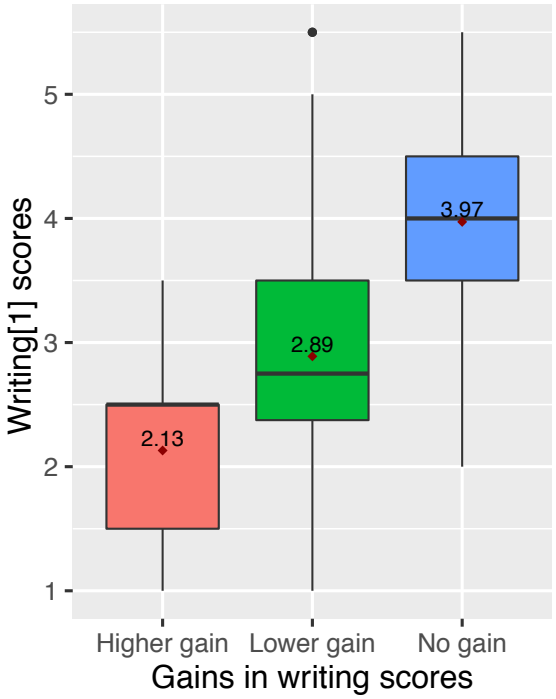


Figure 4.7 Initial English Writing Scores of Three Groups with Different Writing Score Gains
Note. A horizontal line within each box indicates each group's median.; A dot with a number within each box indicates each group's mean.

The finding that lower scorers at Time 1 tended to have greater gains than higher scorers at Time 1 may link to the power law of practice (i.e., performance improves in speed at a decreasing exponential rate). That is, as a power function of amounts of practices, for lower writing scorers at Time 1, improvements in writing may have been large and rapid, while for higher writing scorers at Time 2, improvements in writing may have been small and slow (Kellogg & Whiteford, 2009). In addition, higher scorers at Time 1 might have already been proficient in producing persuasive essays in response to SAT prompts, thus showing smaller gains.

4.2.3 Role of attention in English writing

Attention was found to have no role in predicting either initial English writing scores or English writing score changes, indicating that attention as measured by the Stroop test was not related to English writing ability either cross-sectionally or longitudinally. To visually illustrate the lack of the relationship between attention and writing score changes, the three groups with different levels of writing score gains were compared (see Figure 4.8). The mean scores were similar among the higher-gain group (M = .20, SD = .14), the lower-gain group (M = .25, SD = .20), and the lower-gain group (M = .20, SD = .13).

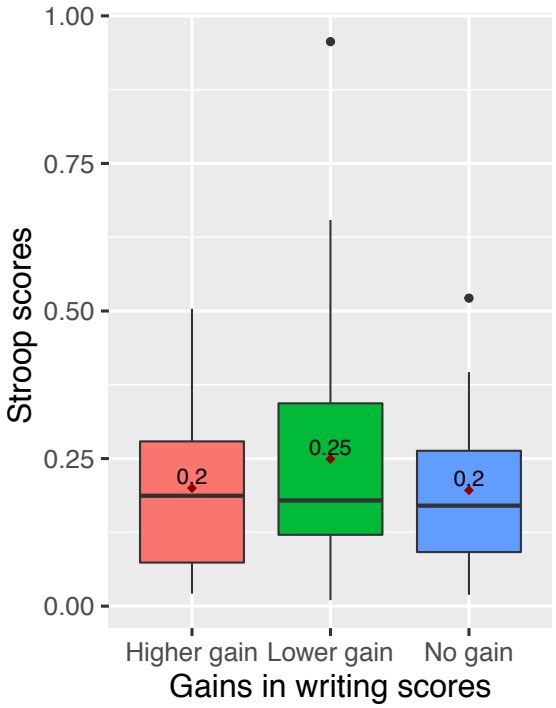


Figure 4.8 Stroop Test Scores of Three Groups with Different Writing Score Gains
Note. A horizontal line within each box indicates each group's median.; A dot within each box indicates each group's mean.

The lack of the role of attention in English writing in multilingual writers can be explained in at least three reasons. First, in terms of processing demands, attention capacity measured by the Stroop test may not be related to writing in general. The Stroop test demands attention to a limited amount of information (i.e., several font colors) for a shorter period at the

lexical level, whereas writing requires attention to higher cognitive processes (e.g., generating and organizing ideas, retrieving a range of words, and reviewing) for a longer period at the discourse level. In addition, many of these higher cognitive processes are writing-specific skills, rather than general cognitive skills. This mismatch in cognitive processing demands between the Stroop test and writing tests may lead to no link between the two.

Second, the Stroop effect may be relevant more closely to writing processes of young learners who increasingly develop resources for attention (Roy et al., 2018) and begin to learn to write narrative and expository writing. These narrative writing and expository writing mainly involve knowledge telling and memory retrieving (Bereiter & Scardamalia, 1987), in which staying focused on writing itself is crucial. Thus, for children, the ability to focus in a shorter span as measured by the Stroop test may link to the ability to write. In contrast, for writing processes of adults who may have already been proficient in performing the Stroop test (MacLeod, 1991) and produce persuasive writing, other abilities, such as critical thinking, argumentation, and evidence providing, may also be crucial beyond the ability to stay focused. Thus, attention as measured by the Stroop test may not be the most useful to measure attentional capacity in the context of persuasive writing in adults.

Lastly, as the Stroop test was measured in the English language, it may not be the most accurate description of attention capacity in multilingual learners. There might have been different relationships between the Stroop effect and writing scores if the Stoop test had been administrated in the participants' more dominant language (e.g., Chinese for Chinese international students).

In sum, the hypothesis of the role of attention as measured by the Stroop test was not supported in the English adult writing context. This does not mean that there is no role of

attention in English writing. More research on the relationship between attention and English writing with different approaches (e.g., using different attention measures behaviors) would merit consideration.

4.2.4 Role of working memory in English writing

Working memory predicted English writing score changes in the general knowledge model and the reading model. On the other hand, working memory did not predict initial English writing scores. However, it should be noted that working memory measured at Time 1 was significantly correlated with writing scores measured at Time 2 ($r = .24$, $p < .05$). This relationship between working memory and writing scores at Time 2 may partly relate to the finding that writing scores gains were related to greater working memory capacity. That is, writing scores at Time 2 consisted of score gains plus writing scores at Time 1, and these gains were related to working memory capacity. These findings indicate that working memory capacity measured a running span task was related to writing scores longitudinally, but not necessarily cross-sectionally.

The finding that no cross-sectional relationship was found between working memory capacity and English writing scores corroborates previous research (Kormos & Sáfár, 2008; Lu, 2010). The lack of such relationship holds across different measures of working memory, including both simple tasks (e.g., a digit span task in Kormos and Sáfár [2008]; and a verbal span task in this dissertation) and complex tasks (e.g., an operational span task in Lu [2010]). Thus, working memory capacity itself may not be the most useful measure to predicting English writing scores.

Potentially, the role of working memory in English writing on the part of multilingual writers may not be linked to raters' evaluation of their writing. Theoretically, working memory

that stores and processes information while performing a task is important in L2 processing (e.g., Linck et al., 2014) as well as in writing (e.g., Hayes & Berninger, 2014). However, this working memory capacity is linked to ongoing writing processes, such as retrieving appropriate language and using knowledge stored in long-term memory, which do not always result in better language use or more persuasive opinion. That is, individuals with higher working memory capacity may hold a large amount of information in working memory during writing, but this does not mean that they can also produce a higher-rated essay. Instead, working memory capacity may be conducive to producing higher-quality writing indirectly when writers also have greater knowledge in long-term memory. In short, although there might be the indirect relationship between working memory as measured by a verbal running span test and English writing scores (e.g., producing better ideas in working memory via using richer prior knowledge), the direct link between two seems flimsy.

Although working memory capacity was not related to English writing scores cross-sectionally at Time 1, it was related to English writing score changes, such that writers with better working memory capacity tended to have greater gains in English writing scores. To illustrate, the three groups with different levels of writing score gains were compared (see Figure 4.9). A pattern was found, such that the higher-gain group tended to receive higher working memory scores at Time 1 (M = 45.65, SD = 11.10) than the lower-gain group (M = 36.53, SD = 9.80). The no-gain group is not of interest because the group tended to have no gains not because of working memory but because of initial higher writing scores (i.e., not enough room for gains).

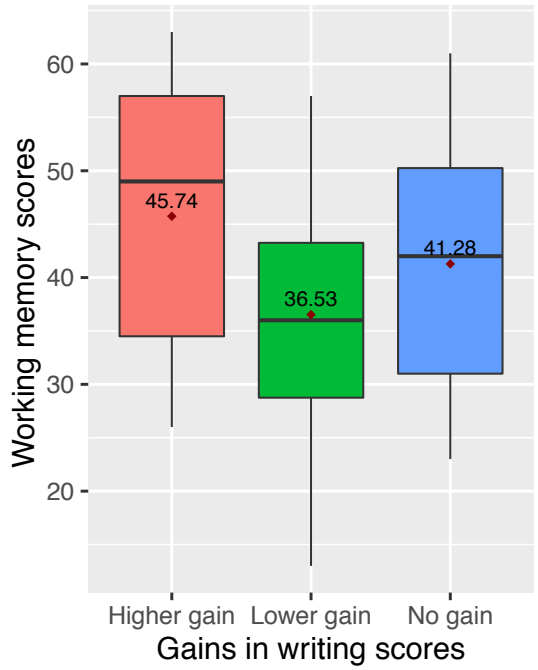


Figure 4.9 Working Memory Scores of Three Groups with Different Writing Score Gains
Note. A horizontal line within each box indicates each group's median.; A dot within each box indicates each group's mean.

The finding that working memory capacity predicted writing score gains can be explained by the notion that as students with higher working memory capacity tend to perform better on learning vocabulary (Williams & Lovatt, 2003), they might also perform better on learning writing. Taking courses in higher education generally involves deliberate and extensive practice of writing (i.e., effortful practice with a goal to improve, Kellogg & Whiteford, 2009), including engaging in English writing practice assigned by an instructor. In this process, students with higher working memory capacity may learn to write an essay in the L2 more quickly than those with lower working memory capacity. Indeed, learning to write a persuasive essay is not a simple task, as producing a successful persuasive essay involves various skills in memory and language, such as the effective management of cognitive demands made on working memory during the writing task (Kellogg & Whiteford, 2009), the rapid retrieval of knowledge related to the writing topic from long-term memory (Kellogg, 2001), the verbal ability to express the

content (McCutchen, 1984), and the automatic search and use of appropriate language forms (Milton et al., 2010). Thus, learning of producing a persuasive essay likely places high demands on working memory, and students with greater working memory capacity resources may learn to manage such high learning demands more rapidly, which may result in greater gains in writing scores over time.

Additionally, it is worth mentioning that working memory capacity was significantly correlated with general knowledge scores ($r = .37$, $p < .01$), English reading scores ($r = .31$, $p < .01$), and English vocabulary scores ($r = .25$, $p < .05$). This result generally supports the importance of working memory capacity in cognitive processing (Hambrick & Engle, 2002) and English-related processing (Jeon & Yamashita, 2014; Linck et al., 2014; Robinson, 2003; Service & Kohonen, 1995; Williams, 2011; Williams & Lovatt, 2003).

In sum, although working memory capacity as measures by a running span task did not predict English writing scores cross-sectionally, it predicted writing score gains longitudinally. This may indicate that working memory capacity is important in learning-to-write processes (e.g., how to generate ideas and how to revise). That is, given that writing processes place high demands on working memory, learning-to-write processes also likely be facilitated by higher working memory capacity.

4.2.5 Role of English vocabulary knowledge in English writing

Initial English vocabulary knowledge was predictive of initial English writing scores, but not changes in English writing scores. The finding that English vocabulary knowledge predicted English writing scores corroborates past research that has reported rich vocabulary knowledge is an important element in successful English writing (Lu, 2010; Milton et al., 2010; Stæhr, 2008; Schoonen et al., 2003, 2011). Higher levels of English vocabulary knowledge likely help writers

express the ideational content more fluently and accurately, which may be linked to higher English writing scores.

However, vocabulary knowledge was not related to writing score changes. This indicates that whether English writing scores increased or not was not associated with the initial level of English vocabulary knowledge. When comparing the high-, low-, and no-grain groups (see Figure 4.10), although all groups showed similar mean initial vocabulary scores, the higher-gain group (M = 560.35, SD = 24.07) tended to receive higher vocabulary scores at Time 1 than the lower-gain group (M = 554.06, SD = 30.31) and the no-gain group (M = 551.22, SD = 28.52), which may indicate that greater vocabulary knowledge may link to higher writing score gains. However, this is a weak presumption, and additional research would need to examine the relationship between vocabulary knowledge and writing score changes.

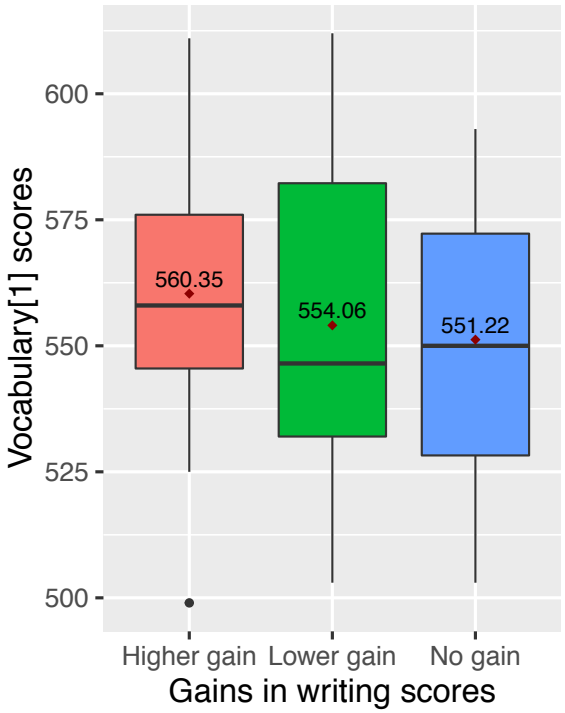


Figure 4.10 English Vocabulary Scores of Three Groups with Different Writing Score Gains
Note. A horizontal line within each box indicates each group's median.; A dot within each box indicates each group's mean.

4.2.6 Role of general knowledge in English writing

Initial general knowledge predicted both initial English writing scores and English writing score changes. Higher-levels of general knowledge were related to higher English writing scores. This may be because greater general knowledge facilitates access to topic-relevant ideas that can enhance planning and idea-generation processes. Thus, multilingual writers with greater general knowledge (that is not directly related to a specific writing topic) tend to produce better essays in the English language. This finding is in line with previous research that has found that L1 writers with greater general knowledge tend to produce higher-quality essays (Dansac & Alamargot, 1999; Hayes & Berninger, 2014), and that L2 writers with greater knowledge on a specific writing topic tend to produce higher quality L2 essays (He & Shi, 2012; Lee & Anderson, 2007).

In addition to initial English writing scores, English writing score changes were also predicted by initial general knowledge, but the relationship between writing score changes and initial general knowledge was negative. This means that multilingual learners with less general knowledge tended to achieve greater gains in writing scores. When comparing the high-, low-, and no-grain groups of different levels of writing score changes (see Figure 4.11), the no-gain group tended to receive higher general knowledge scores (M = 17.94, SD = 4.81) than the higher-gain group (M = 16.39, SD = 4.14) and the lower-gain group (M = 16.17, SD = 4.27). This indicates that the no-gain group with higher initial writing scores also tended to have greater initial general knowledge, while both of the higher-gain and lower-gain groups with lower initial writing scores also tended to less initial general knowledge. That is, multilingual students who have lower levels of both initial English writing scores and initial general knowledge may have a greater potential for gains in English writing scores.

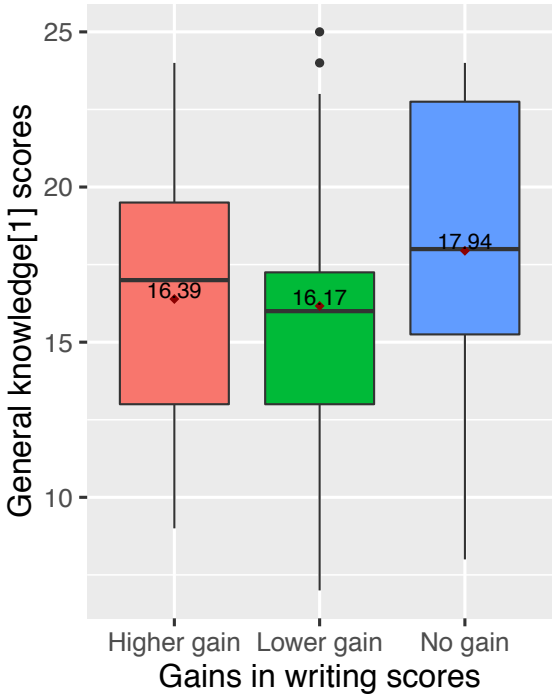


Figure 4.11 General Knowledge Scores of Three Groups with Different Writing Score Gains
Note. A horizontal line within each box indicates each group's median.; A dot within each box indicates each group's mean.

4.2.7 Role of English reading skills in English writing

Initial English reading skills predicted both initial English writing scores and English writing score changes in a manner similar to initial general knowledge predicting the two. Specifically, higher English reading scores predicted higher English writing scores, indicating that skilled English readers also tended to be better English writers (Belcher & Hirvela, 2001; Carson et al, 1990). This is likely because as skilled English readers have the ability to read and understand other authors' texts, they are also likely to have the ability to reread and reflect their own writing during planning and revising processes, which may in turn help construct coherent text (Kaufer et al., 1986; Hayes & Berninger, 2014). For example, if the text written so far is not logical in conveying ideas, skilled English readers are more likely to notice and solve the rhetorical problem.

In addition, initial English reading scores negatively predicted English writing score changes, which indicates that lower levels of initial English reading skills were related to greater gains in English writing scores. When comparing the high-, low-, and no-grain groups of different levels of writing score changes (see Figure 4.12), a pattern was revealed, such that on average, the no-gain group with higher initial English writing scores received the highest initial English reading scores (M = 569.89, SD = 32.43), while the higher-gain group with lower initial English writing scores received the lowest initial English reading scores (M = 558.65, SD = 24.90). These results indicate that greater gains in English writing scores may be a property more of less skilled English writers and readers than of skilled English writers and readers.

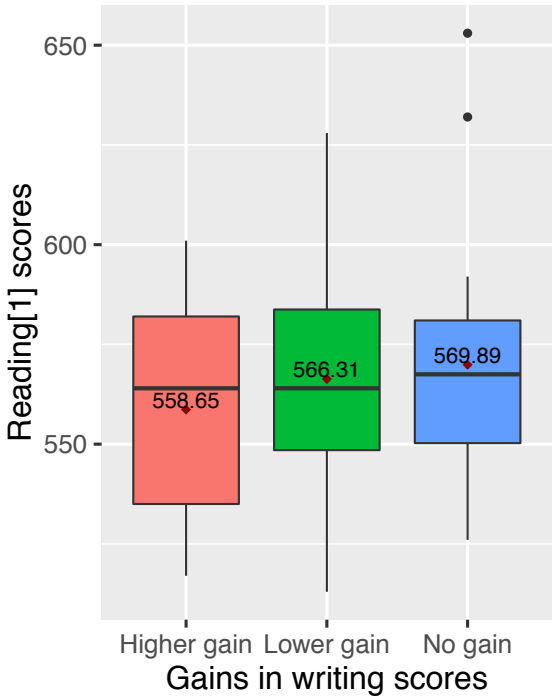


Figure 4.12 Reading Scores of Three Groups with Different Writing Score Gains
Note. A horizontal line within each box indicates each group's median.; A dot within each box indicates each group's mean.

4.2.8 Role of years of English immersion instruction in English writing

No role of years of English immersion instruction was found in predicting either initial English writing scores or English writing score changes. To visually illustrate, two groups were created: one group with multilingual students who had been educated for six years or more in English immersion instruction (henceforth, 'six-years-or-more group'; $n = 41$), and another group with multilingual students who had been educated for five years or less in English immersion instruction (henceforth, 'five-years-or-less group'; $n = 36$). The six-years-or-more group consisted of 13 international students and 28 non-international students (i.e., citizens or residents), while the five-years-or-less group consisted of 33 international students and three non-international students. The average years of English immersion instruction for the six-years-or-more group was 13.68 ($SD = 4.16$), while that for the five-years-or-less group was 2.61 ($SD = 1.38$). Thus, there was a substantial gap in years of English immersion instruction between the two groups.

Figure 4.13 shows histograms of initial English writing scores for each group along with each group's mean score shown in a red line. Despite the substantial differences in English immersion instruction years between the two groups, initial English writing scores ranged from one to 5.5 in both groups. The score distributions of the two groups also look similar. In addition, the six-years-or-more group's mean writing score ($M = 3.00$, $SD = 1.21$) was similar to that of the five-years-or-less group ($M = 2.82$, $SD = 1.10$), though the former's mean score was slightly higher than that of the latter. Similar results were found for L2 writing score changes as well (see Figure 4.14). English writing score changes spread from –1 to 4.5 in both groups with similar distributions of scores. Also, the six-years-or-more group's mean score change ($M = 1.18$, $SD =$

1.11) was similar to that of the five-years-or-less group (M = 1.07, SD = 1.29), though the former's score was slightly higher than that of the latter.

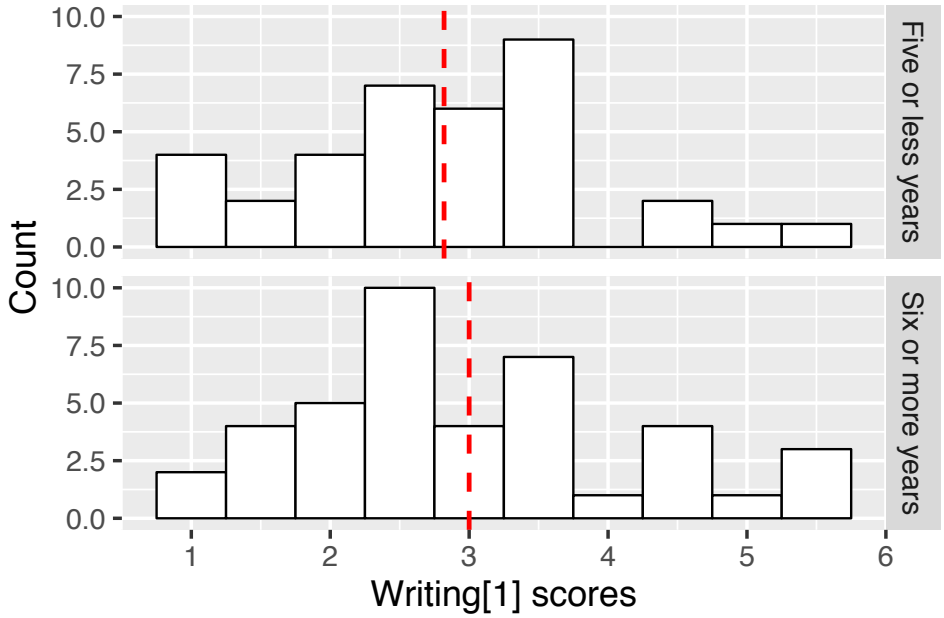


Figure 4.13 Histograms of Initial L2 Writing Scores for Two Groups with Different Years of English Immersion Instruction
Note. [1] = Writing at Time 1; Vertical red lines indicate mean scores.

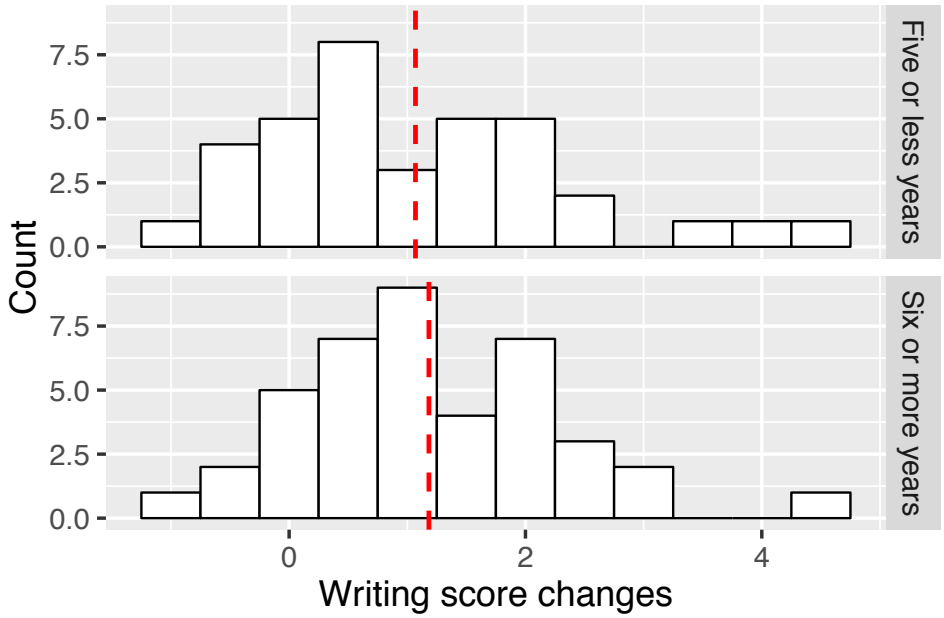


Figure 4.14 Histograms of Writing Score Changes for Two Groups with Different Years of English Immersion Instruction
Note. [1] = Writing at Time 1; Vertical red lines indicate mean scores.

This lack of clear distinctions both in initial English writing scores and English writing score changes between those who had been educated via the English language for six years or more and those who had been educated via the English language for five years or less may indicate that regardless of English immersion experience (mainly through formal schooling), there are individual variations in English writing ability. That is, being immersed and educated in English-speaking immersion contexts cannot be equated with having greater English writing ability. This finding supports Hulstijn's (2015) distinction between Basic Language Cognition (BLC; language related to simple every-day matters) versus Higher Language Cognition (HLC; language related to topics addressed in schools and work places), such that as L1 speakers are not always competent in HLC domains, English immersion instruction does not always lead to competence in HLC domains (e.g., writing a persuasive essay). Thus, it appears that the ability to produce a persuasive essay in English as part of HLC domains differs not as a function of English immersion instruction years but may differ as a function of other factors, such as English writing instruction (Silva & Brice, 2004) and deliberate practice (Kellogg, 2008).

4.2.9 Role of intervals in English writing score change

Intervals in days between the writing test at Time 1 and the writing test at Time 2 were added as an additional, experiment-based predictor of English writing score changes. Results indicated that intervals did not have an effect on score changes. When comparing the high-, low-, and no-grain groups of different levels of writing score gains (see Figure 4.15), although the average intervals of the higher-gain group (M = 226.65, SD = 68.59) were longer than those of the lower-gain (M = 208.03, SD = 42.40) group and the no-gain group (M = 202.67, SD = 34.41), these differences in intervals did not predict writing score changes. Also, the median intervals were similar with a value of approximately 200 days (6.5 months) across the three groups:

higher-gain (205), lower-gain (201), and no-gain (199). This finding suggests that a range of intervals from around four months to one year did not make a difference in writing scores changes.

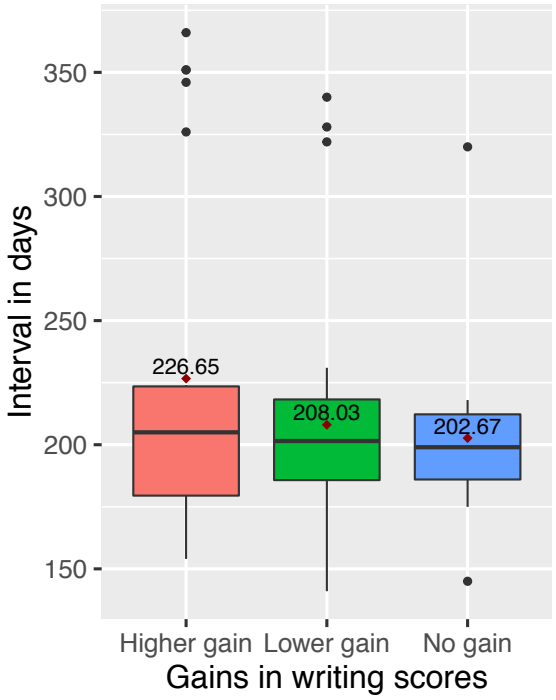


Figure 4.15 Intervals of Three Groups with Different Writing Score Gains
Note. A horizontal line within each box indicates each group's median.; A dot within each box indicates each group's mean.

4.3 Overall Discussion for Research Question 1

The first research question examined the relationship among general cognitive and language resources (i.e., attention, working memory, English vocabulary knowledge, general knowledge, and English reading skills), years of English immersion instruction, and English writing scores over time. Four main overarching findings are discussed below.

First, higher initial English writing scores were predicted by higher levels of English vocabulary knowledge, general knowledge, and English reading skills. This suggests that proficient English writers also tend to have greater English vocabulary knowledge, have greater general knowledge, and be more proficient English readers. In addition, it is worth noting that

English vocabulary knowledge, general knowledge, and English reading skills showed strong correlations with each other ($r > .60$). This strong correlation is interesting because these three are generally considered as part of *crystallized intelligence* (defined as the knowledge learnt through education and experience; Beier & Ackerman, 2005; Cattell, 1943) in L1 speakers in the psychology literature. Thus, the close relationship among the three variables in this study may indicate the presence of the underlying crystallized intelligence in multilingual speakers, which broadly incorporates general knowledge and English language knowledge.[6] In this aspect, being better English writers may not occur in isolation but likely go hand-in-hand with the process of accumulating crystallized intelligence related to general knowledge and English language knowledge through English language education, experience in the English language, and English language practice.

Second, higher English writing score gains were predicted by higher levels of working memory capacity. Becoming a better writer means having better capacity to manage the high degree of cognitive effort related to writing processes using the limited capacity of working memory. Higher working memory capacity likely enable students to quickly learn to be more adept at coordinating planning, sentence generation, and reviewing, which may lead to better writing performance over time (Kellogg, 2008).

Third, higher English writing score gains were related to lower levels of initial English writing scores, initial general knowledge scores, and initial English reading scores. This means that in the context of timed persuasive English writing at the college level, greater writing score gains likely occur to multilingual students who are less skilled in English writing and English

---

[6] Crystallized intelligence is often compared with fluid intelligence (i.e., the processing components of intelligence, such as logical reasoning and math problem solving), which tends to decrease after around age 20. On the other hand, levels of crystallized intelligence (i.e., knowledge) tend to remain stable or increase until at least age 70 (Schaie, 1996).

reading and have less general knowledge at the initial time of measurement. Overall, these findings seem to contradict the existence of a Matthew effect (i.e., cumulative benefit for learning; “rich get richer” scenario; Stanovich, 1986). Rather, a “poor get richer” scenario seems to describe the findings of this study, such that initial lower levels of English writing, English reading, and general knowledge may leave much room and greater potential for growth in better producing persuasive essays by being immersed in English academic contexts in higher education. In addition, a “rich remain the same” scenario also seems to fit the findings of this study in that multilingual students who received initial higher scores in English writing may have already been proficient in producing persuasive essays and remain the same over time.

Lastly, years of English immersion instruction was not related to cross-sectional writing scores or longitudinal writing score gains despite the wide range of years of English immersion instruction among multilingual students (from one year to 21 years). This means that prior extensive experience living in English-speaking countries does not necessarily imply multilingual students’ better ability to produce persuasive essays in English. This finding is in line with previous studies which have reported that Generation 1.5 students who have stayed in the U.S.A. for a longer period do not necessarily produce better English essays than international students who have stayed in the U.S.A. for a shorter period (di Gennaro, 2013; Doolan, 2017). Rather, the finding suggests that the ability to write persuasive essays is something that needs to be learned by multilingual students regardless of whether students have lived longer in English-speaking contexts or whether they are international students or not.

5 LONGITUDINAL RELATIONSHIP BETWEEN THE TRANSLATOR AND ENGLISH WRITING SCORES

The second research question focused on the relationship between the roles of the translator and English writing scores. The roles of the translator were measured in terms of both process and product features. A process feature of the translator was measured by the mean length of language bursts (i.e., the mean number of characters produced between pauses longer than two seconds; Limpo & Alves, 2017). A product feature of the translator was measured by the use of academic words. To measure the use of academic words, the academic word list (AWL; Coxhead, 2000) was used. The number of academic words included in the AWL which were also found in students' essays was counted, normed by text length and multiplied by 100, so that the number indicates the percentage of academic words in the text. The second research question, thus, examined the relationships among English writing scores, burst length, and the use of academic words over time. Various relationships were examined: covariances among initial levels, covariances among changes, covariances between initial levels and changes of each variable, and cross-lagged covariances between initial levels and changes across different variables. Results are provided in Section. 5.1, followed by discussion provided in Section 5.2.

5.1 Research Question 2 Results

5.1.1 Descriptive statistics

Among the 77 participants, one student's data was excluded from the analysis because the student's change in burst lengths between Time 1 (25.80) and Time 2 (141.31) showed an atypical pattern with the substantial change of 115.51. Visual inspection of the scatterplot of burst lengths between Time 1 and Time 2 also indicated that this student's change in burst length was an outlier. Thus, a total of 76 participants were analyzed for the research question 2. Among

them, two students' burst lengths at Time 1 were not recorded due to technical errors. To handle these missing data, a FIML approach was used. Table 5.1 presents the descriptive statistics of writing scores, burst lengths, and academic word percentage at two time points along with their changes across the two time points. Score changes were calculated by subtraction (i.e., scores at Time 2 minus scores at Time 1). The scatter plots and histograms are presented in Figure 5.1.

Table 5.1 Descriptive Statistics for Variables Related to Research Question 2

Variable	*N*	Mean	*SD*	Min.	Max.	Skewness	Kurtosis
Writing score at Time 1	76	2.94	1.16	1	5.50	.37	–.33
Writing score at Time 2	76	4.03	1.00	1.50	6	.10	–.38
Change in writing score	76	1.09	1.18	–1	4.50	.72	.49
Burst length at Time 1	74	53.26	35.83	16.04	200.50	1.75	3.34
Burst length at Time 2	76	49.75	25.91	14.45	158.64	1.35	2.93
Change in burst length	74	–3.24	24.49	–89.11	44.45	–1.09	2.30
Academic word percentage at Time 1	76	6.90	3.48	1.33	18.77	.70	.36
Academic word percentage at Time 2	76	6.15	2.55	1.16	15.23	.71	.88
Change in academic words	76	–.75	3.97	–14.68	7.65	–.63	1.01

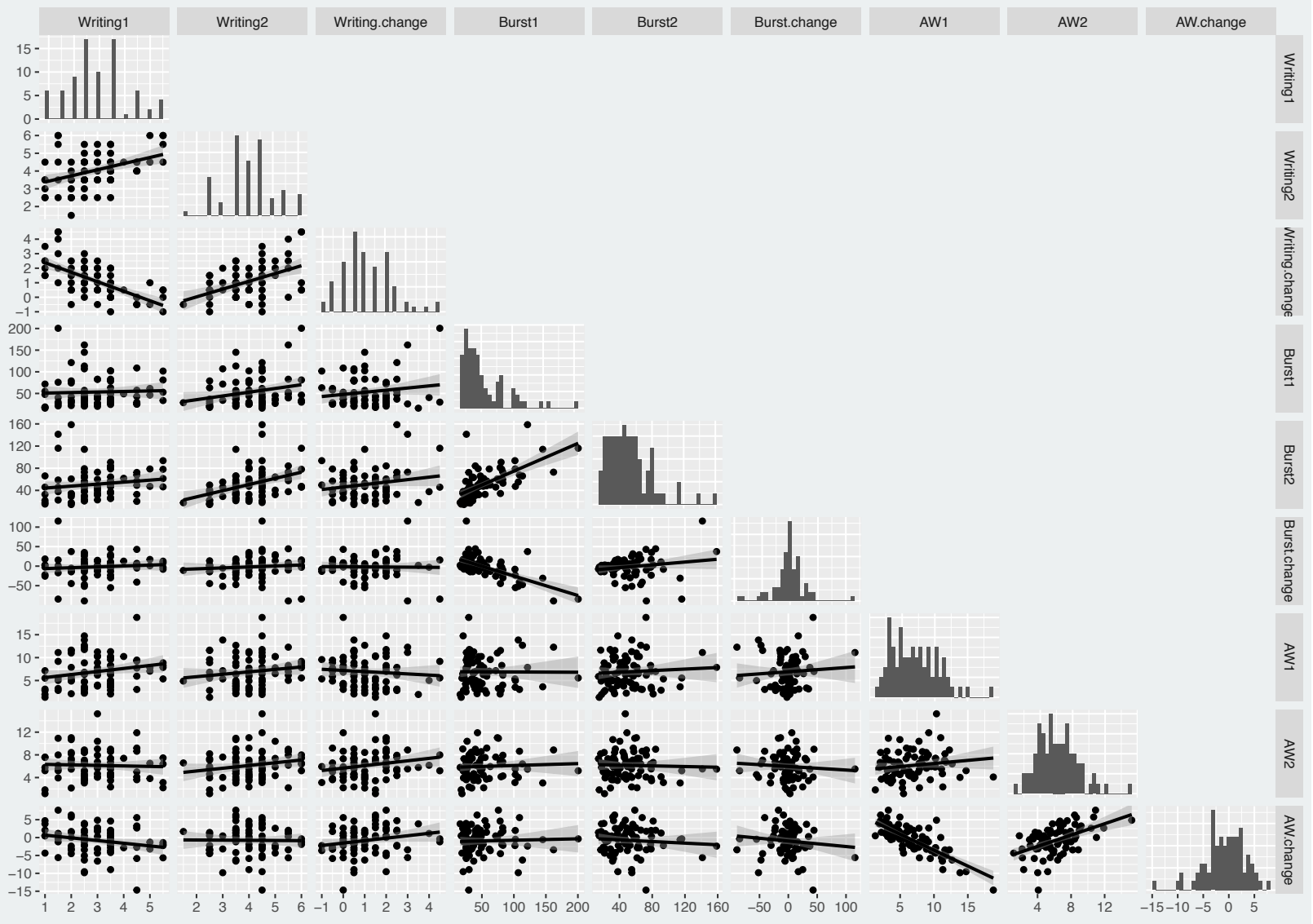


Figure 5.1 Scatter Plots and Histograms of Variables Related to Research Question 2

Note. 1 = Time 1, 2 = Time 2, Burst = mean burst length; AW = percentage of academic percentage.; Diagonal graphs show histograms of each variable. On scatter plots, lines indicate linear predictions from a linear regression model of the two variables.

Overall, average writing scores increased from 2.94 to 4.03 over time with a mean change of 1.09. Average lengths of burst decreased from 53.26 to 49.75 with a mean change of –3.24. The average percentage of academic words decreased from 6.90 to 6.15 with a mean change of –.75. Figure 5.2 graphically shows changes in writing scores, changes in burst length, and changes in the number of academic words over time. For burst length changes and academic word changes, there was a range of variability without clear patterns of changes.

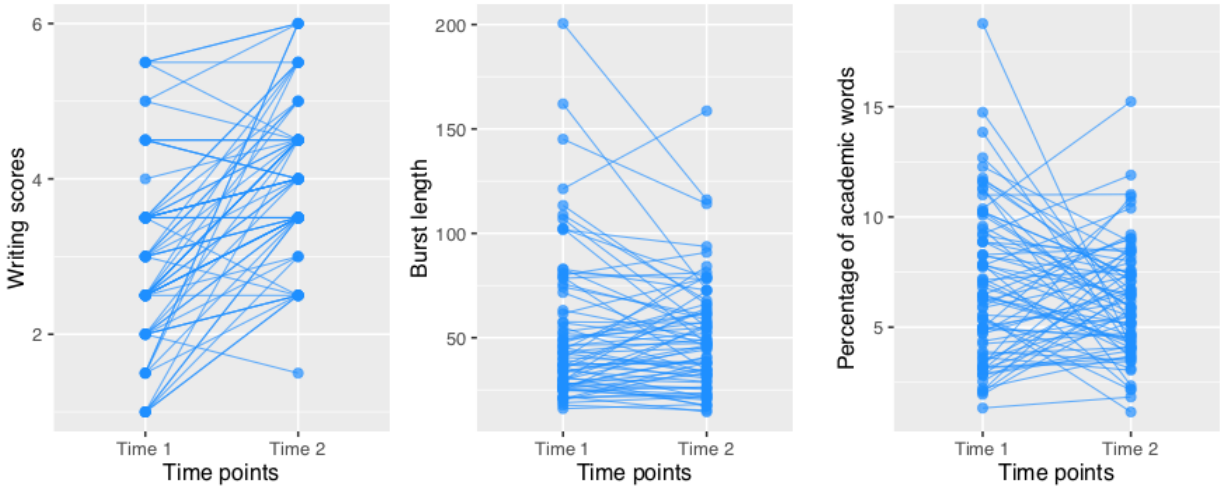


Figure 5.2 Writing Score Changes (Left; N = 76), Burst Length Changes (Center; N = 74), and Academic Word Percentage Changes (Right; N = 76).

5.1.2 Correlation analysis

Correlations among variables related to the research question 2 are shown in Table 5.2. Correlations were significant between writing scores at Times 1 and 2 ($r = .41$, $p < .01$) and betwteen burst lengths at Times 1 and 2 ($r = .73$, $p < .01$), but not between academic word counts at Times 1 and 2 ($r = .16$, $p > .05$). Correlations between intial levels and changes were all negative for writing scores ($r = –.63$, $p < .01$), burst lengths ($r = –.68$, $p < .01$), and academic word counts ($r = –.77$, $p < .01$). On the other hand, positive correlations were found between writing score changes and writing scores at Time 2 ($r = .45$, $p < .01$) and between changes in

academic word counts and academic word counts at Time 2 (r = .50, p < .01), but not between burst length changes and burst length at Time 2 (r = .00, p > .05).

Table 5.2 Correlations among Variables Related to Research Question 2

Variable	1	2	3	4	5	6	7	8	9
1 Writing at Time 1	1								
2 Writing at Time 2	.41**	1							
3 Change in writing score	–.63**	.45**	1						
4 Burst length at Time 1	.03	.25*	.18	1					
5 Burst length at Time 2	.22	.41**	.13	.73**	1				
6 Change in burst length	.19	.08	–.12	–.68**	.00	1			
7 Academic word percentage at Time 1	.25*	.15	–.11	.02	.02	–.00	1		
8 Academic word percentage at Time 2	–.05	.19	.21	.05	–.02	–.07	.16	1	
9 Change in academic word percentage	–.24*	–.01	.23*	.01	–.03	–.04	–.77**	.50**	1

Note. *N*s for burst length at Time 1 and changes in burst length were 74, while *N*s for the other variables were 76.; ** indicates p < .01, and * indicates p < .05.

5.1.3 Results of latent change score modeling

Results of testing the latent change score model as shown in Figure 3.4 are presented in Table 5.3 and Figure 5.3. Change statistics (i.e., intercepts [means] and variances of initial scores and changes) are shown in Table 5.3 but not in Figure 5.3 for clarity of presentation. Significant changes were found in writing scores, but not in mean burst lengths or academic word percentages (see the 'intercept' section in Table 5.3). This indicates that at the group level, while students tended to have gains in writing scores, they tended to produce similar burst lengths during writing and use similar percentages of academic words in their essays across the two time points. On the other hand, significant individual variability was found for writing score changes, burst length changes, and academic word percentage changes (see the 'variance' section in Table 5.3), which suggests that individual students differed in their changes in writing scores, burst lengths, and academic word percentages.

Table 5.3 Results of Latent Change Score Model for Research Question 2

Intercept (mean)	Estimate	*SE*	*z*	*p*	Standardized Estimate
Writing[1]	2.94	.13	22.35	< .01	2.56
ΔWriting	1.09	.13	8.14	< .01	.93
Burst length[1]	53.16	4.07	13.06	< .01	1.50
ΔBurst length	–3.42	2.82	–1.21	.23	–.14
Academic word[1]	6.90	.00	17.41	< .01	2.00
ΔAcademic word	–.75	.01	–1.66	.10	–1.19
Variance	**Estimate**	***SE***	***z***	***p***	**Standardized Estimate**
Writing[1]	1.32	.20	6.61	< .01	1.00
ΔWriting	1.37	.25	5.42	< .01	1.00
Burst length[1]	1249.92	340.47	3.67	< .01	1.00
ΔBurst length	591.88	144.67	4.09	< .01	1.00
Academic word[1]	11.94	.00	5.57	< .01	1.00
ΔAcademic word	15.57	.00	4.93	< .01	1.00
Covariance	**Estimate**	***SE***	***z***	***p***	**Standardized Estimate**
ΔWriting ↔ Writing[1]	–.84	.16	–5.29	< .01	–.63
ΔBurst length ↔ Burst length[1]	–589.78	211.89	–2.78	< .05	–.69
ΔAcademic word ↔ Academic word[1]	–10.54	2.59	–4.07	< .01	–.77
Writing[1] ↔ Burst length[1]	1.26	4.54	.28	.78	.03
Writing[1] ↔ Academic word[1]	.97	.42	2.31	< .05	.25
Burst length[1] ~~ Academic word[1]	.02	.15	.16	.88	.02
ΔWriting ↔ ΔBurst length	–3.42	4.71	–.73	.47	–.12
ΔWriting ↔ ΔAcademic word	1.07	.43	2.48	< .05	.23
ΔBurst length ↔ ΔAcademic word	–4.66	13.60	–.34	.73	–.05
ΔWriting ↔ Burst length[1]	7.43	7.83	.95	.34	.18
ΔWriting ↔ Academic word[1]	–.45	.41	–1.10	.27	–.11
ΔBurst length ↔ Writing[1]	5.27	2.68	1.97	< .05	.19
ΔBurst length ↔ Academic word[1]	–.48	12.82	–.04	.97	–.01
ΔAcademic word ↔ Writing[1]	–1.11	.51	–2.17	< .05	–.24
ΔAcademic word ↔ Burst length[1]	1.21	14.08	.09	.93	.01

Note. Δ = "change in", [1] = Time 1, ↔ = "covariance with"

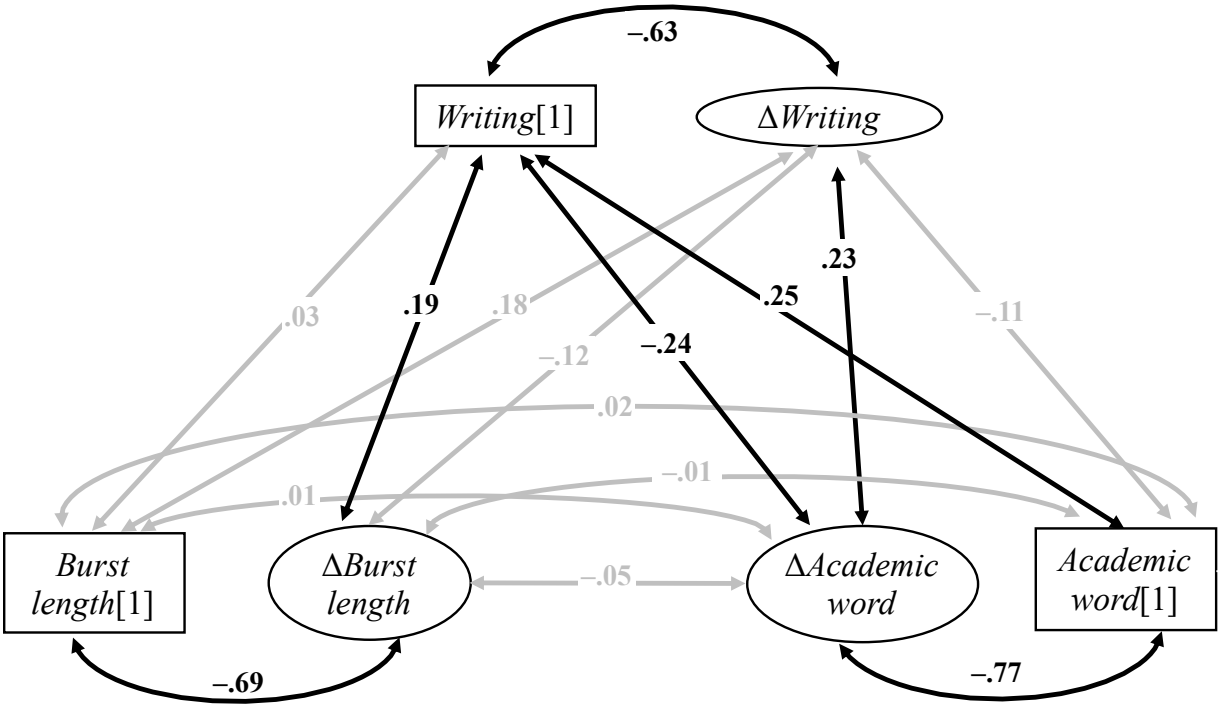


Figure 5.3 Latent Change Score Model for Research Question 2
Note. Δ = "change in", [1] = Time 1. Estimates are standardized coefficients. Significant paths are shown in black color, while nonsignificant coefficients are shown in grey color.

To answer the research question 2 (i.e., longitudinal relationships among English writing scores, burst length, and the use of academic words), four types of covariances are examined (see Figure 5.3 and the 'covariance' section in Table 5.3). These four types were (a) covariances between initial levels and changes of each variable, (b) covariances among initial levels of the three variables, (c) covariances among changes in the three variables, and (d) cross-lagged covariances between initial levels and changes across different variables. These model-based covariances are shown with correlation coefficients (i.e., standardized covariance estimates) in Table 5.4. Results of each covariance type along with standardized estimates are reported below.

Table 5.4 Correlations Based on the Latent Change Score Model for Research question 2

Variable	1	2	3	4	5	6
1 Writing[1]	1					
2 Burst length[1]	.03	1				
3 Academic word[1]	.25*	.02	1			
4 ΔWriting	–.63**	.18	–.11	1		
5 ΔBurst length	.19*	–.69*	–.01	–.12	1	
6 ΔAcademic word	–.24*	.01	–.77**	.23*	–.05	1

Note. [1] = Time 1, Δ = "change in"; ** indicates $p < .01$, and * indicates $p < .05$.

Covariances between initial levels and changes of each variable. Covariances between intial levels and changes were significantly negative for writing scores ($r = –.63$, $p < .01$), burst lengths ($r = –.69$, $p < .05$), and academic word percentages ($r = –.77$, $p < .01$). This indicates that L2 students who received lower writing scores at Time 1, produced shorter burst lengths during writing at Time 1, and used fewer academic words in their essays at Time 1 tended to have greater gains in the respective variables over time.

Covariances among initial levels of the three variables. When covariances among initial levels of the three variables (i.e., writing scores, burst lengths, and academic word percentages) were examined, one significant covariance was found, such that writing scores at Time 1 were positively associated with academic word percentages at Time 1 ($r = .25$, $p < .05$), indicating that higher-rated essays tended to include more academic words than lower-rated essays. The other two covariances were not significant. The covariance between writing scores at Time 1 and burst lengths at Time 1 was nonsignificant ($r = .03$, $p > .05$). The covariance between burst lengtsh at Time 1 and academic word percentages was also nonsigificant ($r = .02$, $p > .05$).

Covariances among changes in the three variables. When covariances among changes in the three variables (i.e., writing scores, burst lengths, and academic word percentages) were examined, one significant covariance was found, such that changes in writing scores were positively related to changes in academic word percentages ($r = .23$, $p < .05$). This indicates that students' greater gains in writing scores are related to their greater use of academic words in essays. On the other hand, the covariance between changes in writing scores and changes in burst length was not significant ($r = –.12$, $p > .05$) In addition, the covariance between changes in burst length and changes in academic word percentages was not significnat ($r = –.05$, $p > .05$).

Cross-lagged covariances between initial levels and changes across different variables. Six cross-lagged covariances between initial levels and changes across the different variables were examined: covariances of initial writing scores with burst length changes and academic word percentage changes, covariances of initial burst lengths with writing score changes and academic word percentage changes, and covariances of initial academic word percentages with writing score changes and burst length changes.

Among the six, two significant cross-lagged covariances were revealed, such that writing scores at Time 1 were positively associated with changes in burst lengths ($r = .19$, $p < .05$), while being negatively associated with changes in academic word percentages ($r = –.24.p < .05$).

The other four cross-lagged covariances were nonsignificant. Initial burst lengths were not associated with either writing score changes ($r = .18$, $p > .05$) or acadmic word percentage changes ($r = .01$, $p > .05$). Also, initial academic word percentages were not related to either writing score changes ($r = –.11$, $p > .05$) or burst length changes ($r = –.01$, $p > .05$).

Lastly, the latent change score model presented in Figure 5.3 was identified without a degree of freedom, and thus its fit measures could not be calculated. To examine whether the

model had a good fit, a model that excluded nonsignificant covariances (n = 8) was tested. The results of the model without the nonsignificant covariances were almost the same (with minor differences in estimates), and the model fit adequately (χ^2 = 6.94, df = 8, p = .54, CFI = 1, SRMR = .064).

5.2 Research Question 2 Discussion

5.2.1 Summary of results

The research question 2 examined cross-sectional and longitudinal relationships among English writing scores, mean language burst lengths during writing, and the use of academic words as found in essays. With respect to cross-sectional relationships, higher English writing scores were related to greater percentages of academic words in essays. In terms of longitudinal relationships, initial levels of writing scores, burst lengths, and academic word percentages were negatively associated with changes in the respective variables. Cross-lagged longitudinal relationships were also found, such that initial English writing scores were positively linked to burst length changes, but negatively linked to changes in academic word percentages in essays. In addition, gains in writing scores were linked to gains in academic word percentages. Lastly, burst length did not show any relationship with academic word percentages. The summary of these results is presented in Table 5.5.

Table 5.5 Summary of the Relationship among Writing Scores, Burst Length, and Academic Word Percentages

Relationship at Time 1		Relationship across time	
Covariance	Significance	Covariance	Significance
Writing[1] ↔ AW[1]	Yes (+)	Writing[1] ↔ ΔWriting	Yes (–)
Writing[1] ↔ Burst length[1]	No	Burst length[1] ↔ ΔBurst length	Yes (–)
AW[1] ↔ Burst length [1]	No	AW[1] ↔ ΔAW	Yes (–)
		Writing[1] ↔ ΔBurst length	Yes (+)
		Writing[1] ↔ ΔAW	Yes (–)
		ΔWriting ↔ ΔAW	Yes (+)
		AW[1] ↔ ΔWriting	No
		Burst length[1] ↔ ΔWriting	No
		Burst length[1] ↔ ΔAW	No
		AW[1] ↔ ΔBurst length	No
		ΔWriting ↔ ΔBurst length	No
		ΔBurst length ↔ ΔAW	No

Note. AW = Academic word percentage, [1] = Time 1, Δ = "change in", ↔ = "covariance with". Directions of significant covariances (positive vs. negative) are shown in parentheses.

Below, changes in writing scores, burst lengths, and academic word percentages are first briefly discussed. Cross-sectional and longitudinal relationships among writing scores, burst lengths, and academic word percentages are then discussed.

5.2.2 Change in writing scores, burst lengths, and academic word percentages

Results of the latent change score modeling indicated that changes in scores writing were significantly positive, while changes in burst lengths and changes in academic word percentages were not significant. The lack of group-level gains in multilingual students' use of academic words over time at the college level have also been reported in previous studies (Knoch, Rouhshad, Oon, & Storch, Neomy, 2015; Knoch, Rouhshad, & Storch, 2014). However, significant variances in the changes in writing scores, burst lengths, and academic word percentages were revealed, suggesting interindividual variability in the patterns of changes.

Thus, although burst lengths and academic word percentages did not change over time as a group, individual students within the group showed interindividual variability in changes in all of the three variables.

5.2.3 Cross-sectional relationship among initial levels of writing scores, burst lengths, and academic word percentages

Results indicated higher English writing scores were related to greater use of academic words as found in essays. To illustrate, two students' intact essays written about an impact of images and impressions on people produced at Time 1 are provided in Table 5.6. One student (Student number 85) produced 30 academic words out of a total of 324 words with an academic word percentage of 9.23, and this student's holistic score on the essay was 4.5 out of 6. Indeed, given that around 10% of an academic text tends to consist of the AWL words (Coxhead, 2000), this student's (Student number 85) essay contained an academic word percentage similar to that of a typical academic text. In contrast, another student (Student number 64) produced 18 academic words out of a total of 496 words with an academic word percentage of 2.12, and this student's holistic score on the essay was 2 out of 6. Thus, although other elements, such as organization, topic development, and coherence, may have led to higher writing scores, it seems that the use of academic words also impacted raters' evaluation of essays, such that essays with more sophisticated words (i.e., more use of academic words in persuasive essays) tended to be rated higher than those with less sophisticated words. This finding supports past research that reported the importance of the use of academic words in academic writing (Coxhead 2012; Douglas, 2013; He & Shi, 2012; Laufer, 2013).

Table 5.6 Two Example Essays Produced at Time 1 and Academic Word Use

Essay from Student Number 85	Essay from Student Number 64
Total word counts: 324 Academic word counts: 30 Academic word percentage: 9.23% Counts of academic word appearing the prompt: 4	Total word counts: 496 Academic word counts: 18 Academic word percentage: 2.12% Counts of academic word appearing the prompt: 4
How many time you have heard this **clause**: "Dress to impress?" More than one time, right? As college students,who learning and working in a semi-professional enviroment, we are constanly reminded to dress up **professional** in important events in ordert to land a **potential** internship or a **job** offer. Have you wonder why? **Psychologists** have scientifically proved that human tend to develop **positive attitudes** toward a stranger if they are impressed by the way that person dress. However, in my opinion, ***images*** and impression have too much of an effect on people. In our society, we are taught to not judge someone based on their race, color, ethinicity or any other **obvious** characteristics. We should get to know the person's background, value, and skills before many any judgement. However, in reality, most of hiring manager make the decision whether to hire an employee **solely** based on their first impression on that person, mostly comes from how they dress. We think we can **assume** a stranger's life story and make **prediction** their work behaviors based on the first ***image***. The fundamential **attribution theory** helps us understand why we make these assumpstions, and still, we ignorantly have **negative attitude** or even **discriminated** action against these **individuals**. **Psychologist** also proved that people tend to like people who are **similar** to them in **physically** or **mentally**. Liking is one of the most important **factor** in persuasion **technique**. If someone like you, you have much higher chance to **convince** them to do something they orginially do not intend to do. That is why companies use influential, famous people to advertise for their products. There public figures has **project *images*** that many people can relate to or dream to be, which give them the power to influent. Lastly, I understand why ***images*** and impression have effects on people. However, it is overused by many companies or politicants. As students, we must listen and **analyze** what they said, not only what the	Appearances are used to **promote** products as well as "famous people" become more famous. ***Images*** and impressions have a huge **impact** in people's lives. If my favorite movie star is ***promoting*** facial products and I absolutely love her than I would buy it because my favorite movie star claimed she/he uses the product and that is why they look the way they look. People have become so gullible that when they see something that looks like the person advertising it is having fun using it and is not a waste of time and money than they will also buy it because they think it is worth their time and money. Most of the times the products shown on television do not work they way it was suppose to work. For example my mother bought a curtain that was suppose to keep bugs out of the house in a hot summer day because there was a lot of bugs outside. We used it for a while and it was great at first but after a while you saw bugs still inside the house. So the product she bought did no **affect** on our house because we still saw bugs there and it was just taking up space and time. My mother could have bought somethineg else instead of **purchasing** this **item**. I believe ***images*** and impressions do have a lot of **affect** on people. I learned from some friends in marketing that colors can make people feel things. Like red and yellow in the McDonald's sign makes customers feel "hunger" and that is good advertisement because people need to go to McDonald's to eat their food not to play in the kids **area**. The way commercials are made are out to get customers to go out and buy **items** they do not really need. Chuchu plants for example is an **item** that is a plant and is done to make look like your favorite thing like Scooby-doo or other actors. It was a big deal back in 2005 but it is not anymore. People went crazy for it because it looked like the person or thing they loved. It was a great advertising **scheme**. The axe clone wants their audience which is **primarily** men to go buy their products because they will "get all the women" and it will make them feel like they are "it" the "big deal" and that is every guys dream but it is not true because women do not stick to men like magnets. It is very false advertising but a very great **scheme** to pull off. Every year some new product is out and every year we have people falling for what it is suppose to "represent". People have this thought that if they possibly buy the product they could just be like that person. Its a marking **strategy** that gets their **items** to be sold out to the public; making the ***creators*** of the products to have lots of profit.

Note. Academic words are shown in bold.; Academic words that also appear in the prompt are shown in italics.; In the essay from Student 85, additional five academic words (*environment, constantly, ethnicity, fundamental*, and *assumption*) were misspelled and thus not calculated toward the number of academic words.

On the other hand, mean burst lengths at Time 1 were not related to writing scores at Time 1. This finding seems to contradict Hayes-Berninger model (2014), which assumes that longer burst lenghts generally indicate the translator functions fluently, thus potentially leading to better text. The finding of this study that higher-rated essays were not linked to longer burst lengths may be explained by at least two reasons. First, longer burst lengths may not always indicate a better production of texts. For example, too many longer mean burst legnths may reflect a stream-of-conciousness writing stype without accompanying much contemplation of contents. Second, while increasing burst lengths may be importat at the initial stages of learning to write, such as childrens' learning to write narratives (Alves & Limpo, 2015; Limpo & Alves, 2017), burst lengths may not be a key element in producing a successful persuasive essay at a more advanced level of writing.

Despite the lack of the relationship between burst lengths and writing scores at Time 1, it should be noted that burst length and writing scores were significantly correlated at Time 2 (r =. 41, p < .01). Based on previous studies which reported the importance of burst length in writing (Chenoweth & Hayes, 2001; Hayes & Berninger, 2014; Limpo & Alves, 2017), the close relationship between burst lengths and writing scores at Time 2 seems to be more reasonable. The relationship between burst lengths and writing scores at Time 1 is discussed in more detail in Section 5.2.4.1.

The lack of the relationship between burst length and writing scores at Time 1 can be explained by some students who produced excessively long burst lengths at Time 1 probably as a result of their stream-of-conciousness writing style. Because these students were not necessarily higher writing scorers at Time 1, these students' presence may have weakend the links between writing scores and burst lengths at Time 1. Indeed, given that a notable pattern of decreases in

burst lengths over time was found in multilingual students who tended to produce excessively longer bursts, burst lengths at Time 2 with this decreasing pattern would be a more reasonable representation of burst length, which was thus significantly correlated with writing scores at Time 2

Mean burst lengths at Time 1 were also not related to academic word percentages in essays produced at Time 1. This indicats that longer burst lengths during writing processes is not associated with the use of more sophisticated words (i.e., academic words). Potentially, burst length may be linked to lexical sophiscation when length of each burst is considered with lexical features as found in that burst. However, this is beyond of the scope of this dissertation.

5.2.4 Longitudinal relationship among writing scores, burst lengths, and academic word percentages

Using trivariate latent change score modeling, a total of 12 longitudinal covariances among writing scores, burst lengths, and academic word percentages were examined. Among the 12, six significant longitudinal relationships were revealed, while the remaining six relationships were nonsignificant. Below, significant covariances are first discussed followed by discussion on nonsignificant covariances.

5.2.4.1 Significant covariances among writing scores, burst lengths, and academic word percentages over time

Among the six significant longitudinal covariances, first of all, English writing scores at Time 1 showed a negative relationship with writing score changes, which suggests that greater gains in writing scores were more linked to multilingual students who received lower initial writing scores (for details, also see Section 4.2.2).

Two significant longitudinal covariances were related to burst length changes. Specifically, changes in burst length showed a negative relationship with burst lengths at Time 1 but a positive relationship with writing scores at Time 1. To illustrate, three groups were created and compared: one group consisting of students whose mean burst lengths increased over time (henceforth, 'burst-increase group', $n = 27$), another group consisting of students whose mean burst lengths decreased over time (henceforth, 'burst-decrease group', $n = 27$), and another group consisting of students whose mean burst lengths remain similar over time (henceforth, 'burst-same group', $n = 22$).[7] Figure 5.4 shows two boxplots for the three groups with different levels of burst length changes: one boxplot for the three groups' burst lengths at Time 1 (left) and another boxplot for the three groups' writing scores at Time 1 (right). Interestingly, the burst-decrease group tended to not only produce longer mean burst lengths at Time 1 (with an average of 77.82) but also receive lower writing scores at Time 1 (with an average of 2.57) than the burst-increase group and the burst-same group. One possible scenario that can explain this result is that students who decreased their burst lengths may have produced longer burst lengths as a result of a stream-of-conciousness writing style, which in turn may have resulted in their lower writing scores at Time 1. Also, this burst-decrease group's mean writing scores increased from 2.57 ($SD = 1.00$) at Time 1 to 3.74 ($SD = 1.06$) at Time 2. Thus, a decrease in burst lengths in initial lower writing scorers may reflect a behavioral change from writing as much as possible to producing more meaningful language strings in a more controlled manner.

---

[7] The burst-increase group consisted of students whose burst lengths increased by five characters or more, the burst-decrease group consisted of students whose burst lengths decreased by five characters or more, and the burst-same group consisted of students whose burst lengths changed within a range between 4.99 and 4.99. The cutoff of five characters were chosen because five characters are approximately equivalent to one word. The burst-increase-group's mean burst length change was 18.95 ($SD = 11.71$), ranging from 6.05 to 44.45. The burst-decrease-group's mean burst length change was –26.10 ($SD = 22.97$), ranging from –89.11 to –5.12. The burst-same-group's mean burst length change was –.39 ($SD = 2.80$) ranging from –4.93 to 4.48.

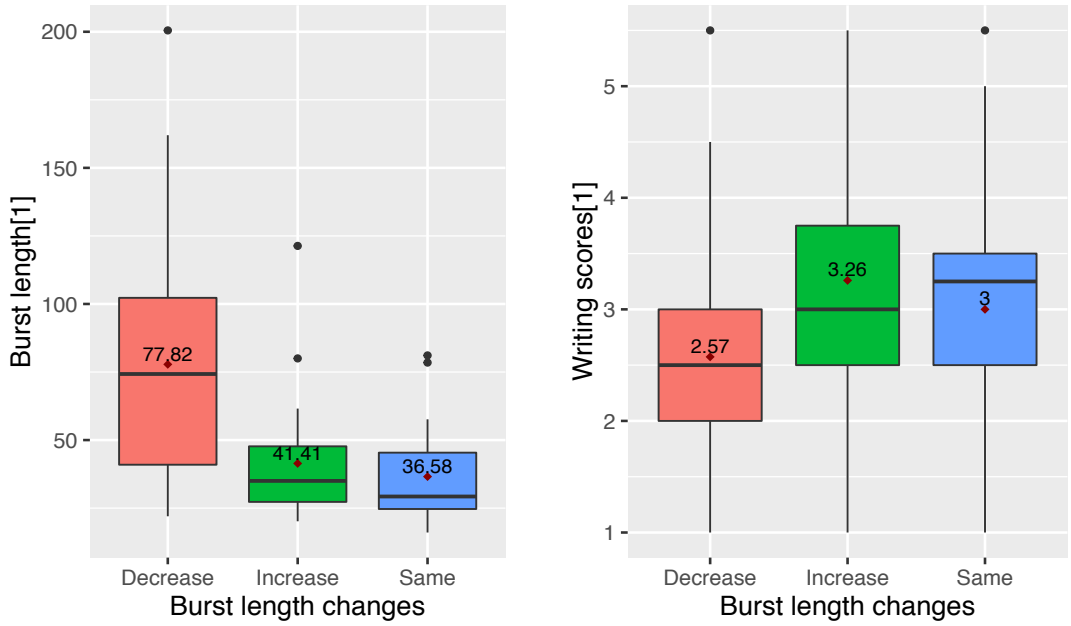


Figure 5.4 Initial burst lengths (Left) and Initial Writing Scores (Right) for Three Groups with Different Burst Length Changes

Note. A horizontal line within each box indicates each group's median.; A dot within each box indicates each group's mean.

The remaining three significant covariances were related to changes in academic word percentages. Specifically, changes in academic word percentages were negatively related to academic word percentages at Time 1 and writing scores at Time 1 but positively related to writing score changes. To illustrate, two groups were created and compared: one group consisting of students whose academic word percentages decreased over time (henceforth, 'academic-word-decrease' group; $n = 41$), and another group consisting of students whose academic word percentages increased over time (henceforth, 'academic-word-increase' group; $n = 35$).[8] Figure 5.5 shows three boxplots for the two groups with different levels of academic word percentage changes: one boxplot for academic word percentages at Time 1 (left), another boxplot for writing scores at Time 1 (center), and another boxplot for writing score changes

[8] The academic-word-decrease-group's average in changes in academic word percentages was –3.53, ranging from –14.68 to –.33, while the academic-word-increase group's one was 2.51 ranging from .03 to 7.65.

(right). Across the three boxplots, interestingly, a similar pattern was revealed, such that students in the academic-word-increase group (shown in aqua boxes) tended to use a smaller percentage of academic words in their essays at Time 1, receive lower writing scores at Time 1, and have greater gains in writing scores than the academic-word-decrease group (shown in orange boxes). This finding suggests that an increase in academic word percentages is more likely to take place for students who tended to receive lower writing scores and make less use of academic words at the initial time of measurement, and these gains in academic word percentages may go hand-in-hand with gains in writing scores.

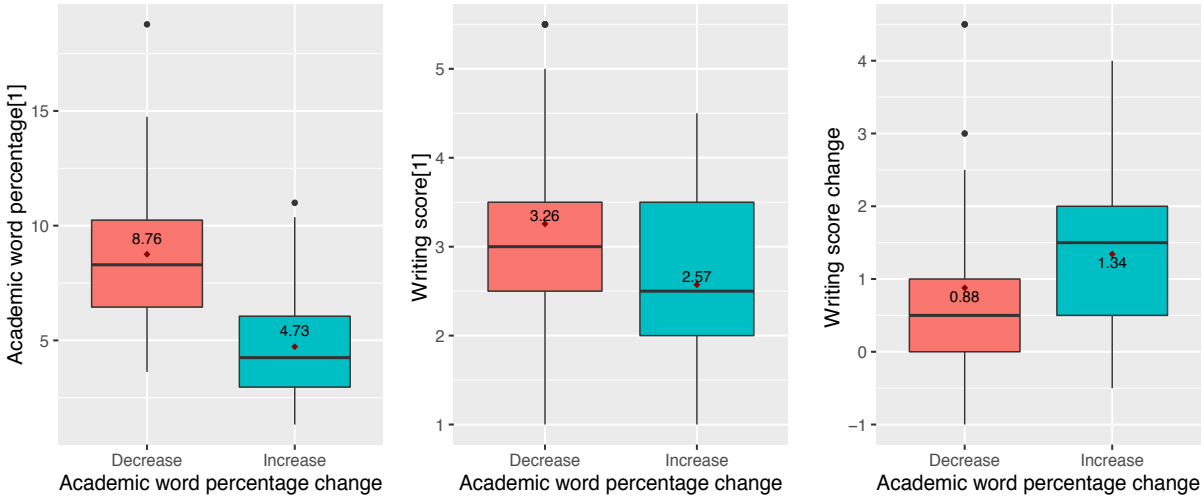


Figure 5.5 Two Groups with Different Academic Word Percentage Changes
Note. [1] = Time 1; A horizontal line within each box indicates each group's median.; A dot within each box indicates each group's mean.

To further illustrate the relationship between gains in academic word percentages and gains in writing scores, one student's two intact essays written at Time 1 and Time 2, respectively, are presented in Table 5.7. This student's academic word percentages increased from 3.37% at Time 1 to 6.76% at Time 2. This student's writing scores also increased from 2.5 at Time 1 to 4 at Time 2. This example shows that greater use of academic words in persuasive essays may positively impact raters' evaluation of the essays.

Table 5.7 Example Essays Produced at Time 1 and Time 2

Essay from Student Number 40 at Time 1	Essay from Student Number 40 at Time 2
Total word counts: 208 Academic word counts: 7 Academic word percentage: 3.37% Counts of academic word appearing the prompt: 1	Total word counts: 340 Academic word counts: 23 Academic word percentage: 6.76% Counts of academic word appearing the prompt: 9
Images and impressions have a lot of effect on people, a lot of people are being mistaken for reality. Many companies uses various **techniques** to attract their customers, and one of them is to take attention of the customers on the way their products are packed. When you go to a supermarket, you will be able to see a **specific** product in many brands and different packaging. What could attract the customer to buy this **item** is the way the product is presented, often the good packages one are the more expensive compared to others. But is it because the package looks good that it means the quality will be good also? This is something society should reflect about. Moreover, the **media** also is an example, where entertainers, politicians and other public figures, try to appear in TVs or social medias other than they are in reality. The viewer of this contents are **normally** thinking wrongly and may have a better opinion about these people as these **individuals** appear something that in the reality they are not. Therefore, people should realize that impressions are not always what they think they are seeing, and that they should question their-selves and be more critical about anything they see or watch.	In my personal opinion people can ***achieve*** more success bu ***cooperating***. i think competition can make anything ugly. Competition brings envy and jealousy, which and sometimes result in pain and disappointment. I agree with the fact that in a ***complex*** world, ***cooperation*** is much more likely to produce ***significant*** and lasting accomplishments because they bring about **mental** peace and happiness. If a person wins something by harsh competing and a sense of entitlement, but that feeling of winning something does not last long as it is **temporary**. Anything and everything in the world is **temporary** and it is the same with competitions. But if you **cooperate** with your opponent or enemy it is more likely to **benefit** you in the long run. There will no room for hatred or jealousy between you and your opponent ages after the ***cooperation***. A sense of urgency and competitions may be the driving force of your life, but it certainly will not last forever as it is driven by something **negative**. I am personally not a bad fan of competition. I feel that every person in this world have different **capabilities** and strengths and it is not fair to to try to **equate** the most strongest and the mild or the mild with the weak. We have often seen this **trend** of ***cooperation*** over competitions in history where a country is trying to get their freedom from a ruling nation. They opt for ***cooperation*** or silent protest rather than direct war or competition as , first they are **aware** of their strengths and shortcomings, secondly they are **aware** of the competition's outcomesee. In **conclusion**, competition in the world does more harm than good, as they make people feel **insecure**, it is dangerous for life long relationships and it also tests people's strength in different **medium**. If in such situation where i will have to choose between competing with someone or simply agreeing to ***cooperate*** in the middle ground, I would always choose for settlement as it will be in both the parties best interest.

Note. Academic words are shown in bold.; Academic words that also appear in the prompt are shown in italics.; In the essay at Time 2, additional one academic word (*outcomes*) was misspelled and thus not calculated toward the number of academic words.

5.2.4.2 Nonsignificant covariances among writing scores, burst lengths, and academic word percentages over time

Six longitudinal covariances among writing scores, burst lengths, and academic word percentages were nonsignificant. First of all, two nonsignificant covariances were related to initial academic word percentages, such that initial academic word percentages did not influence burst length changes or writing score changes. Thus, burst length changes and writing score changes likely happened independently of the use of academic words at the initial time of measurement.

Next, two nonsignificant covariances were related to initial burst lengths, such that initial burst lengths did not influence academic word percentages changes or writing score changes. This finding also indicates that academic word percentage changes and writing score changes may occur independently of burst lengths at the initial time of measurement.

The remaining two nonsignificant covariances were related to burst length changes, such that burst length changes were not linked to academic word percentages changes or writing score changes. Thus, changes in burst lengths may not influence academic word percentage changes or writing score changes, and vice versa.

5.3 Overall Discussion for Research Question 2

The second research question focused on the relationships among English writing scores, burst length, and the use of academic words over time. Mean burst length characterized a process feature of the translator (i.e., transforming ideas into language strings), while the use of academic words characterized a product feature of the translator. Two overarching findings are discussed below.

First, changing patterns in both language burst lengths and academic word percentages which hinted at students becoming better writers were found in multilingual students who received lower initial writing scores. Thus, this finding seems to describe a "poor get richer" scenario as also found in results of the research question 1, rather than "rich get richer" (i.e., the Matthew effect; Stanovich, 1986). More specifically, with respect to language burst changes, a notable pattern of decreases in burst lengths over time was found in multilingual students who tended to receive lower scores on initial writing and produce excessively longer bursts. Importantly, given that these students had gains in writing scores over time, such decreasing pattern in burst lengths may be an indication of improvement that they produced a more reasonable amount of language strings at Time 2 as compared when they had produced a very long stretch of ideas in a single burst at Time 1. On the other hand, with respect to academic word percentages as found in student essays, an increasing pattern in academic word percentages was found in multilingual students who tended to receive lower scores on initial writing and use a smaller number of academic words at Time 1. Thus, it seems that in the context of producing persuasive essays at the college level, multilingual students whose initial writing skills were not fully developed tended to have more potential for improving their writing not only in their translating processes but also their lexical use during writing.

Second, gains in higher writing scores were related to gains in academic word percentages, but not gains in burst lengths. This finding highlights the importance of improving the translator in using more sophisticated words (i.e., academic words as compared to everyday words) in order to have greater gains in writing scores, rather than producing a greater quantity of language strings. Thus, the role of the translator is emphasized in quality (i.e., the use of

sophisticated words) over quantity (i.e., the number of characters produced in a single burst) to have greater gains in the ability to produce college-level persuasive essays in English.

6 LONGITUDINAL RELATIONSHIP AMONG ENGLISH WRITING, READING, AND VOCABULARY

The third research question focused on the longitudinal relationship among three literacy-related variables: English writing ability, English reading comprehension ability, and English vocabulary knowledge, each of which was measured at two time points with intervals of at least five months. Two sub-research questions were examined. For the research question 3a, all of the possible cross-sectional and longitudinal relationships among the three variables were examined: covariances among initial levels, covariances among changes, covariances between initial levels and changes of each variable, and cross-lagged covariances between initial levels and changes across different variables (e.g., covariance between the initial level of English vocabulary and the change in English writing). Research question 3b tested whether a common latent variable of English literacy that was informed by English writing, English reading, and English vocabulary could be constructed, and whether there would be a mean change in the latent variable over time.

6.1 Research Question 3 Results

6.1.1 Descriptive statistics

For the research question 3, 77 participants' data were used. Among them, one student's vocabulary score at Time 2 was unrealistically low with a raw score of two (as compared to this student's vocabulary score of 13 at Time 1). Thus, this student's student vocabulary score at Time 2 was deleted and handled as missing data using a FIML approach. Table 6.1 presents the descriptive statistics of writing scores at two time points, changes in writing scores, vocabulary test scores at two time points, changes in vocabulary scores, reading scores at two time points,

and changes in reading scores. Score changes were calculated by subtraction (i.e., scores at Time 2 minus scores at Time 1). Scatter plots and histograms of these variables are presented in Figure 6.1.

Table 6.1 Descriptive Statistics for Variables Related to Research Question 3

Variable	*N*	Mean	*SD*	Min.	Max.	Skewness	Kurtosis
Writing score at Time 1	77	2.92	1.16	1	5.50	.39	–.35
Writing score at Time 2	77	4.04	1	1.50	6	.08	–.36
Change in writing score	77	1.12	1.19	–1	4.50	.68	.34
Vocabulary score at Time 1	77	555.21	27.93	499	612	.02	–.88
Vocabulary score at Time 2	76	557.96	31.30	478	636	–.06	–.10
Change in vocabulary score	76	2.30	15.29	–30	66	.77	2.43
Reading score at Time 1	77	564.71	26.98	513	653	.49	.68
Reading score at Time 2	77	566.71	25.46	507	643	.21	.07
Change in reading score	77	2.00	20.74	–47	78	.52	1.87

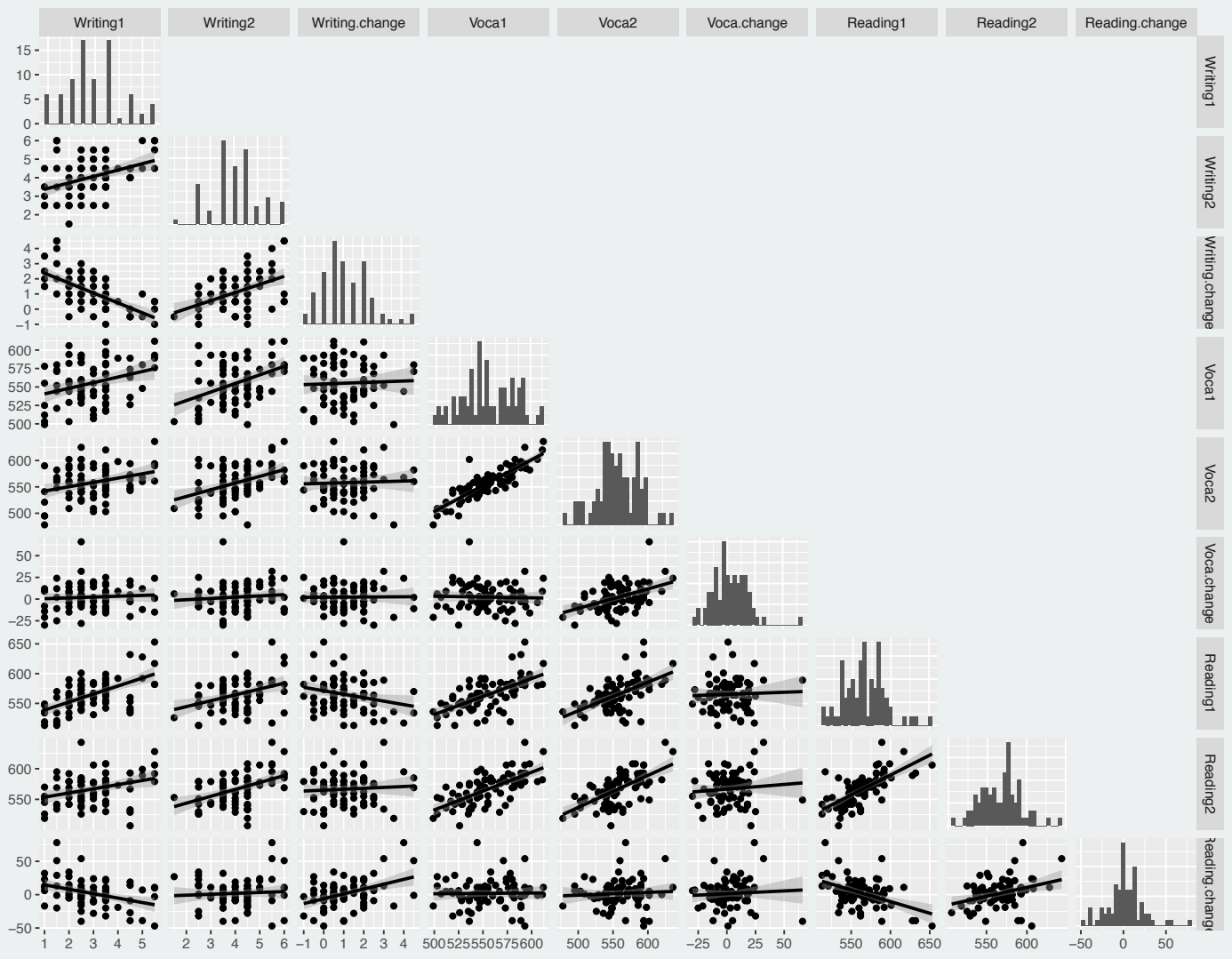


Figure 6.1 Scatter Plots and Histograms of Variables Related to Research Question 3

Note. 1 = Time 1, 2 = Time 2, Voca = vocabulary; Diagonal graphs show histograms of each variable. On scatter plots, lines indicate linear predictions from a linear regression model of the two variables.

The changes in writing, vocabulary, and reading scores showed increasing patterns. Average writing scores increased from 2.92 to 4.04 over time with a mean change of 1.12. Average vocabulary scores increased from 555.21 to 557.96 with a mean change of 2.30. Average reading scores increased from 564.71 to 566.71 with a mean change of 2. Figure 6.2 graphically shows changes in writing scores, changes in vocabulary scores, and changes in reading scores over time.

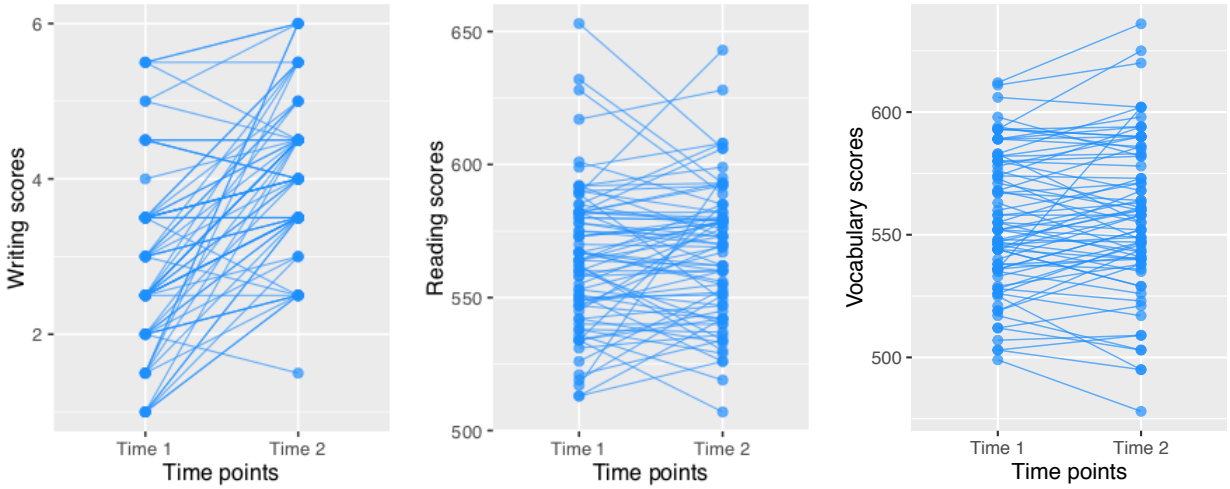


Figure 6.2 Writing Score Changes (Left; N = 77), Reading Score Changes (Center; N = 77), and Vocabulary Score Changes (Right; N = 76).

In reference to norm groups of native-English speaking students, the students' vocabulary mean score was approximately equivalent to grade levels between 11 and 12 of native-English speaking students (MacGinitie et al., 2000b). The students' reading mean score was approximately equivalent to the 12th grade of native-English speaking students (MacGinitie et al., 2000b).

6.1.2 Correlation analysis

Pearson correlations among the variables related to the research question 3 are shown in Table 6.2. Correlations were significant between writing scores at Times 1 and 2 (r = .40, p <

.01), betwteen vocabulary scores Times 1 and 2 (r = .87, p < .01), and between reading scores at Times 1 and 2 (r = .69, p < .01). Corelations between intial levels and changes were negative for writing scores (r = –.64, p < .01) and reading scores (r = –.46, p < .01), but no correlation was found between the intial level and changes for vocabulary scores (r = –.03, p > .05). On the other hand, correlations between score changes and scores at Time 2 were positive for writing (r = .45, p < .01), vocabulary (r = .46, p < .01), and reading (r = .33, p < .01).

Table 6.2 Correlations among Variables Related to Research Question 3

Variable	1	2	3	4	5	6	7	8	9
1 Writing score at Time 1	1								
2 Writing score at Time 2	.40**	1							
3 Change in writing score	–.64**	.45**	1						
4 Vocabulary score at Time 1	.32*	.42**	.04	1					
5 Vocabulary score at Time 2	.30*	.40**	.04	.87**	1				
6 Change in vocabulary score	.07	.09	.01	–.03	.46**	1			
7 Reading score at Time 1	.58**	.37**	–.26**	.63**	.57**	.04	1		
8 Reading score at Time 2	.31*	.44**	.07	.68**	.64**	.09	.69**	1	
9 Change in reading score	–.37**	.07	.42*	.01	.06	.06	–.46**	.33**	1

Note. *N*s for vocabulary scores at Time 1 and changes in vocabulary scores were 76, while *N*s for the other variables were 77.; ** indicates p < .01, and * indicates p < .05.

6.1.3 Longitudinal relationship among writing, reading, and vocabulary in English

For the research question 3a (i.e., relationships among English writing, reading, and vocabulary), results of testing the latent change score model as shown in Figure 3.5 are presented in Table 6.3 and Figure 6.3. Change statistics (i.e., intercepts [means] and variances of initial scores and changes) are shown in Table 6.3 but not in Figure 6.3 for clarity of presentation. Change were significant in writing scores, but not in vocabulary scores or reading scores (see the 'intercept' section in Table 6.3), which indicates that students' writing scores increased over

time, while their reading and vocabulary scores did not differ over time. On the other hand, there was significant variability in changes in writing scores, vocabulary scores and reading scores (see the ‘variance’ section in Table 6.3), which indicates that individual students showed different degrees of changes in writing, vocabulary, and reading scores over time.

Table 6.3 Results of Latent Change Score Model for Research Question 3a

Intercept	Estimate	*SE*	*z*	*p*	Standardized Estimate
Writing[1]	2.92	.13	22.28	< .01	2.54
ΔWriting	1.12	.14	8.30	< .01	.95
Vocabulary[1]	555.21	3.16	175.60	< .01	20.01
ΔVocabulary	2.29	1.73	1.32	.19	.15
Reading[1]	564.71	3.06	184.84	< .01	21.07
ΔReading	2.00	2.35	.85	.39	.10
Variance	**Estimate**	***SE***	***z***	***p***	**Standardized Estimate**
Writing[1]	1.33	.20	6.69	< .01	1.00
ΔWriting	1.40	.25	5.64	< .01	1.00
Vocabulary[1]	769.78	95.20	8.09	< .01	1.00
ΔVocabulary	230.65	56.69	4.07	< .01	1.00
Reading[1]	424.39	96.70	4.39	< .01	1.00
ΔReading	718.70	136.46	5.27	< .01	1.00
Covariance	**Estimate**	***SE***	***z***	***p***	**Standardized Estimate**
ΔWriting ↔ Writing[1]	–.87	.16	–5.46	< .01	–.64
ΔVocabulary ↔ Vocabulary[1]	–13.27	47.57	–.28	.78	–.03
ΔReading ↔ Reading[1]	–251.77	96.76	–2.60	< .05	–.46
Writing[1] ↔ Vocabulary[1]	10.06	4.05	2.49	< .05	.32
Writing[1] ↔ Reading[1]	17.82	4.79	3.72	< .01	.58
Reading[1] ↔ Vocabulary[1]	465.36	88.38	5.27	< .01	.63
ΔWriting ↔ ΔReading	10.15	4.18	2.43	< .05	.42
ΔWriting ↔ ΔVocabulary	.18	1.96	.09	.93	.01
ΔReading ↔ ΔVocabulary	18.68	51.62	.36	.72	.06
ΔWriting ↔ Vocabulary[1]	1.37	3.75	.37	.71	.04
ΔWriting ↔ Reading[1]	–8.10	3.83	–2.12	< .05	–.26
ΔVocabulary ↔ Writing[1]	1.24	1.98	.63	.53	.07
ΔVocabulary ↔ Reading[1]	16.69	43.97	.38	.70	.04
ΔReading ↔ Writing[1]	–8.81	3.11	–2.83	< .05	–.37
ΔReading ↔ Vocabulary[1]	8.38	62.70	.13	.89	.02

Note. [1] = Time 1, Δ = "change in", ↔ = "covariance with"

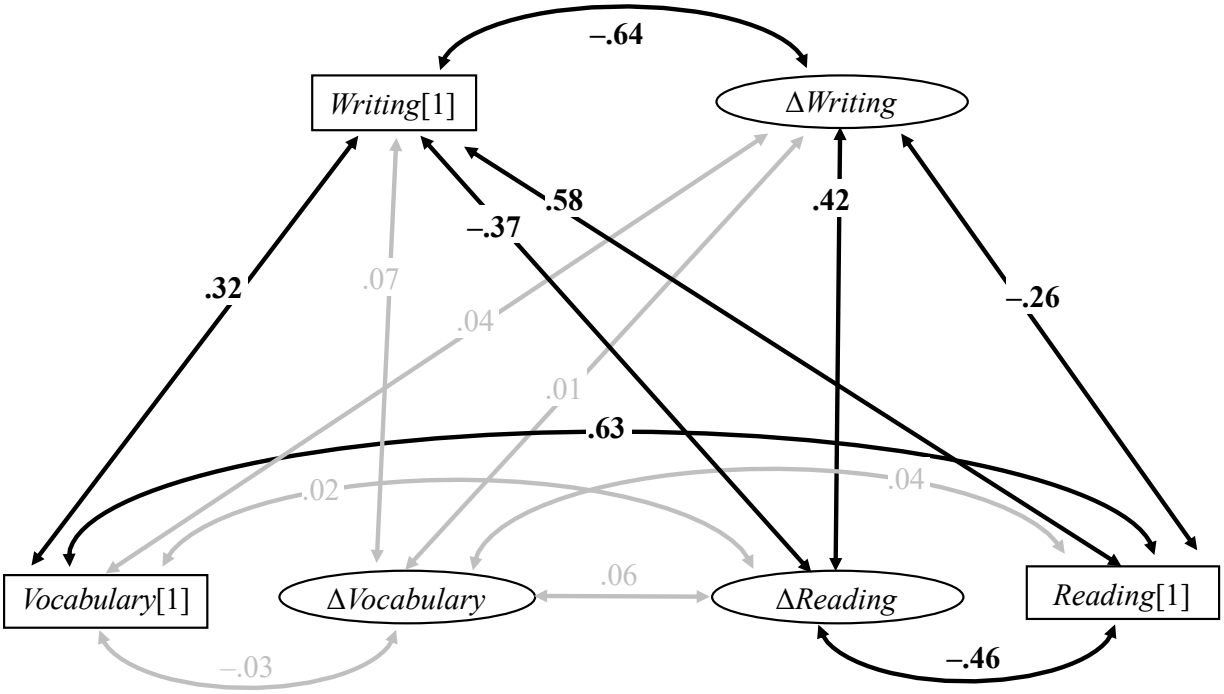


Figure 6.3 Latent Change Score Model for Research Question 3a
Note. Δ = "change in", [1] = Time 1; Estimates are standardized coefficients. Significant paths are shown in black color, while nonsignificant paths are shown in grey color.

To answer the research question 3b (i.e., longitudinal relationships among English writing scores, English vocabulary scores, and English reading scores), four types of covariances are examined (see Figure 6.3 and the 'covariance' section in Table 6.3). These four types were (a) covariances among initial levels of the three variables, (b) covariances among changes in the three variables, (c) covariances between initial levels and changes of each variable, and (d) cross-lagged covariances between initial levels and changes across different variables. These model-based covariances are shown with correlation coefficients (i.e., standardized covariance estimates) in Table 6.4. Results of each covariance type are reported below.

Table 6.4 Model-Based Correlations related to Research Question 3a

Variable	1	2	3	4	5	6
1 Writing[1]	1					
2 Vocabulary[1]	.32*	1				
3 Reading[1]	.58**	.63**	1			
4 ΔWriting	–.64**	.04	–.26*	1		
5 ΔVocabulary	.07	–.03	.04	.01	1	
6 ΔReading	–.37*	.02	–.46*	.42*	.06	1

Note. [1] = Time 1; Δ = "change in"; ** indicates $p < .01$, and * indicates $p < .05$.

Covariances among initial levels of the three variables. Initial levels of writing, vocabulary, and reading scores were related to each other, ranging *r* values from .32 ($p < .05$) to .63 ($p < .01$). This indicates that students with higher writing scores at Time 1 also tended to receive higher scores on reading and vocabulary at Time 1. Also, students with higher reading scores at Time 1 also tended to receive higher vocabulary scores at Time 1.

Covariances among changes in the three variables. When covariances among changes in the three variables (i.e., writing, vocabulary, and reading scores) were examined, one significant covariance was found, such that changes in English writing scores were positively related to changes in English reading scores ($r = .42$, $p < .05$). This suggests that gains in writing scores tended to go hand-in-hand with gains in reading scores. On the other hand, the covariance between writing score changes and vocabulary score changes was not significant ($r = .01$, $p > .05$). In addition, the covariance between reading score changes and vocabulary score changes was nonsignificant ($r = .06$, $p > .05$).

Covariances between initial levels and changes of each variable. Covariances between intial levels and changes were significantly negative for writing scores ($r = –.64$, $p < .01$), and reading scores ($r = –.46$, $p < .05$), but no significnat relation was found between initial levels of,

and changes in, vocabulary scores ($r = –.03$, $p > .05$). These results indicate that L2 students who received lower writing scores at Time 1 tended to have greater gains in writing scores over time, and those who received lower reading scores at Time 1 tended to have greater gains in reading scores over time. On the other hand, vocabulary scores at Time 1 did not relate to changes in vocabulary scores.

Cross-lagged covariances between initial levels and changes across different variables. Six cross-lagged covariances between initial levels and changes across the different variables were examined: covariances of initial writing scores with vocabulary score changes and reading score changes, covariances of initial reading scores with writing score changes and vocabulary score changes, and covariances of initial vocabulary scores with writing score changes and reading score changes.

Among the six, two significant cross-lagged covariances were found, such that reading scores at Time 1 were negatively associated with changes in writing scores ($r = –.26$, $p < .05$), and writing scores at Time 1 were negatively related to changes in reading scores ($r = –.37$, $p < .05$). These results indicate that multilingual students with lower initial English reading scores tended to have greater gains in English writing scores, and multilingual students with lower initial English writing scores tended to have greater gains in English reading scores.

The other four cross-lagged covariances that involved vocabulary scores were nonsignificant. Initial vocabulary scores were not realted with writing score changes ($r = .04$, $p > .05$) or reading score changes ($r = .02$, $p > .05$). Also, vocabulary score changes were not related with initial writing scores ($r = .07$, $p > .05$) or initial reading scores ($r = .04$, $p > .05$). These results indicate that initial vocabulary scores did not influence reading or writing score changes, and initial writing and reading scores did not influence vocabulary score changes.

Lastly, the latent change score model presented in Figure 6.3 was identified without a degree of freedom. To examine the model fit, a model that excluded nonsignificant covariances (n = 7) was tested. The results of the model without the nonsignificant covariances were almost the same (with minor differences in estimates), and the model fit was excellent (χ^2 = 2.34, df = 7, p = .94, CFI = 1, SRMR = .02).

6.1.4 Latent variable of English literacy over time

The research question 3b tested whether a common latent variable of English literacy that was informed by English writing, English reading, and English vocabulary could be constructed, and if so, whether there would be a mean change in the latent variable over time. Longitudinal measurement invariance analysis was conducted by three steps: Configural, weak, and strong invariance. Table 6.5 presents the results of goodness-of-fit statistics and model comparisons.

Table 6.5 Fit Statistics for Longitudinal Measurement Invariance Assessment

Model	χ^2	df	$\Delta\chi^2$	Δdf	Δp	CFI	ΔCFI	SRMR
Configural	19.35	5	-	-	-	.946	-	.045
Metric	18.86	7	.85	2	.66	.952	.006	.048
Scalar	62.23	9	43.75	2	< .01	.784	–.168	.159
$\text{Scalar}_{\text{partial}}$	18.90	8	.02	1	.88	.956	.004	.048

Note. χ^2 and CFI are robust measures.

First, the baseline model for configural invariance was tested. Configural invariance was met based on acceptable model fit indices (χ^2 = 19.35, df = 5, CFI = .946, SRMR = .045). Invariance at the configural level supported the notion that that the latent variable was formed by the same number of observed variables across time.

Second, given the evidence of configural invariance, metric measurement invariance was tested by constraining factor loadings to be equal across time. The metric invariance model was acceptable based on the goodness-of-fit statistics (χ^2 = 18.86, df = 7, CFI = .952, SRMR = .048).

In addition, the χ^2 difference test between the configural and metric invariance models was not significant ($\Delta\chi^2[2] = .85$, $\Delta p = .66$) and the ΔCFI value was greater than −.01, which revealed that metric invariance was supported. Metric measurement invariance supported the notion that each observed variable contributed to the latent variable to a similar degree across time.

Third, given the evidence of metric invariance, scalar measurement invariance was tested by constraining intercepts of the observed variables to be equal across time. The scalar invariance was not supported ($\chi^2 = 62.23$, $df = 9$, CFI = .784, SRMR = .159). Also, the χ^2 difference test between the metric and weak invariance models was significant ($\Delta\chi^2[2] = 43.75$, $\Delta p < .01$) and the ΔCFI value was smaller than −.01, which revealed that scalar invariance was not supported. Failing to meet scalar invariance indicates that mean differences in the latent variable did not capture all of the mean differences in the shared variance of the observed variables across time

The partial strict invariance was then tested by eliminating the constraints that might have added substantial chi-square values to the model. By this procedure, it was found that the equality constraints on the intercepts for writing scores resulted in a substantial increase in chi-square values to the model. This was probably because the degrees of the increase in writing scores were greater than those of the increases in reading and vocabulary scores, and that increase in writing scores was not related to the increased level of the *English literacy* latent factor. Thus, the constraints on the intercepts of writing scores were removed, and the intercepts of writing scores were freely estimated. The subsequent model supported partial scalar measurement invariance. Its goodness-of-fit statistics were acceptable ($\chi^2 = 18.90$, $df = 8$, CFI = .956, SRMR = .048). In addition, the χ^2 difference test between the metric and partial scalar invariance models was not significant ($\Delta\chi^2[1] = .02$, $\Delta p = .88$) and the ΔCFI value was greater

than −.01, which supported partial scalar invariance. This result indicates that mean differences in the latent variable captured the mean differences in the shared variance of reading and vocabulary scores but of writing scores across time.

Given that the partial scalar invariance model was supported, latent factor means and variances across time were compared. With the factor mean at Time 1 being set to zero and the factor variance at Time 1 being set to one, the estimate for the factor mean at Time 1 was .102 (*SE* = .07, *z* = 1.50, *p* = .13), which was not significantly different from zero. In addition, the covariance of the latent *English Literacy* variable between the two occasions was quite high (estimate = .99, *SE* = .09, *z* = 10.50, *p* < .01), which indicates that the *English Literacy* latent variable at Time 1 is closely related to (almost identical with) the *English Literacy* latent variable at Time 2. The results of the partial scalar invariance model are presented in Figure 6.4 and Table 6.6. Overall, these findings indicate that while the equality of the latent structure of the *English Literacy* factor was supported across time via partial scalar measurement invariance (with the exception of the intercepts of writing scores across time), the scores of the *English Literacy* latent factor did not change over time.

Additionally, when comparing factor loadings of the three observed variables on the latent variable, the standardized factor loadings of writing scores (.46 at Time 1 and .53 at Time 2) were much lower than those of vocabulary scores (.77 at Time 1 and .71 at Time 2) and those of reading scores (.83 at Time 1 and .88 at Time 2). In addition, the standardized residual variance (i.e., unexplained variance) of writing scores (.79 at Time 1 and .72 at Time 2) was much higher than that of vocabulary scores (.41 at Time 1 and .49 at Time 2) and that of reading scores (.31 at Time 1 and .23 at Time 2). Thus, writing was a weaker indicator of the common literacy factor at both times than vocabulary and reading. This result indicates that when creating

the latent variable of *English Literacy*, reading and vocabulary scores were more closely linked to each other than they were with writing scores.

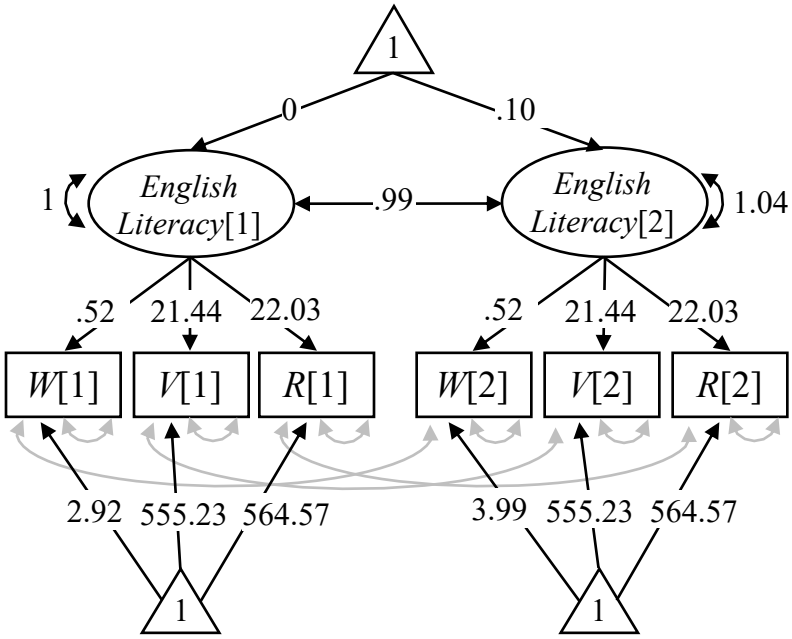


Figure 6.4 Partial Scalar Measurement Invariance Model
Note. W = Writing, V = Vocabulary, R = Reading, [1] = Time 1, [2] = Time 2; Estimates are unstandardized coefficients. Arrows beginning from △1 indicate estimates for mean scores. Residual variances and covariances of the indicator variables are shown in grey and their estimates are presented in Table 6.6.

Table 6.6 Results of Partial Scalar Measurement Invariance Model

Factor loading	Estimate	*SE*	*z*	*p*	Standardized Estimate
Writing[1] loading on *Literacy*[1]	.52	.12	4.52	< .01	.46
Vocabulary[1] loading on *Literacy*[1]	21.44	3.08	6.97	< .01	.77
Reading[1] loading on *Literacy*[1]	22.03	3.12	7.06	< .01	.83
Writing[2] loading on *Literacy*[2]	.52	.12	4.52	< .01	.53
Vocabulary[2] loading on *Literacy*[2]	21.44	3.08	6.97	< .01	.71
Reading[2] loading on *Literacy*[2]	22.03	3.12	7.06	< .01	.88
Indicator intercept	Estimate	*SE*	*z*	*p*	Standardized Estimate
Writing[1] intercept	2.92	.13	22.28	< .01	2.58
Vocabulary[1] intercept	555.23	3.17	175.23	< .01	19.98
Reading[1] intercept	564.57	2.77	204.06	< .01	21.36
Writing[2] intercept	3.99	.11	34.90	< .01	3.95
Vocabulary[2] intercept	555.23	3.17	175.23	< .01	18.12
Reading[2] intercept	564.57	2.77	204.06	< .01	22.08
Latent factor intercept, variance, and covariance	Estimate	*SE*	*z*	*p*	Standardized Estimate
Literacy[1] intercept	0				.00
Literacy[2] intercept	.10	.07	1.50	.13	.10
Literacy[1] variance	1				1.00
Literacy[2] variance	1.04	.18	5.69	< .01	1.00
Literacy[1] ↔ *Literacy*[2]	.99	.09	10.50	< .01	.97
Indicator residual variance	Estimate	*SE*	*z*	*p*	Standardized Estimate
Writing[1] residual variance	1.01	.15	6.56	< .01	.79
Vocabulary[1] residual variance	312.72	97.47	3.21	< .01	.41
Reading[1] residual variance	213.06	107.23	1.99	.05	.31
Writing[2] residual variance	.74	.13	5.69	< .01	.72
Vocabulary[2] residual variance	460.06	116.23	3.96	< .01	.49
Reading[2] residual variance	148.20	105.27	1.41	.16	.23
Indicator residual covariance	Estimate	*SE*	*z*	*p*	Standardized Estimate
Writing[1] ↔ Writing[2]	.20	.12	1.62	.11	.23
Vocabulary[1] ↔ Vocabulary[2]	285.30	107.94	2.64	< .01	.75
Reading[1] ↔ Reading[2]	–13.92	92.14	–.15	.88	–.08

Note. [1] = Time 1, [2] = Time 2, ↔ = "covariance with"

6.2 Research Question 3a Discussion

6.2.1 Summary of results of research question 3a

The research question 3a focused on cross-sectional and longitudinal relationships among English writing, English reading, and English vocabulary scores. With respect to cross-sectional relationships, all of English writing, reading, and vocabulary scores measured at Time 1 were related to each other, indicating their close relationships. On the other hand, longitudinal relationships were found in reading and writing scores, but not in vocabulary scores in relation to writing or reading. Specifically, writing scores at Time 1 were negatively related to both writing score changes and reading score changes, while reading scores at Time 1 were negatively related to both writing score changes and reading score changes. These findings indicate that multilingual students who received lower reading and writing scores at the initial time of measurement tended to have greater gains in reading and writing scores over time. In addition, writing score changes were positively related to reading score changes, indicating that gains in reading and writing scores may go hand-in-hand. Lastly, vocabulary scores did not show any longitudinal relationship with reading or writing scores. The summary of these results is presented in Table 6.7.

Table 6.7 Summary of the Relationship among Writing, Reading, and Vocabulary scores

Relationship among initial scores		Relationship among scores across time	
Covariance	Significance	Covariance	Significance
Writing[1] ↔ Reading[1]	Yes (+)	Writing[1] ↔ ΔWriting	Yes (–)
Writing[1] ↔ Vocabulary[1]	Yes (+)	Writing[1] ↔ ΔReading	Yes (–)
Vocabulary[1] ↔ Reading[1]	Yes (+)	Reading[1] ↔ ΔReading	Yes (–)
		Reading[1] ↔ ΔWriting	Yes (–)
		ΔWriting ↔ ΔReading	Yes (+)
		Vocabulary[1] ↔ ΔVocabulary	No
		Vocabulary[1] ↔ ΔWriting	No
		Vocabulary[1] ↔ ΔReading	No
		ΔVocabulary ↔ Writing[1]	No
		ΔVocabulary ↔ Reading[1]	No
		ΔVocabulary ↔ ΔWriting	No
		ΔVocabulary ↔ ΔReading	No

Note. [1] = Time 1, Δ = "change in", ↔ = "covariance with". Directions of significant covariances (positive vs. negative) are shown in parentheses.

Below, changes in writing, reading, and vocabulary scores are first briefly discussed. Cross-sectional and longitudinal relationships among writing, reading, and vocabulary scores are then discussed.

6.2.2 Changes in writing, reading, and vocabulary scores

Results of the latent change score modeling indicated that score changes in writing were significantly positive, while score changes in reading and writing were not significant. This indicates that on average, multilingual students tended to improve their English writing skills, but not reading skills and vocabulary scores. However, it should be mentioned that significant variances of the changes scores in writing, reading, and vocabulary were revealed, suggesting interindividual variability in the patterns of changes. That is, although reading and vocabulary scores did not change over time as a group, individual students within the group showed

interindividual variability in changes in all of the three variables. As changes in writing, reading, and vocabulary scores were addressed at the latent level in the research question 3b, more detailed discussion related to change statistics is provided in Section 6.3.

6.2.3 Cross-sectional relationship among initial levels of writing, reading, and vocabulary scores

Initial levels of writing, vocabulary, and reading scores in English were related to each other. These findings corroborate previous research that has reported close relationships between L2 writing and L2 vocabulary (Lu, 2010; Milton, Wade & Hopkins, 2010; Stæhr, 2008; Schoonen et al., 2003, 2011), between L2 wrting and L2 reading (Belcher & Hirvela, 2001; Carson et al, 1990; Pae, 2018), and between L2 reading and L2 vocabulary (Jeon & Yamashita, 2014; Koda, 2005; van Gelderen et al., 2004).

6.2.4 Longitudinal relationship among writing, reading, and vocabulary scores

Using trivariate latent change score modeling, a total of 12 longitudinal covariances among writing, and reading, and vocabulary scores were examined. Among the 12, five significant longitudinal relationships were revealed, while the remaining seven relationships were nonsignificant. Below, significant covariances are first discussed followed by discussion on nonsignificant covariances.

6.2.4.1 Significant covariances among writing, reading, and vocabulary scores over time

The five significant longitudinal relationships involved English reading and writing scores but not vocabulary scores. These five relationships were found between writing scores at Time 1 and writing score changes, between reading scores at Time 1 and writing score changes, between reading scores at Time 1 and reading score changes, between reading scores at Time 1 and writing score changes, and between writing score changes and reading score changes.

Results showed that writing score changes were negatively related with initial writing scores and initial reading scores, indicating that multilingual students who received lower initial English writing scores and initial lower English reading scores tended to receive greater gains in writing scores over time. These results have already discussed in Section 4.2, and thus are not discussed in this section.

Similar to writing score changes, reading score changes were also negatively related with initial reading scores and initial writing scores. This result suggests that multilingual students who received lower initial English reading scores and lower initial writing scores tended to have greater gains in English reading scores. To illustrate, three groups with different levels of reading score changes were created and compared: one group of students whose reading scores increased by six or more (henceforth, 'reading-increase-group'; n = 30), another group of students whose reading scores decreased by six or more (henceforth, 'reading-decrease-group'; n = 23), and another group of students whose reading scores did not change or changed fewer than six ('reading-same-group'; n =24).[9]

Figure 6.5 shows two boxplots for the three groups' initial reading scores (left) and the three groups' initial writing scores (right). The reading-increase group tended to receive lower reading scores at Time 1 (M = 556.17, SD = 27.51) than the reading-decrease group (M = 574.78, SD = 31.43) and the reading-same group (M = 565.75, SD = 17.57). In addition, the reading-increase-group tended to receive lower writing scores at Time 1 (M = 2.62, SD = 1.35) than the reading-decrease-group (M = 3.37, SD = 1.09) and the reading-same group (M = 2.88, SD = .82).

[9] The cut-off of the score difference of six was chosen because norm groups' ESS score differences between adjacent grades (e.g., between grade 10 and grade 11) were around six (MacGinitie et al., 2000b). The reading-increase-group's mean reading score change was 20.50 (SD = 16.21), ranging from 6 to 78. The reading-decrease-group's mean reading score change was –19.87 (SD = 12.51), ranging from –47 to –6. The reading-same-group's mean reading score change was –.17 (SD = 3.25) ranging from –5 to 5.

These results clearly show that multilingual students who had greater gains in English reading scores tended to receive lower initial English reading scores and lower initial writing scores.

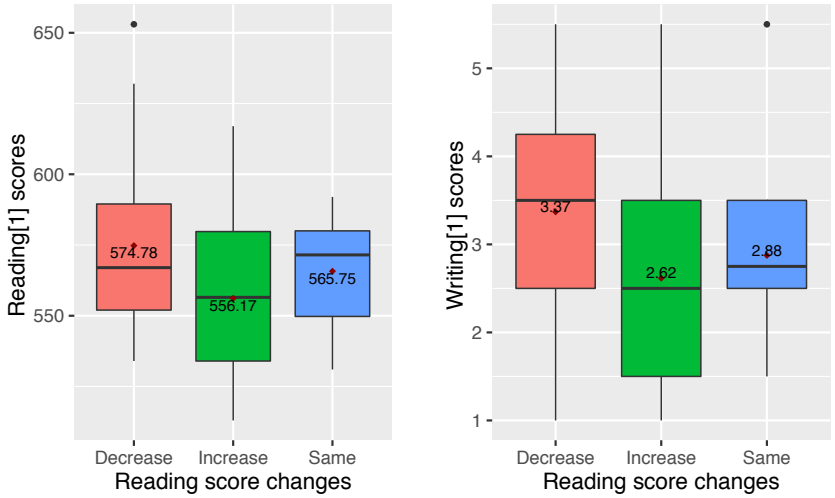


Figure 6.5 Initial Reading Scores (left) and Initial Writing Scores (right) of Three Groups with Different Reading Score Changes

Note. A horizontal line within each box indicates each group's median.; A dot within each box indicates each group's mean.

Lastly, reading score changes were positively related to writing score changes, which suggests that multilingual students who tended to have greater gains in reading scores also tended to have greater gains in writing scores over time. To illustrate, the three groups with different levels of reading score changes were compared (see Figure 6.6). The reading-increase group tended to have greater gains in writing scores (M = 1.53, SD = 1.26) than the reading-decrease group (M = .76, SD = 1.09) and the reading-same group (M = .94, SD = 1.10).

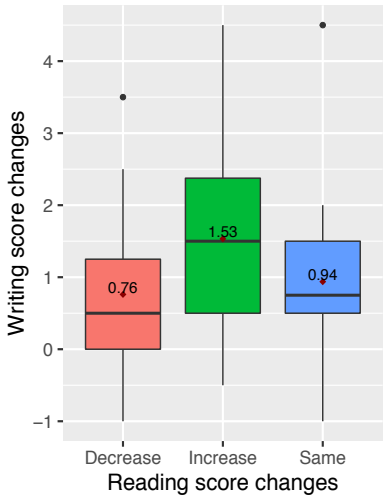


Figure 6.6 Writing Score Changes of Three Groups with Different Reading Score Changes
Note. A horizontal line within each box indicates each group's median.; A dot within each box indicates each group's mean.

Taken together, findings indicate longitudinal associations between English reading ability and English writing ability. Specifically, less proficient English readers and writers in higher education tended to improve their English reading and writing skills (as evidenced by greater gains in L2 reading and writing test scores over time) at a greater degree than more proficient English readers and writers. These findings indicate that for college multilingual students, initial lower levels of English writing and English reading may leave greater potential for improvement in better producing persuasive essays and better comprehending the author's messages.

Importantly, improvements in English reading and writing also tended to go hand-in-hand. This finding may be explained by the unique features shared by reading and writing processes at the discourse level (Berninger et al., 2002; Fitzgerald & Shanahan, 2000). Specifically, being more proficient English readers generally means that they become more skilled at understanding the author's words and logics and constructing meanings of the text, which may in turn help them create their own meanings in English in a coherent and logical manner. By the same token,

being more proficient English writers generally means having greater linguistic skills in English at their disposition, which may also help them understand another person's language.

6.2.4.2 Nonsignificant covariances among writing, reading, and vocabulary scores over time

Among the 12 longitudinal covariances among writing, and reading, and vocabulary scores, seven were nonsignificant. Interestingly, all of these seven insignificant covariances involved vocabulary scores. Specifically, two were related to initial vocabulary scores, while the other five were related to vocabulary score changes. Each is briefly discussed below.

With respect to the two nonsignificant covariances related to initial vocabulary scores, initial vocabulary scores were not related to writing score changes, suggesting that initial vocabulary scores did not influence writing score changes (see also Section 4.2.5). Initial vocabulary scores were also not related to reading score changes, indicating that initial vocabulary scores did not influence reading score changes. To illustrate, initial vocabulary scores of the three reading groups created in Section 6.2.4.1 were compared. The three groups with different levels of reading score changes did not show noticeable differences in initial vocabulary scores (see Figure 6.7).

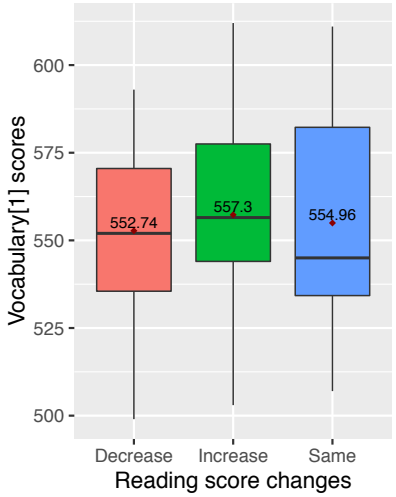


Figure 6.7 Initial Vocabulary Score Changes of Three Groups with Different Reading Score Changes

Note. A horizontal line within each box indicates each group's median.; A dot within each box indicates each group's mean.

With respect to the other five nonsignificant covariances related to vocabulary score changes, three groups with different levels of vocabulary score changes were created and compared: one group of students whose vocabulary scores increased by six or more (henceforth, 'vocabulary-increase-group'; $n = 29$), another group of students whose vocabulary scores decreased by six or more (henceforth, 'vocabulary-decrease-group'; $n = 22$), and another group of students whose vocabulary scores did not change or changed by fewer than six ('vocabulary-same-group'; $n = 25$).

Vocabulary score changes were not related to initial vocabulary scores, initial reading scores, or initial writing scores. That is, regardless of whether students' vocabulary scores increased (the vocabulary-increase group) or decreased (the vocabulary-decrease group), their initial vocabulary scores were similar (see Figure 6.8.a), their initial reading scores were similar (see Figure 6.8.b), and their initial writing scores were similar (see Figure 6.8.b). In addition, vocabulary score changes were also not related to either reading score changes or writing score changes. Regardless of whether students' vocabulary scores increased or decreased, their reading score changes did not differ (see Figure 6.8.d), and their writing score changes did not differ (see figure 6.8.e).

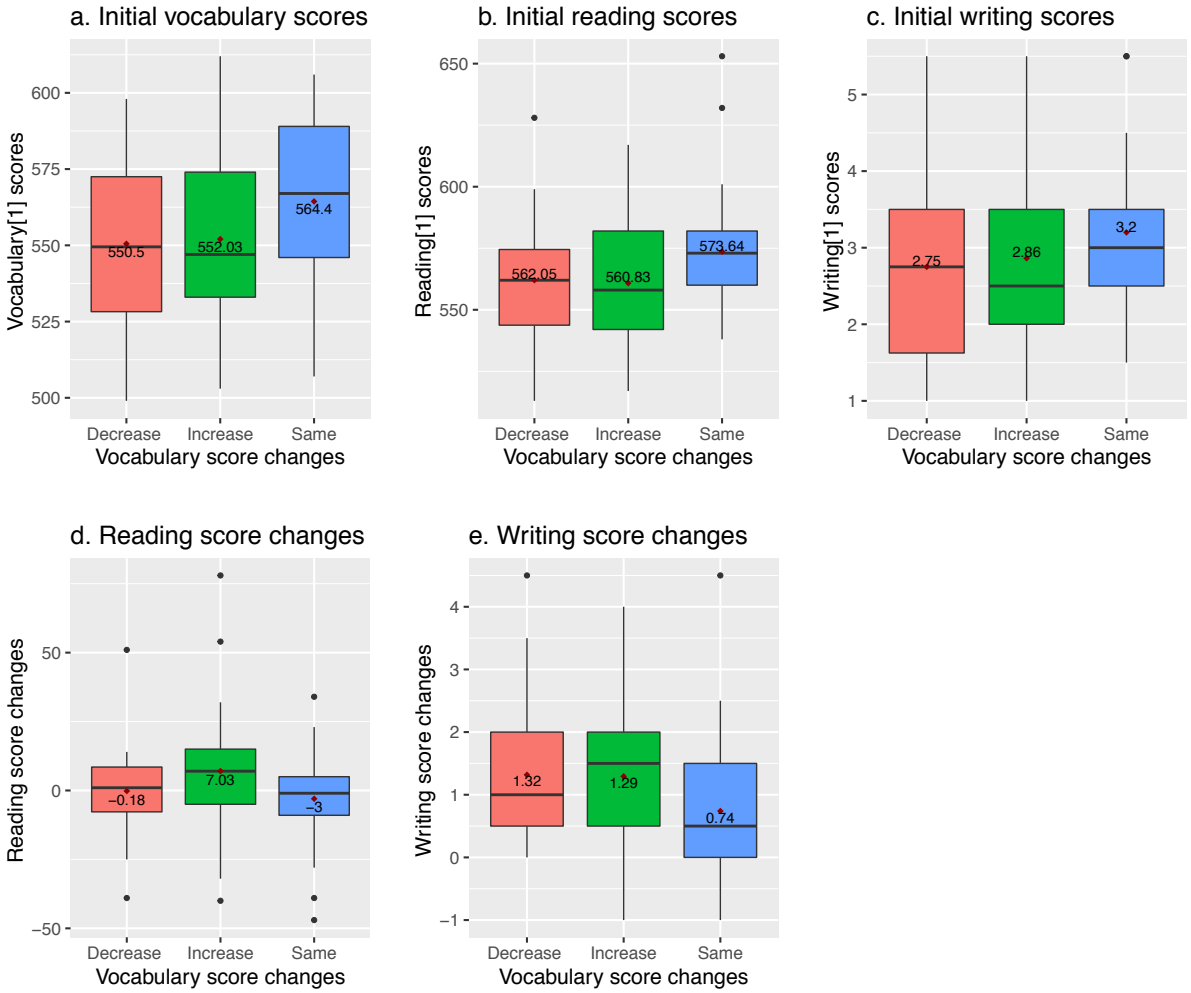


Figure 6.8 Three Groups with Different Vocabulary Score Changes
Note. A horizontal line within each box indicates each group's median.; A dot within each box indicates each group's mean.

In sum, while vocabulary scores were linked to both reading and writing scores cross-sectionally, they did not have any longitudinal relationship with reading or writing scores. This lack of the longitudinal relationship of initial vocabulary knowledge with reading and writing gains is surprising given the reported close cross-sectional relationships between L2 writing and L2 vocabulary knowledge (Schoonen et al., 2003; Stæhr, 2008) and L2 reading and L2 vocabulary (Jeon & Yamashita, 2014; Koda, 2005; van Gelderen et al., 2004). One potential reason for the lack of this relationship is that multilingual students at the college level may have

reached a certain level of their receptive vocabulary size at which they can read passages and produce text without difficulty. In this respect, having greater English vocabulary knowledge at the initial time of measurement may have neither benefited nor penalized students' English reading and writing gains over time. Thus, at the college level, improving English reading and writing skills may not necessarily pertain to becoming more vocabulary-savvy. Rather, to be better English writers and readers, they may need to learn how to make arguments and provide evidence in English writing and how to better understand the author's messages and read between lines at the discourse level.

In addition, vocabulary score changes were not related to any of the initial vocabulary, reading, and writing scores. This suggests that vocabulary changes likely occur independently of the existing levels of English knowledge and skills.

6.3 Research Question 3b Discussion

The research question 3b tested whether a common latent variable of English literacy informed by English writing, reading, and vocabulary could be constructed, and whether there was a mean score difference in the latent variable over time. When the latent *English Literacy* variable was constructed over time, partial scalar measurement invariance was obtained, which allowed for comparisons between latent mean scores across time. The *English Literacy* variable did not increase significantly over time, which suggests that the latent mean score of the latent *English Literacy* variable remained the same across time.

Results of testing longitudinal measurement invariance showed that the latent variable of *English Literacy* was constructed as a unidimensional construct over time and measured in the same manner across the two measurement times with the exception of the intercepts of writing scores. The lack of equality constraints on the intercepts of writing scores indicates that an

increase in writing scores was not related to the increased level of the *English literacy* latent factor potentially due to greater increases in writing scores than those in reading and vocabulary scores over time. In addition, the latent mean of the *English Literacy* variable did not increase significantly over time potentially because vocabulary and reading scores did not change much. This result is in line with the finding of the research question 3a that writing scores significantly increased but reading and vocabulary scores remained unchanged over time. Greater changes in writing scores than those in reading and vocabulary scores can be explained by characteristics of college-level courses. Specifically, taking academic courses in higher education may facilitate multilingual undergraduate students' improvement in English writing skills (specifically in producing persuasive essays), but not much in English reading skills and vocabulary knowledge. Indeed, given that an essay is one of the most frequent genres produced in higher education for assessment purposes (Gardner & Nesi, 2013), the students may have produced various essays, such as term papers, which likely led to their gains in writing scores over time. On the other hand, although academic courses in higher education typically have reading assignments, these reading materials tend to be discipline-specific. Thus, reading discipline-specific texts may have not been conducive to improving general reading skills and vocabulary knowledge as measured by Gates-MacGinitie tests.

Additionally, in creating the latent *English Literacy* variable, it is worthy of noting that among the three observed variables, writing scores were weaker indicators of the common literacy factor at both times with smaller factor loadings and considerable unexplained residual variances. In contrast, reading and vocabulary scores were stronger indicators of the common literacy factor with higher factor loadings and smaller unexplained residual variances at both times, which supports the greater shared variances between reading and vocabulary in the latent

structure across time. That is, in the latent structure, reading and vocabulary were more tightly coupled at each time of measurement as compared to reading-writing and writing-vocabulary.

In short, tests of longitudinal measurement invariance support the presence of the underlying, common *English Literacy* variable that captures English writing, reading, and vocabulary across time. Also, this latent *English Literacy* variable tended to be stable over time

6.4 Overall Discussion for Research Question 3

The third research question examined the longitudinal relationship English writing ability, English reading ability, and English vocabulary knowledge via latent change score modeling and longitudinal measurement invariance. Two overarching findings are discussed below.

First, longitudinally, English reading ability and writing ability are closely linked to each other over time, while having no longitudinal relationship with vocabulary knowledge (see Table 6.3). For example, greater gains in reading scores were related to greater gains in writing scores, indicating the longitudinal positive relationship between reading and writing in English. In addition, the "poor get richer" scenario seems to fit into the findings, such that initial lower levels of English writing and reading may leave much room for growth in better producing persuasive essays and better understanding text.

Another main finding is that English reading ability and vocabulary knowledge tend to be more tightly related with each other than English writing ability across time (see Table 6.6). This was supported by the finding that reading and vocabulary were stronger indicators of the latent *English Literacy* variable than writing at both times of measurement. This closer relationship between reading and vocabulary may be explained by the nature of the shared processes of understanding input (as compared to writing processes of producing output), and,

methodologically, by the nature of the shared test formats of multiple-choice questions (as compared to open-ended essays).

7 CONCLUSION AND OUTLOOK

The purpose of the dissertation was to investigate the longitudinal development of English writing in undergraduate multilingual students in the U.S.A. in relation to language skills (i.e., vocabulary and reading), cognitive skills (i.e., attention, working memory, and general knowledge), and language features (i.e., lexical sophistication and language bursts). A total of 77 multilingual undergraduate students participated with intervals of at least five months. In this section, a summary of the results is first presented followed by overall discussion. Then, implications are presented followed by limitations and future research directions.

7.1 Summary of the Results

The first research question asked how initial levels of general cognitive/language resources and years of English immersion instruction predicted the initial level of English writing scores and changes in English writing scores in multilingual undergraduate students. Results indicated that higher initial levels of English writing scores in multilingual undergraduate students were predicted by higher levels of English vocabulary knowledge, general knowledge, and English reading skills. In addition, greater gains in English writing scores in multilingual students were predicted by higher levels of working memory capacity, but lower levels of general knowledge and lower levels of English reading skills. However, working memory capacity did not predict English scores cross-sectionally. Also, attentional capacity as measured by the Stroop test and years of English immersion instruction did not predict either initial writing scores or writing score changes. Intervals also did not predict writing score changes.

The second research question focused on the longitudinal relationships among English writing score, burst length, and academic word use in multilingual undergraduate students. Cross-sectionally, higher English writing scores were related to greater percentages of academic words in essays. Longitudinally, initial levels of writing scores were negatively associated with writing score changes and academic word percentage changes, but positively linked to burst length changes. In addition, gains in writing scores were linked to gains in academic word percentages.

The third research question examined the longitudinal relationship among English writing, English reading, and English vocabulary in multilingual undergraduate students, and the existence of a latent variable of English literacy that was informed by English writing, reading, and vocabulary over time. Results indicated that cross-sectionally, English writing, reading, and vocabulary scores were positively related to each other. Longitudinally, initial English writing scores were negatively related to writing score changes and reading score changes, while initial reading scores were negatively related to both writing score changes and reading score changes. In addition, gains in writing scores were related to gains in reading scores. Furthermore, vocabulary scores did not show any longitudinal relationship with reading or writing scores. Lastly, the construction of the *English Literacy* latent variable that was informed by English writing, reading, and vocabulary was supported over time (with the exception of writing score intercepts), though no gains in the *English Literacy* latent variable mean scores were found over time.

Table 7.1 shows a summary of overall results across the three research questions, which includes variables that showed significant relations with initial writing scores and writing score changes. Initial writing scores were positively linked to initial vocabulary scores, initial general

knowledge scores, initial reading scores, initial academic word percentages, and burst length changes, but negatively linked to academic word percentage changes and reading score changes. On the other hand, writing score changes were positively related to initial working memory test scores, academic word percentage changes, and reading score changes, but negatively related to initial general knowledge scores, initial reading scores, initial writing scores, and initial academic word percentages.

Table 7.1 Summary of the Results Across the Three Research Questions

Relation	Initial writing scores	Writing score change
Positive relation	Initial vocabulary scores Initial general knowledge scores Initial reading scores Initial academic word percentages Burst length changes	Initial working memory test scores Academic word percentage changes Reading score changes
Negative relation	Academic word percentage changes Reading score changes	Initial general knowledge scores Initial reading scores Initial writing scores Initial academic word percentages

7.2 Overall Discussion

Based on the results across the three main research questions, eight main findings are discussed. First of all, multilingual students' gains in writing scores tended to rise as a function of lower initial levels of English writing scores, English reading scores, general knowledge scores, and academic word percentages found in essays. That is, greater gains in writing scores were related to multilingual students who received lower writing scores, lower reading scores, and lower general knowledge scores and produced lower percentages of academic words at the initial time of measurement. This finding may extend L1 reading research that has supported the notion that children with initially lower reading skills may show faster growth in reading, while those with initially higher skills may grow slowly (Parrila, Aunola, Leskinen, Nurmi, & Kirby, 2005; Pfost, Hattie, Dörfler, & Artelt, 2014) into multilingual writing research at the college

level. That is, students with poorer initial performance level tend to show faster growth than students that start higher. Also, this finding supports the power law of practice which suggests that learning increases at a constant rate depending on how much it is left to be learned (Kellogg & Whiteford, 2009). In other words, in the context of learning to write persuasive essays in English at the college level, multilingual students who show initial lower levels in English reading and writing, general knowledge, and the use of academic words tend to show greater gains in writing scores because much is left to be learned. In contrast, multilingual students who show initial higher levels in English reading and writing, general knowledge, and the use of academic words are likely to show fewer gains in writing scores because little is left to be learned. Thus, this dissertation is in line with "poor get richer" and "rich remain the same" scenarios rather than "rich get richer" (i.e., the Matthew effect; Stanovich, 1986), such that initial lower levels may leave greater potential for growth in better producing persuasive essays in the process of being immersed in English academic contexts in higher education.

Second, multilingual students' gains in English writing scores co-occurred with their increases in academic words and their gains in English reading scores. This finding expands previous cross-sectional research that has reported the importance of L2 reading in L2 writing (Belcher & Hirvela, 2001; Carson et al, 1990; Pae, 2018) and the importance of the use of academic words in academic writing (Coxhead 2012; Douglas, 2013; He & Shi, 2012; Laufer, 2013). This study also expands L2 writing research in that although no group-level gains in multilingual students' use of academic words over time at the college level were reported in this study and previous studies (Knoch et al., 2014, 2015), gains in the use of academic words are found to be important in gains in English writing scores. Thus, as better English writing ability is related to better English reading ability and greater use of English academic words in essays, this

dissertation reports that greater gains in writing ability tend to go hand-in-hand with greater gains in reading ability and greater gains in the use of academic words.

Third, with respect to general cognitive resources (i.e., attention as measured by the Stroop test and working memory as measured by the running span test), findings indicated that these cognitive resources may not be important in predicting initial writing scores. This finding does not seem to corroborate the Hayes-Berninger's (2014) cognitive model of writing that emphasizes the roles of attention and working memory in writing. The lack of the roles of attention and working memory in initial English writing scores at the college level may be due to differences between writers' cognitive processes linked to attention and working memory and raters' evaluation of the essays. Specifically, attention and working memory likely relate to holding and processing information, which is not directly linked to better argumentation and better language use in persuasive essays that raters mostly focus on when scoring essays. Additionally, despite the lack of direct links between working memory and writing scores, it is worth noting that working memory capacity was significantly correlated with English vocabulary knowledge, general knowledge, and English reading skills, all of which are considered as part of *crystallized intelligence* (the knowledge learnt through education and experience; Cattell, 1943). Importantly, all of these three knowledge/skills predicted initial English writing scores. This may hint at the indirect relationship between working memory and writing scores via crystallized intelligence (i.e., English knowledge and general knowledge), such that working memory is linked to crystallized intelligence, which in turn predicts writing scores. Thus, higher working memory capacity may be indirectly linked to higher writing scores via higher levels of crystallized intelligence, though no direct link between working memory and writing scores at Time 1 was found.

Fourth, in relation to cognitive resources and writing score gains over time, working memory predicted writing score changes, such that higher levels of working memory capacity tended to relate to greater gains in writing scores. This indicates that higher working memory capacity may help students quickly learn how to coordinate writing processes including planning, sentence generation, and reviewing, which may facilitate better writing performance over time (Hayes, 2009; Kellogg, 2008).

Fifth, no role of years of English immersion years was found in predicting initial English writing scores or English writing score changes in multilingual undergraduates who included both international students and Generation 1.5 students. This finding indicates that although differences in writing between international students and Generation 1.5 students have been reported (Doolan, 2017; Levi, 2004; di Gennaro, 2009; Mikesell, 2007), no noticeable differences in initial English writing scores and writing score changes between the two groups of students were found when years of English immersion years were considered. Thus, when defining L2 learners in terms of writing ability, using a simple distinction between international students and Generation 1.5 students may not be the best approach because longer years of English immersion instruction (which relate to international students educated in the English language and Generation 1.5 students) do not imply higher levels of English writing ability in producing persuasive essays. Thus, as L1 speakers need to learn academic writing (Connerty, 2009; Hulstijn, 2015), multilingual writers also need to learn academic writing regardless of their years of English immersion instruction.

Sixth, both initial English vocabulary scores or vocabulary score changes were not related to English reading score changes or English writing score changes. This finding may contradict previous L1 reading research which has found the important role of initial vocabulary

knowledge levels in reading comprehension growth in young students (Lervåg & Aukrust, 2010; Quinn, Wagner, Petscher, & Lopez, 2015). For undergraduate multilingual students, gains in English reading and writing skills at the discourse level may occur independently of previous levels of English vocabulary knowledge. Also, at the college level, improving English reading and writing skills may not necessarily pertain to knowing more words.

Seventh, language burst lengths at Time 1 were not related to writing scores Time 1 but negatively related to burst length changes. Despite the lack of the relationship between burst length and writing scores at Time 1, it should be noted that burst length and writing scores were significantly correlated at Time 2. The lack of the relationship between burst length and writing scores at Time 1 may be explained by some students who produced excessively long burst lengths at Time 1, potentially as a result of their stream-of-conciousness writing style. Because these students were not necessarily higher writers at Time 1, these students' presence may have weakend the links between writing scores and burst lengths at Time 1. Indeed, given that a notable pattern of decreases in burst lengths over time was found in multilingual students who tended to produce excessively longer bursts, burst lengths at Time 2 with this decreasing pattern would be a more reasonable representation of burst length, which were thus significantly correlated with writing scores at Time 2. Thus, findings of this dissertation not only supports previous study that has found the importance of burst in L1 writing (Chenoweth & Hayes, 2001; Hayes & Berninger, 2014; Limpo & Alves, 2017) and L2 writing (Révész, Michel, & Lee, 2017), but also add an additional observation that some multilingual students showed a hint of improvement in writing behavior that they produced a more reasonable amount of language strings over time.

Lastly, the presence of a latent variable of English *Literacy* informed by English writing, reading, and vocabulary was supported over time. In addition, no gains in the latent mean scores were found, though writing scores showed an increase over time. Thus, while writing scores increased over time, reading and vocabulary scores remained the same. This may be because taking academic courses in higher education may facilitate students' learning of English writing skills (specifically in producing persuasive essays) but may not facilitate English reading skills and vocabulary knowledge.

7.3 Implications

7.3.1 Theoretical implications

Three are three main theoretical implications. First, the dissertation was informed by a recent model of writing (i.e., Hayes & Berninger, 2014) in selecting variables with a focus on the resource level and the process level in a multilingual context at the college level. At the resource level, cross-sectionally, English vocabulary knowledge, English reading skills, and general knowledge was linked to English writing scores, while longitudinally, working memory capacity, English reading skills, and general knowledge were related to English writing scores. Attention, however, was not related to writing scores either cross-sectionally and longitudinally. Thus, in a multilingual college-level context in which students' academic writing skills (such as persuasive essays) are important, the role of attention can be minimized. On the other hand, at the process level, the roles of the translator were measured by the use of academic words and burst lengths. Both of the use of academic words and burst lengths were found to be related to writing scores, which supports the important role of the translator in English writing in multilingual writers in a college-level context.

Second, extending previous studies on longitudinal writing development in young L2 learners (Schoonen et al., 2011), the dissertation investigated the longitudinal development of English writing in young adult multilingual learners in higher education. Overall, this dissertation provides insights that greater gains in writing scores tend to be a property more of less proficient English writers and readers, supporting the "poor gets richer" scenario.

Lastly, the dissertation sheds light on the longitudinal relationship among English writing, reading, and vocabulary knowledge in young adult multilingual learners. Findings support that a latent construct of English literacy is constructed to explain English writing, reading, and vocabulary equivalently over time with the exception of mean writing scores. This provides a parsimonious understanding of English-literacy related variables and evidence of the presence of an underlying latent trait that represents English literacy in multilingual students.

7.3.2 Pedagogical implications

Based on findings of the dissertation, two main pedagogical implications are discussed. First of all, findings suggest the importance of diagnosing multilingual undergraduates' writing skills after matriculation, so that that multilingual students who have lower levels of English writing skills can benefit from writing instructions in their earlier academic years. To do so, the first necessary step would be for institutions to assess all incoming multilingual students' writing ability after matriculation. Assessing all multilingual students, including Generation 1.5 and international students who have been educated in the English language, is important because longer lengths of English immersion instruction do not guarantee proficient English writing skills as found in this dissertation. Thus, institution-level support systems that can cater for various multilingual students' needs in academic writing would be important (Andrade, 2006; Lee, 2018).

Second, given that writing score gains are related to reading score gains and academic word percentage gains, multilingual writing classes may do well to focus not only on English writing itself, but also reading and academic vocabulary. For example, writing assignments can include reading elements. Also, explicit teaching of academic words may help students make the greater use of academic words (Laufer, 1994). In addition, considering that higher English writing scores were predicted by greater English vocabulary knowledge and greater general knowledge, multilingual students may benefit from learning more English vocabulary and having more general knowledge (including that related to English-related literature and history) for more effective use of knowledge resources stored in long-term memory (Chenoweth & Hayes, 2001).

7.4 Limitations and Future Research Directions

Interpretation of current findings is contingent on at least three main areas of limitations, which also provide future research directions. First, limitations related to measuring gains in writing scores should be noted. While the dissertation examined gains in writing scores in relation to cognitive and language resources, other factors that may have influenced writing score gains, such as students' academic experiences (e.g., courses taken), motivation to learn English writing, and engagement in English writing, were not considered. In addition, the dissertation measured gains over time without considering any interventions. Furthermore, when examining gains between variables, covariances were considered, and thus causality inferences should be avoided. Future studies would provide a more comprehensive and complete understanding of English writing development by considering students' academic experience and motivations and implementing pedagogical interventions that may facilitate learn-to-writing processes.

Second, methodological decisions in this dissertation have limitations. Two time points were included only. Also, when assessing skills and knowledge, such as writing and working

memory, a single measure (indicator) was used. Future studies may benefit from including multiple measurement points for a longer period. In addition, the sample size was relatively small, and thus future studies can include larger samples. Furthermore, this dissertation used a typical essay type of writing using SAT-based prompts along with reading and vocabulary tests at 10-12 grade levels, and these tests may not capture students' discipline-specific writing skills that students are more likely to encounter in higher education. Using discipline-specific reading and writing tasks may better capture students' writing development over time. Additionally, essays written at Time 1 were separately scored from essays written at Time 2, which means that the raters were aware of when the essays were written. Also, the potential interaction effect between prompts and academic words was not considered. Potentially, prompts may lead to the different use of academic words (Lavallée & McDonough, 2015). Future studies examining whether and how different prompts lead writers to use specific types of academic words would merit consideration. Lastly, the general knowledge test used in this dissertation may reflect a Euro-American cultural knowledge test because many test items were related to American and European history and literature. Perhaps, general knowledge tests may need to include culture-neutral items mainly related to science.

Lastly, this dissertation has limitations in investigating writing processes. The dissertation did not consider working memory capacity and attention in relation to writing process behaviors (e.g., pause durations). There might be a relationship of writing processes with working memory and attention because writers who can store a larger amount of information in working memory and are more able to maintain focus may fluently produce longer stretches of ideas without longer pause durations. Also, typing speed which may influence burst lengths was not considered. Furthermore, burst lengths were considered in terms of bursts interrupted by pauses,

but not those interrupted by revisions. Future studies may conduct a more fine-grained analysis of writing processes by examining the relationship of working memory and attention with micro-level writing processes, the effects of typing speed on burst length and writing scores, and bursts interrupted by revisions.

REFERENCES

Abbott, R. D., & Berninger, V. W. (1993). Structural equation modeling of relationships among developmental skills and writing skills in primary- and intermediate-grade writers. *Journal of Educational Psychology, 85*, 478–508.

Acheson, D., Wells, J., & MacDonald, M. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods, 40*, 278–289.

Alderson, J. C., & Kremmel, B. (2013). Re-examining the content validation of a grammar test: The (im)possibility of distinguishing vocabulary and structural knowledge. *Language Testing, 30*(4), 535–556.

Allen, L. K., Snow, E. L., & McNamara, D. S. (2016). The narrative waltz: The role of flexibility on writing performance. *Journal of Educational Psychology, 108*, 911–924.

Allen, L. K., Snow, E. L., Crossley, S. A., Jackson, G. T., & McNamara, D. S. (2014). Reading comprehension components and their relation to writing. *L'Annee Psychologique, 114*(4), 663–691.

Alves, R. A., & Limpo, T. (2015). Progress in written language bursts, pauses, transcription, and written composition across schooling. *Scientific Studies of Reading, 19*, 374-391.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369–406

Andrade, M. S. (2006). International students in English-speaking universities: Adjustment factors. *Journal of Research in International education, 5*(2), 131–154.

Andrade, M. S. (2009). The effects of English language proficiency on adjustment to university life. *International Multilingual Research Journal, 3*(1), 16–34.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language, 59*(4), 390–412.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417–423.

Baird, A. D., & Babb, K. A. (2014). First language, reading skills in the language of instruction, and first-semester grades in a Canadian university. *Journal of The First-Year Experience & Students in Transition, 26*(1), 63–83.

Berninger, V. (2000). Development of language by hand and its connections to language by ear, mouth, and eye. *Topics of Language Disorders, 20*, 65–84.

Berninger, V., Vaughan, K., Abbott, R. D., Begay, K., Coleman, K. B., Curtin, G., et al. (2002). Teaching spelling and composition alone and together: Implications for the simple view of writing. *Journal of Educational Psychology, 94*, 291–304

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly, 45*(1), 5–35.

Broadway, J. M., & Engle, R. W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods, 42*(2), 563–570.

Buhi, E. R., Goodson, P., & Neilands, T. B. (2008). Out of sight, not out of mind: Strategies for handling missing data. *American Journal of Health Behavior, 32*(1), 83–92.

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing, 26*, 42–65.

Canagarajah, A. S. (2002). Multilingual writers and the academic community: towards a critical relationship. *Journal of English for Academic Purposes, 1*, 29–44.

Carretti, B., Borella, E., & De Beni, R. (2007). Does strategic memory training improve the working memory performance of younger and older adults?. *Experimental psychology, 54*(4), 311–320.

Carson, J. E., Carrell, P. L., Silberstein, S., Kroll, B., & Kuehn, P. (1990). Reading-writing relationships in first and second language. *TESOL Quarterly, 24*(2), 245–266.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40,* 153–193.

Chen, H. C., & Graves, M. F. (1995). Effects of previewing and providing background knowledge on Taiwanese college students' comprehension of American short stories. *TESOL Quarterly, 29,* 663–686.

Chen, H. C. & Ho, C. (1986). Development of Stroop interference in Chinese-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 397–401.

Chenoweth, N., & Hayes, J. R. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication, 18*, 80–98.

Clark, M. R. (2005). Negotiating the freshman year: Challenges and strategies among first-year college students. *Journal of College Student Development, 46*(3), 296–316.

Collier, V. P. (1987). Age and rate of acquisition of second language for academic purposes. *TESOL quarterly, 21*(4), 617–641.

Connelly, V., Dockrell, J. E., Walter, K., & Critten, S. (2012). Predicting the quality of composition and written language bursts from oral language, spelling, and handwriting skills in children with and without specific language impairment. *Written Communication, 29*, 278– Re: Pairwise score analysis 302.

Conway, K. M. (2010). Educational aspirations in an urban community college: Differences between immigrant and native student groups. *Community College Review, 37*(3), 209–242.

Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 99*(2), 311–325.

Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading, 35*(2), 115–135.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing, 32*, 1–16.

Crossley, S. a, Yang, H. S., & Mcnamara, D. S. (2014). What's so simple about simplified texts? A computational and psycholinguistic investigation of text comprehension and text processing. *Reading in a Foreign Language, 26*(1), 92–113.

Cumming, A. (2001). Learning to write in a second language: Two decades of research. *International Journal of English Studies, 1*(2), 1–23.

Daller, H., Milton, J. and Treffers-Daller, J. (2007) Editors' introduction: Conventions, terminology and an overview of the book. In H. Daller, J. Milton and J. Treffers-Daller (eds) *Modelling and Assessing Vocabulary Knowledge* (pp. 1–32). Cambridge: Cambridge University Press.

Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*, 422–433.

Dansac, C., & Alamargot, D. (1999). Accessing referential information during text composition: When and why? In M. Torrance, & D. Galbraith (Eds). *Knowing what to write: Conceptual processes in text production* (pp. 79–97). Amsterdam: Amsterdam University Press

DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 97–113). Mahwah, NJ: Lawrence Erlbaum.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22*(4), 499–533.

di Gennaro, K. (2009). Investigating differences in the writing performance of international and generation 1.5 students. *Language Testing, 26*, 533–559.

di Gennaro, K. (2013). How different are they? A comparison of Generation 1.5 and international L2 learners' writing ability. *Assessing Writing, 18*(2), 154–172.

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development, 43*, 121–149.

Doolan, S. M. (2017). Comparing patterns of error in generation 1.5, L1, and L2 FYC writing. *Journal of Second Language Writing, 35*, 1–17.

Doolan, S. M., &Miller, D. (2012). Generation 1.5 written error patterns: A comparative study. *Journal of Second Language Writing, 21*,1–22.

Eckstein, G., & Ferris, D. (2018). Comparing L1 and L2 Texts and writers in first-year composition. *TESOL Quarterly, 52*(1), 137–162.

Ellis, N. (1997) Vocabulary acquisition: Word structure, collocation, word-class, and meaning. In N. Schmitt and M. McCarthy (eds) *Vocabulary: Description Acquisition and Pedagogy* (pp. 122-139). Cambridge: Cambridge University Press.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19–23.

Evans, N. W., Anderson, N. J. & Eggington, W. G. (Eds). (2015). *ESL readers and writers in higher education: Understanding challenges, providing support*. New York, NY: Routledge.

Faigley, L. (1986, October). Competing theories of process: A critique and a proposal. *College English, 48*, 527–542.

Finnie, R., & Qiu, H. (2008). *The patterns of persistence in post-secondary education in Canada* (MESA Project research paper, Educational Policy Institute). Retrieved from MESA Project website: http://mesa-project.org/pub/pdf /MESA_Finnie_Qiu_2008Aug12. pdf

Fox, H. (1994). *Listening to the World: Cultural issues in academic writing*. National Council of Teachers of English, Urbana, IL.

Galloway, F. J., & Jenkins, J. R. (2009). The adjustment problems faced by international students in the United States: A comparison of international students and administrative perceptions at two private, religiously affiliated universities. *NASPA Journal, 46*(4), 661–673.

Gardner, S., & Nesi, H. (2013). A classification of genre families in university student writing. *Applied Linguistics, 34*(1), 25–52.

Gass, S. & J. Lee. (2011). Working memory capacity, inhibitory control, and proficiency in a second language. In M. Schmid & W. Lowie (Eds.), *Modeling bilingualism: From structure to chaos*, (pp. 59–84). Amsterdam: John Benjamins.

Ghisletta, P., & McArdle, J. J. (2012). Latent curve models and latent change score models estimated in R. *Structural equation modeling: a multidisciplinary journal, 19*(4), 651–682.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Golden, C., & Freshwater, S. (2002). *The Stroop color and word test: A manual for clinical and experimental uses.* Wood Dale, IL: Stoelting Co.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6–10.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. New York, NY: Cambridge University Press.

Grabe, W., & Kaplan, R. B. (1996). *Theory & practice of writing: An applied linguistic perspective*. London: Longman.

Grabe, W., & Stoller, F.L. (2011). *Teaching and researching reading* (2nd Ed.). New York, NY: Routledge.

Graham, S. & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*, 445–476.

Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing, 9*, 123–145.

Grimm, K. J., An, Y., McArdle, J. J., Zonderman, A. B., & Resnick, S. M. (2012). Recent changes leading to subsequent changes: Extensions of multivariate latent difference score models. *Structural equation modeling: a multidisciplinary journal, 19*(2), 268–292.

Guo, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing, 18*(3), 218–238.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th Ed.). Upper Saddle River, NJ: Prentice Hall.

Halliday, M. A. K.., & Hasan, R. (1976). *Cohesion in English*. London: Longman Group.

Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. *Cognitive psychology, 44*(4), 339–387.

Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal of Educational Psychology, 93*, 103–128.

Harklau, L. (2000). From the "good kids" to the "worst": Representations of English language learners across educational settings. *TESOL quarterly, 34*(1), 35–67.

Haswell, R. H. (2000). Documenting Improvement in College Writing: A Longitudinal Approach. *Written Communication, 17*(3), 307–352.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 1–27). Mahwah, NJ: Erlbaum.

Hayes, J. R. (2009). From idea to text. In R. Beard, D. Myhill, J. Riley, & M. Nystrand (Eds.), The *SAGE handbook of writing development* (pp. 65–79). London, UK: Sage

Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication, 29*(3), 369–388.

Hayes, J. R., & Berninger, V. W. (2014). Cognitive processes in writing: A framework. In B. Arfe, J., Dockrell, & V. W. Berninger (Eds), *Writing development in children with hearing loss, dyslexia, or oral language problems* (pp. 3–15). Oxford: Oxford University Press.

Hayes, J. R., & Chenoweth, N. A. (2006). Is working memory involved in the transcribing and editing of texts? *Written Communication, 23*(2), 135–149.

Hayes, J. R., & Chenoweth, N. A. (2007). Working memory in an editing task. *Written Communication, 24*(4), 283–294.

Hayes, J., & Flower, L. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3–30). Hillsdale, NJ: Erlbaum.

He, L., & Shi, L. (2012). Topical knowledge and ESL writing. *Language Testing, 29*, 443–464.

Holmes, P. (2004). Negotiating differences in learning and intercultural communication: Ethnic Chinese students in a New Zealand university. *Business Communication Quarterly, 67*(3), 294–307.

Holten, C. (2009). Creating an inter-departmental course for generation 1.5 ESL writers: Challenges faced and lessons learned. In: M. Roberge, M. Siegal, & L. Harklau (Eds.), *Generation 1.5 in college composition: Teaching academic writing to U.S.-educated learners of ESL* (pp. 170–184). New York: Routledge.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2,* 127–160.

Housen, A., Kuiken, F., & Vedder, I. (Eds.), (2012). *Dimensions of L2 performance and proficiency: complexity, accuracy and fluency in SLA.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55.

Institute of International Education (2016). *Open Doors® 2016 report on international educational exchange.* Retrieved from https://p.widencdn.net/p8fxny/Open-Doors-2016-Presentation.

Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing, 19*(1), 57–84.

Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning, 63*(SUPPL. 1), 87–106.

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning, 64*(1), 160–212.

Johnstone, K. M., Ashbaugh, H., & Warfield, T. D. (2002). Effects of repeated practice and contextual-writing experiences on college students' writing skills. *Journal of Educational Psychology, 94(*2), 305–315.

Kaplan, D., Kim, J.-S., & Kim, S.-Y. (2009). Multilevel latent variable modeling: Current research and recent developments. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The Sage handbook of quantitative methods in psychology* (pp. 592–612). Thousand Oaks, CA: Sage.

Kaufer, D. S., Hayes, J. R., & Flower, L. (1986). Composing written sentences. *Research in the Teaching of English, 20, 121–140.*

Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research, 1(*1), 1–26.

Kellogg, R. T., & Whiteford, A. P. (2009). Training advanced writing skills: The case for deliberate practice. *Educational Psychologist, 44*(4), 250–266.

Kim, Y., Payant, C., & Pearson, P. (2015). The intersection of task-based interaction, task complexity, and working memory: L2 question development through recasts in a laboratory setting. *Studies in Second Language Acquisition. 37*, 549–581.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, England: Cambridge University Press.

Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. J. Snowling & C. Hulme (Eds.), The *science of reading: A handbook* (pp. 209–226). Oxford: Wiley-Blackwell.

Knoch, U., Rouhshad, A., Oon, S. P., & Storch, N. (2015). What happens to ESL students' writing after three years of study at an English medium university? *Journal of Second Language Writing, 28*, 39–52.

Kobrin, J. L., Deng, H., & Shaw, E. J. (2007). Does quantity equal quality? The relationship between length of response and scores on the SAT essay. *Journal of Applied Testing Technology, 8*, 1–15.

Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). *Validity of the SAT for predicting first-year college grade point average* (College Board Research Report No. 2008-5). Retrieved from http://professionals.collegeboard.com/data-reports-research/ sat/validity-studies

Koda, K. (2005). *Insights into second language reading*. New York, NY: Cambridge University Press.

Koda, K. (2007). Reading and language learning: Crosslinguistic constraints on second language reading development. *Language Learning, 57*(SUPPL. 1), 1–44.

Kormos, J. (1999). Monitoring and self repair in L2. *Language Learning, 49*, 303–342.

Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing, 21*(4), 390–403.

Kormos, J., & Sáfár, A. (2008). Phonological short term-memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition, 11*, 261–271.

Kramsch, C. (2009). *The multilingual subject: what foreign language learners say about their experience and why it matters*. Oxford. UK: Oxford University Press.

Kyle, K., & Crossley, S. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing, 34*, 12–24.

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly, 49*(4), 757–786.

Landi, N. (2005). Behavioral and electrophysiological investigations of the relationship between semantic processing and reading comprehension. *Brain and Language, 102,* 30–45.

Larsen-Freeman, D. (1991). Teaching grammar. In M. Celce-Murcia (ed.), *Teaching English as a Second or Foreign Language* (pp. 279–296). Boston, MA: Heinle and Heinle Publishers.

Laufer, B. (1994). The Lexical Profile of Second Language Writing: Does It Change Over Time? *RELC Journal, 25*(2), 21–33.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics, 16*(3), 307–322.

Lawrick, E., & Esseili, F. (2015) Familiar strangers: International students in the U.S. composition course. In N. W. Evans, N. J. Anderson, & W. G. Eggington. (Eds.). *ESL readers and writers in higher education: Understanding challenges, providing support* (pp. 80–94). New York, NY: Routledge.

Lavallée, M., & McDonough, K. (2015). Comparing the lexical features of EAP students' essays by prompt and rating. *TESL Canada Journal, 32*(2), 30-44.

Lee, S. (2018). Frameworks for failure in L2 writing: What we can learn from "failures" of Chinese international students in the US. *Journal of Second Language Writing, 41*, 98–105.

Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using inputlog to analyze and visualize writing processes. *Written Communication, 30*(3), 358–392.

Leki, I., Cumming, A., & Silva, T. (2008). *A synthesis of research on second language writing in English.* London: Routledge.

Levi, E. I. (2004). *A study of linguistic and rhetorical features in the writing of non-English language background graduates of U.S. schools.* Pennsylvania: University of Pennsylvania [Unpublished doctoral dissertation]

Lewis, M. (1993). *The lexical approach: the state of ELT and a way forward*. London: Language Teaching Publications.

Limpo, T., & Alves, R. A. (2017). Written language bursts mediate the relationship between transcription skills and writing performance. *Written Communication, 34*(3), 306–332.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review, 21*(4), 861–883.

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford Press.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15*(4), 474–496.

Lu, Y. (2010). *Cognitive factors contributing to Chinese EFL learners' L2 writing performance in timed essay writing*. Doctoral Dissertation, Georgia State University, USA.

Ludwig, C., Borella, E., Tettamanti, M., & de Ribaupierre, A. (2010). Adult age differences in the Color Stroop Test: A comparison between an Item-by-item and a Blocked version. *Archives of Gerontology and Geriatrics, 51*(2), 135–142.

MacArthur, Charles A., Graham, Steve, & Fitzgerald, John (Eds.). (2008). *Handbook of writing research*. New York, NY: The Guilford Press.

MacDonald, S. (1994). *Professional academic writing in the humanities and social sciences*. Carbondale: Southern Illinois University Press."

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2002). *Gates-MacGinitie Reading Tests, fourth edition technical report for Forms S & T.* Itasca, IL: Riverside.

MacLeod, C. M. (1992). The Stroop task: The 'gold standard' of attentional measures. *Journal of Experimental Psychology General, 121*, 12–14

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual review of psychology, 60,* 577–605.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14,* 1–43.

McNamara, T. F. (1996). *Measuring second language performance*. London/New York: Longman.

Miele, C. (2003). Bergen community college meets Generation 1.5. Community College. *Journal of Research & Practice, 27*(7), 603–612.

Milton, J. (2009). *Measuring second language vocabulary acquisition* (Vol. 45). Multilingual Matters.

Mrazek, M. D., Franklin, M. S., Phillips, D. T., Baird, B., & Schooler, J. W. (2013). Mindfulness training improves working memory capacity and GRE performance while reducing mind wandering. *Psychological science, 24*(5), 776–781.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*(2), 133–142.

Nassaji, H. (2007). Schema theory and knowledge-based processes in second language reading comprehension: A need for alternative perspectives. *Language Learning, 57*, 79–113.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*(4), 555–578.

Ortega, L. (2015). Syntactic complexity in L2 writing: Progress and expansion. Journal of *Second Language Writing, 29*, 82–94.

Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2015). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods, 47*(4), 1343–1355.

Ozuru, Y., Rowe, M., O'Reilly, T., & McNamara, D. S. (2008). Where's the difficulty in standardized reading tests: the passage or the question? Be*havior Research Methods, 40*(4), 1001–1015.

Pae, T. (2018a). Effects of task type and L2 proficiency on the relationship between L1 and L2 in reading and writing: An SEM approach. *Studies in Second Language Acquisition, 40*, 63–90.

Parrila, R., Aunola, K., Leskinen, E., Nurmi, J. E., & Kirby, J. R. (2005). Development of individual differences in reading: Results from longitudinal studies in English and Finnish. *Journal of Educational Psychology, 97*(3), 299–319

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357–383.

Perfetti, C., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Oxford, UK: Blackwell.

Pfost, M., Hattie, J., Dörfler, T., & Artelt, C. (2014). Individual differences in reading development: A review of 25 years of empirical research on Matthew effects in reading. *Review of Educational Research, 84*(2), 203–244.

Pienemann, M. (1998) *Language Processing and Second Language Development: Processability Theory*. Amsterdam: John Benjamins.

Purpura, J. E. (2004). *Assessing grammar*. Cambridge, UK: Cambridge University Press.

Quinn, J. M., Wagner, R. K., Petscher, Y., & Lopez, D. (2015). Developmental relations between vocabulary knowledge and reading comprehension: A latent change score modeling study. *Child development, 86(*1), 159–175.

R Core Team (2018). R*: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ramburuth, P. (2001). Language diversity and the first-year experience: Implications for academic achievement and language skills acquisition. *Journal of the First-Year Experience, 13*(2), 75–93.

Ramsay, S., Barker, M., & Jones, E. (2006). Academic adjustment and learning processes: A comparison of international and local students in first-year university. *Higher Education Research & Development, 18*(1), 129–144.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Révész, A., Michel, M., & Lee, M. (2017). *Investigating IELTS Academic Writing Task 2: Relationships between cognitive writing processes, text quality, and working memory*. IELTS Research Reports Online Series, 44.

Robertson, M., Line, M., Jones, S., & Thomas, S. (2000). International students, learning environments and perceptions: A case study using the Delphi technique. Higher *Education Research & Development, 19*(1), 89–102.

Robinson, P. (2003). Attention and memory during SLA. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 631-678). Oxford: Blackwell.

Römer, U. (2009). The inseparability of lexis and grammar: corpus linguistic perspectives. *Annual Review of Cognitive Linguistics, 7*, 141–163.

Roscoe, R. D., Crossley, S. A., Snow, E. L., Varner, L. K., & McNamara, D. S. (2014). Writing quality, knowledge, and comprehension correlates of human and automated essay scoring. In W. Eberle & C. Boonthum-Denecke (Eds.), *Proceedings of the 27th International Florida Artificial Intelligence Research* Society (FLAIRS) Conference (pp. 393–398). Palo Alto, CA: AAAI Press.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2), 1-36.

Roy, A., Kefi, M. Z., Bellaj, T., Fournet, N., Le Gall, D., & Roulin, J. L. (2018). The Stroop test: A developmental study in a French children sample aged 7 to 12 years. *Psychologie Francaise, 63*(2), 129–143.

Sanders, A. (1998). *Elements of Human Performance.* Mahwah, NJ: Lawrence Erlbaum Associates.

Schaie, K. W. (1996). *Intellectual development in adulthood. The Seattle longitudinal study*. New York: Cambridge University Press.

Schmidt, R. W. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1–63). Honolulu: University of Hawai'i Press"

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime User's Guide*. Pittsburgh: Pschology Software Tools, Inc.

Schoonen, R., Van Gelderen, A., De Glopper, K., Hulstijn, J., Simis, A., Snellings, P. & Stevenson, M. (2003). First language and second language writing: The role of linguistic fluency, linguistic knowledge and metacognitive knowledge. *Language Learning, 53*(1), 165–202

Schoonen, R., Van Gelderen, A., Stoel, R. D., Hulstijn, J., & De Glopper, K. (2011). Modeling the development of L1 and EFL writing proficiency of secondary school students. *Language Learning, 61*(1), 31–79.

Service, E. & Kohonen, V. (1995). Is the relation between phonological memory and foreign-language learning accounted for by vocabulary acquisition. *Applied Psycholinguistics, 16*, 155–172.

Sherry, M., Thomas, P., & Chui, W. H. (2010). International students: A vulnerable student population. *Higher Education, 60*(1), 33–46.

Shiotsu, T., & Weir, C. J. (2007). The relative significance of syntactic knowledge and vocabulary breadth in the prediction of reading comprehension test performance. *Language Testing, 24*, 99–128.

Shum, K. K. man, Ho, C. S. H., Siegel, L. S., & Au, T. K. Fong. (2016). First-language longitudinal predictors of second-language literacy in young L2 learners. *Reading Research Quarterly, 51*(3), 323–344.

Silva, T., & Brice, C. (2004). Research in teaching writing. *Annual Review of Applied Linguistics, 24*, 70-106.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal, 36*(2), 139–152.

Storch, N., & K. Hill. (2008). What happens to international students' English after one semester at university? *Australian Review of Applied Linguistics, 31*, 1–17.

Stoynoff, S. (1997). Factors associated with international students' academic achievement. *Journal of instructional psychology, 24*(1), 56.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology, 18*(6), 643–662.

Sumiya, H. & Healy, A. F. (2004). Phonology in the bilingual Stroop effect. *Memory & Cognition, 32,* 752–758

Tabachnick, B. G., & Fidell, L. S. (2012). *Using multivariate statistics* (6th Ed.). Boston: Allyn & Bacon.

Tang, R. (Ed.). (2012). *Academic writing in a second or foreign language: Issues and challenges facing ESL/EFL academic writers in higher education contexts*. London: Bloomsbury Academic.

Teaching English to Speakers of Other Languages, Inc. (TESOL) (2010). *Position statement on the acquisition of academic proficiency in English at the postsecondary level.* Washington, DC. Retrieved January 11, 2018, from https://www.tesol.org/about-tesol/press-room/position-statements/higher-education-position-statements

Terui, S. (2012). Second language learners' coping strategy in conversations with native speakers. *Journal of International Students, 2*(2), 168–183.

Thonus, T. (2003). Serving Generation 1.5 learners in the university writing center. *TESOL Journal, 12(*1), 17–24.

Uccelli, P., Galloway, E. P., Barr, C. D., Meneses, A., & Dobbs, C. L. (2015). Beyond vocabulary: Exploring cross-disciplinary academic-language proficiency and its association with reading comprehension. *Reading Research Quarterly, 50*(3), 337–356.

Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition, 131*(3), 373–403.

van Gelderen, A., Schoonen, R., de Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2004). Linguistic knowledge, processing speed, and metacognitive knowledge in first- and second-language reading comprehension: A componential analysis. *Journal of Educational Psychology, 96*(1), 19–30.

Van Gelderen, A., Schoonen, R., de Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2004). Linguistic knowledge, processing speed, and metacognitive knowledge in first- and second-language reading comprehension: A componential analysis. *Journal of Educational Psychology, 96*(1), 19–30.

Van Gelderen, A., Schoonen, R., Stoel, R. D., de Glopper, K., & Hulstijn, J. (2007). Development of adolescent reading comprehension in language 1 and language 2: A longitudinal analysis of constituent components. *Journal of Educational Psychology, 99*(3), 477–491.

Van Waes, L., & Leijten, M. (2015). Fluency in writing: A multidimensional perspective on writing fluency applied to L1 and L2. *Computers and Composition, 38,* 79–95.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–69.

Wan, G. (2001). The learning experience of Chinese students in American universities: A cross-cultural perspective. *College Student Journal, 35*(1), 28–44.

Weigle, S. C. (2002). *Assessing writing*. New York, NY: Cambridge University Press.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. *Ariel, 165*, 220–314.

Wilkins, D. A. (1972) *Linguistics in language teaching*. London: Arnold.

Williams, J. N. (2011). Working memory and SLA. *The Routledge Handbook of Second Language Acquisition*, 427–441.

Williams, J. N. & Lovatt, P. (2003). Phonological memory and rule learning. *Language Learning, 53*,67–121.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. (1998). *Second language development in writing: Measures of fluency, accuracy and complexity*. Honolulu, HI: Second Language Teaching & Curriculum Center, University of Hawai'i.

Wu, H., Garza, E., & Guzman, N. (2015). International student's challenge and adjustment to college. *Education Research International*, 1–9.

Yamashita, J., & Shiotsu, T. (2017). Comprehension and knowledge components that predict L2 reading: A latent-trait approach. *Applied Linguistics, 38(*1), 43–67.

Yang, W., Lu, X., & Weigle, S. C. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing, 28*, 53–67.

Yeh, C. J., & Inose, M. (2003). International students' reported English fluency, social support satisfaction, and social connectedness as predictors of acculturative stress. *Counselling Psychology Quarterly, 16*(1), 15–28.

Yeh, C., & Inose, M. (2003). International students' reported English fluency, social support satisfaction, and social connectedness as predictors of acculturative stress. *Counselling Psychology Quarterly, 16,* 15–28.

Yuill, N., & Oakhill, J. (1991). *Children problems text comprehension*. Cambridge: Cambridge University Press.

Zhang, D. (2012). Vocabulary and grammar knowledge in second language reading comprehension: A structural equation modeling study. The *Modern Language Journal, 96*(4), 558–575.

Zhang, D. (2012). Vocabulary and grammar knowledge in second language reading comprehension: A structural equation modeling Study. *The Modern Language Journal, 96(*4), 558–575.

Zhang, Y. & Mi, Y. (2010). Another look at the language difficulties of international students. *Journal of Studies in International Education, 14*(4), 371–388.

APPENDIX A: BACKGROUND SURVEY

Demographic information

(a) Major
(b) Age
(c) First language
(d) Country of citizenship
(e) Gender
(f) Academic year

English language learning backgrounds

(a) How old were you when you started to learn English?
(b) How many years have you studied English?
(c) Have you lived in an English-speaking environment (i.e., countries whose mother tongue is English or schools whose official language is English)?

(c-1) If YES, please specify where, when, and how long you lived, and which school/institute you attended.

(d) Have you taken an English proficiency test, such as IELTS and TOEFL?

(d-1) If YES, please write down the most recent scores you received.

APPENDIX B: HOLISTIC ESSAY RATING RUBRIC

After reading each essay and completing the analytical rating form, assign a holistic score based on the rubric below. For the following evaluations you will need to use a grading scale between 1 (minimum) and 6 (maximum). As with the analytical rating form, the distance between each grade (e.g., 1-2, 3-4, 4-5) should be considered equal.

SCORE OF 6: An essay in this category demonstrates clear and consistent mastery, although it may have a few minor errors. A typical essay effectively and insightfully develops a point of view on the issue and demonstrates outstanding critical thinking, using clearly appropriate examples, reasons, and other evidence to support its position is well organized and clearly focused, demonstrating clear coherence and smooth progression of ideas exhibits skillful use of language, using a varied, accurate, and apt vocabulary demonstrates meaningful variety in sentence structure is free of most errors in grammar, usage, and mechanics.

SCORE OF 5: An essay in this category demonstrates reasonably consistent mastery, although it will have occasional errors or lapses in quality. A typical essay effectively develops a point of view on the issue and demonstrates strong critical thinking, generally using appropriate examples, reasons, and other evidence to support its position is well organized and focused, demonstrating coherence and progression of ideas exhibits facility in the use of language, using appropriate vocabulary demonstrates variety in sentence structure is generally free of most errors in grammar, usage, and mechanics.

SCORE OF 4: An essay in this category demonstrates adequate mastery, although it will have lapses in quality. A typical essay develops a point of view on the issue and demonstrates competent critical thinking, using adequate examples, reasons, and other evidence to support its position is generally organized and focused, demonstrating some coherence and progression of ideas exhibits adequate but inconsistent facility in the use of language, using generally appropriate vocabulary demonstrates some variety in sentence structure has some errors in grammar, usage, and mechanics.

SCORE OF 3: An essay in this category demonstrates developing mastery, and is marked by ONE OR MORE of the following weaknesses: develops a point of view on the issue, demonstrating some critical thinking, but may do so inconsistently or use inadequate examples, reasons, or other evidence to support its position is limited in its organization or focus, or may demonstrate some lapses in coherence or progression of ideas displays developing facility in the use of language, but sometimes uses weak vocabulary or inappropriate word choice lacks variety or demonstrates problems in sentence structure contains an accumulation of errors in grammar, usage, and mechanics.

SCORE OF 2: An essay in this category demonstrates little mastery, and is flawed by ONE OR MORE of the following weaknesses: develops a point of view on the issue that is vague or seriously limited, and demonstrates weak critical thinking, providing inappropriate or insufficient examples, reasons, or other evidence to support its position is poorly organized and/or focused, or demonstrates serious problems with coherence or progression of ideas displays very little facility in the use of language, using very limited vocabulary or incorrect word choice

demonstrates frequent problems in sentence structure contains errors in grammar, usage, and mechanics so serious that meaning is somewhat obscured.

SCORE OF 1: An essay in this category demonstrates very little or no mastery, and is severely flawed by ONE OR MORE of the following weaknesses: develops no viable point of view on the issue, or provides little or no evidence to support its position is disorganized or unfocused, resulting in a disjointed or incoherent essay displays fundamental errors in vocabulary demonstrates severe flaws in sentence structure contains pervasive errors in grammar, usage, or mechanics that persistently interfere with meaning.