

THE EFFECTS OF AGE OF IMMERSION AND WORKING MEMORY ON SECOND
LANGUAGE PROCESSING OF ISLAND CONSTRAINTS
: AN EYE-MOVEMENT STUDY

By

Sehoon Jung

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Second Language Studies—Doctor of Philosophy

2017

ABSTRACT

THE EFFECTS OF AGE OF IMMERSION AND WORKING MEMORY ON SECOND LANGUAGE PROCESSING OF ISLAND CONSTRAINTS: AN EYE-MOVEMENT STUDY

By

Sehoon Jung

One of the central questions in recent second language processing research is whether the types of parsing heuristics and linguistic resources adult L2 learners compute during online processing are qualitatively similar or different from those used by native speakers of the target language. While the current L2 processing literature provides evidence for both qualitative similarities and differences between L1 and adult L2 processing, Clahsen and Felser (2006a, 2006b, 2006c) claimed that the types of syntactic representations adult L2 learners apply during online processing are shallower and hierarchically less detailed, and adult L2 learners rely more on other types of linguistic resources available to them, such as lexical-semantic and pragmatic information. This dissertation aimed to explore these issues to provide more insight into the nature of adult L2 syntactic processing, by investigating how advanced ESL learners varying in their ages of arrival—early learners: ages of arrival between 2 and 9 years old; adult learners: ages of arrival between 18 and 31 years old—deal with relative clause island constructions while processing filler-gap dependencies online in a natural reading environment. In addition, the present study also sought to examine how individual differences in working memory capacity (WMC) influence learners' processing behaviors and use of target language grammars in real time. Twenty-eight advanced adult ESL learners with either Korean or Chinese background and 21 early ESL learners, as well as 24 native English speaker controls participated in an eye-tracking reading experiment and took two different types of automated complex working

memory span tests (operation-span & symmetry-span). Results suggested that while the early and adult ESL learners made use of active filler strategies to fill the gap as early as possible in the non-island environment, they rapidly deployed relevant syntactic knowledge of island constraints, thereby avoiding illicit filler-gap formation inside the relative clause islands from early stages of processing, as measured by first-pass reading time, first-pass regression, and regression path duration. Results also suggested that the early ESL learners and native English speaker controls were sensitive to structural cues and gap identifications at the ultimate gap, initiating filler-gap reanalysis processes from early stages of processing, as measured by first fixation duration and first-pass reading time. On the other hand, the adult ESL learners exhibited filler-gap reanalysis effects only during later stages of processing, suggesting that they were not as efficient and immediate as the early ESL learners and native English speaker controls in detecting the need for filler-gap reanalysis. Lastly, individual differences in WMC did not show any significant effect on early and adult ESL learners' processing of island constructions, in that both ESL groups successfully blocked gap postulations in the island environment, by and large, irrespective of their working memory capacities. However, it was found that different WMCs among the adult learners influenced their reading behaviors during filler-gap reanalysis at the ultimate gap, in that adult learners with higher WMC were more sensitive to gap identifications than those with lower WMC, showing more immediate filler-gap reanalysis effect from early stages of processing, as measured by first fixation duration and first-pass regression. These results suggest that early and adult ESL learners' processing of structurally complex filler-gap dependencies in the L2 is not qualitatively different from that of native English speakers.

Copyright by
SEHOON JUNG
2017

This dissertation is dedicated to my family
To my wife, Soojin Ryu, for her love, support, encouragement, and sacrifice.
To my parents, Jong-Sung Jung and Kyung-Ja Shin, and
To my parents-in-law, Hyun-Joo Ryu and Chung-Mi, Kim,
For their support and belief in me
To my grandma, Chun-Nam Jung, for her prayers.
To my sister Se-Jung and my brother Myoungwhon
For their encouragement and support.
But above all, this dissertation is dedicated to the Lord whom I love and trust

ACKNOWLEDGEMENTS

This dissertation is the outcomes of work and support of many people. I first would like to express my deepest appreciation to Professor Patti Spinner who is my teacher, advisor, and mentor. She has always supported me, taught me to see a big picture of what I am doing in my research, and helped me feel more confident in becoming a professional since my first days in East Lansing. I am deeply indebted to her for her encouragement and belief in me during my time in the Second Language Studies program. I am also grateful to all the members of my dissertation committee for their support, encouragement, and respecting my ideas.

I am thankful to Professor Susan Gass who has given me wonderful opportunities to work as an eye-tracking research assistant. The research and work experience I gained through my first four years of this assistantship certainly provided a great foundation on which I could plan, conduct, and write this dissertation study. Professor Paula Winke was my supervisor during my first two years in the eye-tracking lab as a graduate student lab coordinator. I owe her my current expertise in eye-tracking research. I really appreciate her caring attitude and support from Day 1 until the last moment in my Ph.D. Journey.

I am also thankful to Professor Aline Godfroid, Professor Susan Gass, Professor Shawn Lowen, and Professor Rod Ellis, for giving me a wonderful opportunity to work with them on the eye-tracking GJT research project, through which I was able to learn a lot and establish my research competence in the field. The enthusiasm and passions they brought to every single meeting over the two-year span was such an inspiration for me.

I also express my best wishes to all SLS colleagues who made life easier when we came together and shared our thoughts and challenges. I appreciate the good times, conversations and conference journeys I shared with Jens, Roman, and Hyung-Jo, and all the old and new members

of the SLS friends. I particularly extend my gratitude for Ayman Mohamed with whom I shared many days of so-called ‘Pseudo-research meting.’ I will always cherish our memories together and truly appreciate his friendship.

But above all, I want to thank my Lord who allowed me to meet these people above, gave me the knowledge, strength, and ability to finish my dissertation study. Blessed are those who trust in the name of the Lord.

TABLE OF CONTENTS

LIST OF TABLES	x
LIST OF FIGURES	xi
CHAPTER 1. INTRODUCTION	1
CHAPTER 2: REVIEW OF LITERATURE	14
2.1. Grammatical representations of wh-structures	14
2.1.1 Filler-gap dependency representations in English	14
2.1.2. Island Constraints: Violation of the movement constraints	15
2.1.3. Grammatical representations of wh-structures in wh-in-situ languages	17
2.1.4. Summary	18
2.2. Incremental processing: Evidence from processing of garden-path sentences	18
2.3. Processing of filler-gap dependencies: Incremental gap search processes	21
2.4. Processing of filler-gap dependencies by nonnative speakers	26
2.4.1. A review of early research on L2 processing.....	26
2.4.2. Shallow structure hypothesis	29
2.4.3. A Review of the empirical research testing the SSH.....	30
2.4.3.1. L2 processing of ambiguous RC constructions	30
2.4.3.2. L2 processing of long-distance filler-gap dependencies.....	31
2.4.3.3. L2 processing of island constraints.....	35
2.5. The effect of age of immersion (or acquisition) and critical period hypothesis.....	38
2.6. The role of working memory on L2 parsing	41
2.7. Research Questions	44
CHAPTER 3: METHOD	47
3.1. Participants	47
3.2. Materials	50
3.2.1. English proficiency measures	50
3.2.2. Working memory capacity measures.....	53
3.2.3. Main experiment: Eye-tracking reading	58
3.2.3.1. Reading materials.....	58
3.2.3.2. Areas of Interest for analyses.....	60
3.2.3.3. Eye-tracking reading task design and procedures.....	63
3.2.3.4. Eye-tracking dependent variables	65
3.3. Overall procedures	68
3.4. Data Analysis	68
3.4.1. Preparation of the data for analyses	68
3.4.2. Main Statistical analyses.....	71
CHAPTER 4. RESULTS	75
4.1. Comprehension Accuracy	75
4.2. Overview of reading profiles.....	76

4.3. The effect of age of immersion on L2 processing of filler-gap dependencies	81
4.3.1. Active filler strategy and application of island constraints: Initial gap	81
4.3.1.1. Analysis of reading patterns at the first critical region (Region1).....	81
4.3.1.2. Analysis of reading patterns at the spillover region (Region2)	89
4.3.1.3. Interim summary of the results— Initial gap.....	97
4.3.2. Filler-gap reanalysis: Ultimate gap	98
4.3.2.1. Analysis of reading patterns at the second critical region (Region3)	98
4.3.2.2. Analysis of reading patterns at the spillover region (Region4)	108
4.3.2.3. Interim summary of the results— Ultimate gap	117
4.4. The effect of individual differences in working memory capacity	118
4.4.1. The effect of WMC at the earliest gap Region1 and spillover Region2.....	118
4.4.2. The effect of WMC at the ultimate gap at Region3 and spillover Region4	128
4.4.3. Summary of the results— the effect of WM.....	135
CHAPTER 5: DISCUSSION.....	136
5.1. The effect of age of acquisition.....	137
5.2. The role of working memory in L2 processing of island constraints.....	152
CHAPTER 6: CONCLUSION	157
6.1. Limitations and future research.....	159
APPENDICES	161
Appendix A. Language background questionnaire	162
Appendix B. List of test items in the LexTALE English proficiency measure.....	166
Appendix C. Materials for the eye-tracking experiment.....	167
REFERENCES	170

LIST OF TABLES

Table 1. Biodata and English learning background of the ESL learners	48
Table 2. Self-rated English proficiency of the ESL learners for each language skill	50
Table 3. LexTALE scores (in percent) of the native speakers and the ESL learners	52
Table 4. Summary of the WM span test results in percent	71
Table 5. Mean comprehension accuracy in percent in the reading task	75
Table 6. Descriptive statistics for RTs in and first-pass regressions in percent at Region1	81
Table 7. Summary of the results of preliminary analyses at Region1	83
Table 8. Descriptive statistics for RTs and first-pass regressions at Region2	89
Table 9. Summary of the results of preliminary analyses at Region2	91
Table 10. Summary of the major findings at the initial gap	97
Table 11. Descriptive statistics for RTs and first-pass regression at Region3.....	98
Table 12. Summary of the results of preliminary analyses at Region3	103
Table 13. Descriptive statistics for RTs and first-pass regressions at Region4	108
Table 14. Summary of the results of preliminary analyses at Region4	110
Table 15. Summary of the findings at the ultimate gap	117
Table 16. Summary of the WM effect analyses at Region1 and Region2	120
Table 17. First-pass RT and first-pass regressions by higher- and lower-WM early ESL	122
Table 18. Summary of the WM effect analyses at Region3 and Region4	129
Table 19. Summary of the findings— The WM effect	135

LIST OF FIGURES

Figure 1. A screenshot of the LexTALE Test.....	52
Figure 2. Processing and storage component of the operation span test	56
Figure 3. Processing and storage component of the symmetry span test.....	56
Figure 4. An illustration of eye-movements during reading.....	65
Figure 5. Fixation map: Reading profiles of the NS English speakers	77
Figure 6. Fixation map: Reading profiles of the early ESL learners	78
Figure 7. Fixation map: Reading profiles of the adult ESL learners	78
Figure 8. Reading patterns of the three groups during early stages of processing at Region1	85
Figure 9. Reading patterns of the three groups during late stages of processing at Region1	86
Figure 10. Reading patterns of the three groups during early stages of processing at Region2 ...	92
Figure 11. Reading patterns of the three groups during late stages of processing at Region2	93
Figure 12. Reading patterns of the three groups during early stages of processing at Region3.	101
Figure 13. Reading patterns of the three groups during late stages of processing at Region3 ...	102
Figure 14. Reading patterns of the three groups during early stages of processing at Region4.	111
Figure 15. Reading patterns of the three groups during late stages of processing at Region4 ...	112

CHAPTER 1. INTRODUCTION

How do people come to understand what others say or what they read? One might assume that the process of understanding linguistic input (e.g., written texts or utterances) is very simple and straightforward, when considering how frequently and quickly it happens even without much conscious effort. However, language comprehension works through a sequence of highly complex linguistic analyses that map linguistic input onto a variety of different types and levels of mental representations in real time during comprehension (Clahsen, 2007; Mazuka, 1998). That is, within a limited amount of time, one must deploy different components of linguistic knowledge (e.g., lexical, syntactic, semantic, pragmatic, discourse, and world knowledge) to process the input in a linguistically and contextually appropriate way. All these mapping and application processes must be computed efficiently to achieve comprehension. In light of this, comprehension requires suitable linguistic knowledge at various levels, as well as sufficient processing skills that allow the processor to carry out such demanding linguistic operations efficiently. Of the various types and levels of linguistic representations and processing, this dissertation specifically focuses on second language (L2) learners' syntactic processing during L2 comprehension.

A critical part of understanding linguistic input is creating a licit grammatical representation that can accommodate the processed input strings. In this respect, syntactic processing (or parsing¹) is mainly responsible for conducting moment by moment computations making structural inferences from word strings in the input, creating associations between and

¹ The term *parse* or *parsing* specifically refers to online applications of grammatical information, namely syntactic processing in real time, whereas *processing* is used as a more general term that covers all types of linguistic operations and interfaces between them (VanPatten & Jegerski, 2010).

among words in a sentence to structure constituents and assign syntactic categories. In doing so, it has been well attested that the parser builds a series of representations incrementally on the basis of syntactic information provided by the grammar. In this sense, parsing can be understood as a laboratory where the current state of grammar is tested. Parsing should likely end up being successful if learners' existing grammar is mature and adequate to license a representation for the target language input. On the other hand, if the deployed information is not appropriate, either because it has not been fully acquired yet or because it is impaired due to deviations from the target language norms, then parsing may not be successful, resulting in comprehension breakdown. Such parsing failures may necessitate the need for an update of the current grammar system and trigger acquisition of the representation in the long term through repeated processing practice on the structure (e.g., Dekydtspotter & Renaud, 2014; Fodor, 1998; Gregg, 2003; White, 2003; but cf. Klein, 1999).

In order for parsing to be efficient and successful, it also requires suitable and “least effort” parsing strategies attuned to specific grammatical properties of the target language (e.g., wh-movement, grammatical gender, and relative clause attachment preferences), accompanied with sufficient processing abilities that allow the parser to integrate needed grammatical information efficiently in a consistent manner (e.g., Dussias, 2003; Gibson, Pearlmutter, Canseco-Gonzalez, & Hickok, 1996; Juffs, 2005; Juffs & Harrington, 1995, 1996; Keating, 2009; Marinis, Roberts, Felser, & Clahsen, 2005; Williams, 2006; Williams, Moebius, & Kim, 2001). Note that parsing is a processing component (i.e. linguistic performance) guided by the grammatical information (i.e., linguistic competence) under pressure. In this regard, Juffs and Rodriguez (2015) analogized the grammar to *the engine at rest* while comparing parsing to *the engine in motion*, further explaining:

the grammar is the engine at rest, not driving the vehicle, but with the potential to do so.

Parsing is the engine in motion, subject to stresses and possible breakdowns allowable by the system, and driving production or comprehension in real time..... the operation of the grammar during processing may be affected by the quality of input, memory limitations, and interference from outside influenced not related to the architecture of the grammar itself. (pp. 15)

What they mean is that, even if one has acquired a relevant representation, it does not necessarily guarantee that the acquired representation will be fully utilized during comprehension. If the parser is burdened for some reason (e.g., slower processing speed, complexity of the input, and limited working memory capacity), and/or if learners' parsing strategies are not proficient enough to rapidly extract and unload the detailed syntactic information of the target language structure in real time (e.g., Juffs, 2005; Juffs & Harrington, 1996), it may then result in misanalyses of the input. Additionally, learners might be led to rely on some other alternative sources of information to compensate for the lack of parsing abilities (e.g., use of semantic information, context surrounding the input being currently processed, or world knowledge) (Clahsen & Felser, 2006a; 2006b). Taken together, as far as syntactic processing is concerned, successful comprehension requires L2 learners to acquire not only the grammar of the target language, but also processing heuristics that allow the parser to make use of the acquired target language grammar efficiently in real time (Marinis, 2003).

Now, the question to ask is, to what extent can learners do this job? It is perhaps needless to say that native speakers of a language come with a full list of relevant syntactic representations and fully optimized and proficient parsing strategies to get the job done reliably well. However, this does not always seem to be the case for adult second language (L2) learners,

especially when taking into account the observation that many adult learners, even those who are highly proficient in their L2s, tend to be less accurate, less efficient, and more prone to errors and processing breakdown during their L2 performance, compared to native speakers of the language they are acquiring. Given that successful parsing in the L2 presupposes adequate knowledge of the target language grammar as discussed above, one possibility to account for relatively inconsistent and deficient L2 performance is to assume that the interlanguage grammar acquired by adult learners is incomplete, thus deviating from the target language norm (e.g., Bley-Vroman, 1990; 2009). In this regard, a large body of research, from a formal perspective in particular, has investigated whether or not learners can ultimately acquire L2 grammar to a degree that is qualitatively comparable to that of native speakers. Years of accumulated L2 acquisition research conducted to explore this question has provided much information about the processes of interlanguage grammar development, but it also has shown that even highly proficient L2 learners often display a wide range of (meta)linguistic performance variability within and across individuals (e.g., varying accuracy rates on grammaticality judgment tests (GJTs)), thus making it difficult to conclude what aspects of L2 grammar can or cannot be learned (e.g., Bley-Vroman, 1990, 2009; Clahsen & Muysken, 1996; Epstein, Flynn, & Martohardjono, 1996; Hawkins & Chan, 1997; Hawkins & Hattori, 2006; Johnson & Newport, 1989, 1991; Schwartz & Sprouse, 1996; White, 1992; White & Genesee, 1996; White & Juffs, 1998; see White, 2003, for review and discussion).

Another limitation in this line of research is its methodology. That is, with the type of data obtained from those studies, which mostly consist of intuitional judgment data collected during offline tasks (e.g., grammaticality judgments, acceptability judgments, or truth-value judgments; see Gass & Selinker, 2008, for discussion), it is difficult to pinpoint how, at which

point, and on what basis during reading learners come to accept or reject certain sentences in question. This is a very critical piece of information because there may be perhaps more than one factor driving learners to certain judgments. In other words, learners' judgments may not necessarily be based on their knowledge of the target language grammar that the test attempts to tap into. Birdsong (1992), for example, used an acceptability judgment test, and found that the L2 French learners in his study made correct judgments similarly to the native French controls, but the reasons participants gave for their decisions often exhibited variation within the L2 group and differed from those of native speakers, raising a potential issue of task validity (e.g., see Ellis, 1991; Tremblay, 2005 for discussion). Therefore, it is important to look into learners' processing performance more in detail to better understand how they use their interlanguage grammar to construct syntactic structures in real time. This information may in turn provide insight into learners' underlying L2 grammar (Jiang, 2007; Juffs & Rodriguez, 2015).

Taking into account learners' performance variability and the methodological limitations discussed above, there is now a growing body of research dealing with real-time L2 processing, with a question as to how learners process target language input and what kinds of processing (or parsing) mechanisms and information resources (e.g., lexical, syntactic, and semantic) L2 learners access during their reading and listening comprehension. Employing online time course measures such as cross-modal priming (Swinney, 1979), self-paced reading (Just, Carpenter, & Wolley, 1982), or eye-tracking (for a review of this method, see Clifton, Staub, & Rayner, 2007; Dussias, 2010; Roberts & Siyanova-Chanturia, 2013), L2 processing research mainly aims at capturing learners' moment-by-moment parsing decisions to examine exactly how they construct structural representations. Of particular interest in this line of research has been whether the ways nonnative speakers and native speakers process (or parse) incoming L2 input online are

qualitatively the same or different. In addressing this big question, a number of studies have also examined whether the extent to which L2 processing converges on or diverges from L1 processing is modulated by some other variables, such as different L1 morpho-syntactic properties and related processing strategies (e.g., Aldwayan, Fiorentino, & Gabriele, 2010; Dussias, 2003; Frenck-Mestre, 1997, 2002; Hopp, 2010; Jegerski, VanPatten, & Keating, 2011; Juffs, 1998, 2005; Juffs & Harrington, 1995, 1996; Marinis, et al., 2005; Jegersky, VanPatten, & Keating, 2011; Jiang, Novokshanova, Masuda, & Wang, 2011; Keating, 2009; Papadopoulou & Clahsen, 2003; Sagarra & Ellis, 2013; Omaki & Schulz, 2011, Trenkic, Mirkovic, & Altmann, 2014; White & Juffs, 1998), L2 proficiency (Fernandez, 1999; Frenck-Mestre, 2002; Hopp, 2006; Jackson, 2008; Jackson & Dussias, 2007), L2 exposure (e.g., Cuetos, Mitchell, & Corely, 1996; Dussias & Sagarra, 2007; Ha, 2005; Pliatsikas & Marinis, 2013), or individual working memory capacity of learners (e.g., Dussias & Piñar, 2010; Felser & Roberts, 2007; Juffs, 2004, 2005; Williams, 2006). In search of answers to these questions, processing of filler-gap dependency constructions has received much attention in recent L2 processing literature (e.g., Cunnings et al., 2010; Dekydtspotter & Miller, 2009; 2013; Dussias & Piñar, 2010; Juffs, 2005; Kim et al., 2015; Marinis, Roberts, & Felser, 2005; Miller, 2015; Omaki & Schulz, 2013; Williams, 2006; Witzel, Witzel, & Nicole, 2002; see Clahsen and Felser, 2006a, 2006b; Juffs & Rodriguez, 2015).

What makes filler-gap dependency constructions intriguing from a processing perspective is that for some languages such as English and Spanish, there is a non-canonically positioned constituent overtly moved out of its original theta position, for example, a fronted wh-phrase in the wh-question and relative clause in (1) and (2) respectively.

(1) *Who_i* did the police know *t_i* the pedestrian killed *t_i*? (Dussias & Piñar, 2010)

(2) The nurse_{*i*} *who_i* the doctor argued *t_i* that the rude patient had angered *t_i* is refusing
to work late. (Marinis et al., 2005)

When processing constructions containing a displaced wh-phrase (called a *filler* or *wh-filler*) for comprehension, the parser needs to track down where the filler was originally positioned (called the *gap*), and figure out how it is associated with the other part of the sentence both syntactically and semantically. This procedure is referred to as filler-gap processing. As will be discussed in more detail in the next chapter, it is important to note here that there is cross-linguistic variation with respect to the way the wh-phrases are treated in their grammatical representations. That is, different from English or Spanish, languages such as Korean, Chinese, and Japanese do not require such movements, and a wh-phrase stays in its base-generated position (wh-in-situ), thus not necessitating such filler-gap processing procedures at least at the level of syntax. With this in mind, there is ample evidence in the L1 processing literature that for wh-movement languages, the parser universally constructs grammar structures incrementally (e.g., immediacy hypothesis, Just & Carpenter, 1980) and actively seeks to fill the gap as early as possible by releasing the filler to every structurally possible trace position (marked as *t_i*) until it finally finds its home (i.e., gap) (e.g., active filler strategy by Frazier & Clifton, 1989; trace reactivation hypothesis by Swinney, Ford, Frauenfelder, & Bresnan, 1988; but cf. Pickering & Berry, 1991). However, relatively less has been uncovered and no converging evidence seems to have been established yet when it comes to processing such structures in L2 contexts, especially regarding what types of linguistic resources and representations learners compute to locate the non-canonically positioned filler's place during real-time processing.

In this regard, several different positions have been put forward to account for characteristics of L2 processing. One position states that adult L2 learners may experience greater parsing difficulties than native speakers especially when the parser is expected to face momentarily heavier parsing loads, but that they still may process the target language structures in a qualitatively similar way to native speakers (e.g., Dussias & Piñar, 2010; Juffs, 2005; Juffs & Harrington, 1995; 1996; Williams, Mobius, & Kim, 2001). In the same vein, a body of research has also argued for L1-L2 parsing similarities under certain conditions, such as when learners have high proficiency (e.g., Hopp, 2006; Omaki & Schulz, 2011; Sagarra & Herschensohn, 2010; Williams, 2006), extensive L2 exposure (e.g., Dussias & Sagarra, 2007; Frenck-Mestre, 2005; Pliatsikas & Marinis, 2013), and when there is closeness of L1-L2 syntactic properties (e.g., Zawiszewski, Gutierrez, Fernandez, & Laka, 2011). Omaki and Schulz (2011), for example, implemented a self-paced reading test to examine how advanced L1 Spanish learners of English process long-distance filler-gap dependency constructions. Omaki and Schulz found that the nonnative speakers in their self-paced reading study not only searched for the gap actively in a comparable fashion to native speakers, but they also demonstrated that they apply relevant grammatical constraints into processing with “substantial grammatical precision” during filler-gap processing (p. 585). Zawiszewski et al. (2011) reported the role of L2 proficiency and L1-L2 distance in syntactic features through their event-related potential (ERP) study with L1 Spanish learners of Basque. Zawiszewski et al. tested three syntactic parameters: the head parameter, argument alignment, verb agreement. The first two had diverging syntactic features between the L1 and L2 whereas the same verb-agreement feature was shared by both L1 and L2 systems. They found that divergence in syntactic parameters between L1 and L2 may yield L2 processing patterns that are different from L1 processing, but

that native-like L2 processing is possible as a function of increased L2 proficiency (see also, Aldwayan et al., 2010, for similar findings). A more recent study by Pliatsikas and Marinis (2013) that tested L2 learners' use of intermediate gaps during online filler-gap processing found that learners with substantial immersion experience in the L2 process the target language structures similarly to the native speakers employing detailed syntactic representations (see also, Dussias & Sagarra, 2007, for a similar finding).

On the other hand, there are a number of other researchers who view adult L2 processing as something essentially different compared to native processing. Based on empirical findings that mostly come from research on L2 processing of filler-gap dependencies and ambiguous relative clause constructions, these researchers claim that L2 learners are limited in their use of grammatical information when processing in the L2, by and large irrespective of learners' L1s, L2 proficiency, and available cognitive resources such as working memory capacity (e.g., Cummings et al., 2010; Felser et al., 2003, 2012; Felser & Roberts, 2007; Marinis et al, 2005; Papadopoulou & Clahsen, 2003). This position eventually led Clahsen and Felser (2006a, 2006b) to propose their shallow structure hypothesis (SSH). According to the SSH, L2 processing by adult learners is fundamentally different from L1 processing, in that nonnative speakers are much less likely to be able to utilize rich and fully detailed syntactic representations when processing the L2 online, either because of their deficient and inadequate interlanguage grammar representations (e.g., Bley-Vroman, 1990) and/or their inability to compute sufficiently detailed syntactic information in the input stream in real time. This is the case even if learners have acquired relevant syntactic representations of the target language either via their L1 grammar or through the development of their L2 learning. On the basis of these assumptions, the SSH suggests that adult L2 learners, even at a highly proficient level, are led to take the shallow

processing route, which is fed by less detailed and incomplete syntactic representations, and instead rely heavily on lexical, semantic, pragmatic and discourse information.

In the light of what has been discussed above, the present study aimed to test the validity of the shallow structure hypothesis by investigating how proficient adult ESL learners with either L1 Korean or L1 Chinese background process long-distance filler-gap dependency constructions in English during real time processing. Specifically, the focus of the study was on L2 learners' use of relative clause island constraint during online reading. As will be discussed more in details in the next chapter, the relative clause island is a type of syntactic structure that does not allow the formation of filler-gap dependencies inside the island constructions. In other words, when the parser encounters this relative clause island structure in the input, it should avoid postulating a gap inside the island as there is no grammatically licit gap for the filler in the grammatical representations. From a SSH point of view, however, adult L2 learners would not be able to employ such abstract and hierarchically detailed syntactic constraints in real time processing (cf. Cunnings et al., 2010). Consequently, L2 learners may attempt to postulate a gap inside the relative clause island, constructing a representation that lacks sufficiently detailed hierarchical configurations. Therefore, looking into how adult L2 learners handle the filler when encountering an island structure in the middle of filler-gap processing may provide an important piece of information whether L2 processing is indeed guided by the shallower processing routes as the SSH would predict. To this end, this study implemented an eye-tracking reading experiment and analyzed learners' eye-movement patterns to provide more insight into the ways learners handle non-canonically positioned fillers under an island environment during reading.

The present study also explored the role of age of acquisition (or age of immersion) by including L2 learners who have been immersed in an ESL environment from early ages

(hereafter, early ESL). As laid out above, the SSH assumes that adult learners are restricted in making use of detailed syntactic representations during real-time structure building, either because of the representational problem (i.e., grammatical knowledge) or application problem (i.e., processing ability). Of the two possibilities, if we assume that adult L2 learners hold grammatical representations that are qualitatively comparable to those of native speakers of the target language (e.g., Rothman, 2008; Schwarz & Sprouse, 1996; White, 2003; White & Juffs, 1998, among others), the question to ask is, “Is this applicational limitation of adult L2 learners on syntactic processing due to age-related issues, for instance delayed ages of acquisition or exposure to a target language environment past critical (or sensitive) period for language learning?” The SSH does not provide a clear reason as to what precisely causes adult learners’ limited ability to access the full parsing route if it is not a representational problem, although they (adult learners) are capable of computing other linguistic types of resources (e.g., lexical and semantic information) (Dekydtspotter, Schwartz, & Sprouse, 2006), as well as some other domains of (morpho-)syntactic processing (e.g., Ojima et al., 2005; Sabourin & Haverkort, 2003). In this regard, comparing how highly advanced adult and early L2 learners process the target language structures may provide further insight into the nature of adult L2 processing.

There is a general consensus that L2 learners who have been immersed to a L2 environment at earlier ages before puberty are generally more efficient, fluent, and arrive at a native-like end-state grammar stage in a more consistent way than adult learners, who often display a wide range of L2 grammar knowledge and/or performance variability. The fact that adult learners demonstrate more performance variability than early learners makes it particularly crucial to examine the way these two L2 groups process the target language structures online. Even if syntactic knowledge of the two distinct groups is comparable, processing could differ,

meaning that there could be a critical or sensitive period that restricts adult learners' application of the grammar in real time. When such a critical or sensitive period would fall is unknown, but it is possible that early learners (if early enough) would have an advantage regarding the acquisition of full, native-like processing strategies. Thus far, however, there has been only a very small volume of L2 sentence processing research that directly examined the role of age of acquisition on L2 processing by comparing early and adult L2 learners in the same study (e.g., Ha, 2005; Weber-Fox & Neville, 1996), whereas most studies tested only adult L2 learners. Thus, it has not been clearly revealed yet whether or not, and how, acquisition of proficient parsing ability to apply detailed parsing mechanisms is affected by the age that learners have been exposed to the target language. Taken together, delving into how early-immersed and adult L2 learners perform filler-gap processing using the same test materials may provide valuable information not only for evaluating the claims of the SSH, but also for increasing our understanding of the nature of adult learners' acquisition of L2 grammar and processing, and how different ages of immersion influence learners' development of grammatical knowledge and processing abilities.

Lastly, it is another goal of this study to explore the role of individual differences in working memory capacity (WMC). It has been well attested in the current L2 literature that adult L2 learners are generally slower and less efficient in their L2 processing than L1 speakers (e.g., Juffs, 2005; Segalowitz & Segalowitz, 1993. Ullman, 2004; Williams et al, 2001; 2006). If this is the case, it may be reasonable to assume that learners' cognitive resources during online processing are more taxed because the amount of time the processor should hold unanalyzed constituent information such as wh-fillers will increase due to delayed integrations of the processed information (e.g., Kann, Ballantyne, & Wijnen, 2015). As a result, adult L2 learners

may be more prone to processing difficulties due to their shortage of cognitive resources relative to the amount of processing cost the parser has to pay. This seems more likely especially when learners process highly demanding target language structures such as long-distance filler-gap dependencies in real time. As a result, L2 learners, those with lower WMC in particular, may have less chance to access detailed syntactic information to a degree that native speakers of the target language would do. Note however, that the SSH does not predict any effect of different individual WMC in L2 processing assumedly because L2 grammatical representations that are computed by the parser are more likely shallower and less detailed regardless of one's WMC. While there is growing interest in the role of WMC on L2 processing, the current literature has not yet provided enough data to elicit any clear conclusion as to how individual differences in learners' WMC affect the way learners process the target language structures during online sentence processing. Therefore, it seems that further observations of the role of WMC are obviously needed.

The rest of this dissertation is organized as follows: Chapter 2 provides the theoretical background for the filler-gap dependency representations/processing and working memory, and reviews recent L2 processing research, followed by the research questions that guide the present study. Chapter 3 provides the information of the participants, research design, materials, and overview the data collection and analysis processes. Chapter 4 reports the results, and Chapter 5 discusses of the results more in detail in light of the research questions. Finally, Chapter 6 provides a brief summary of the research findings, and addresses some of the limitations of this study, and makes suggestions for future research.

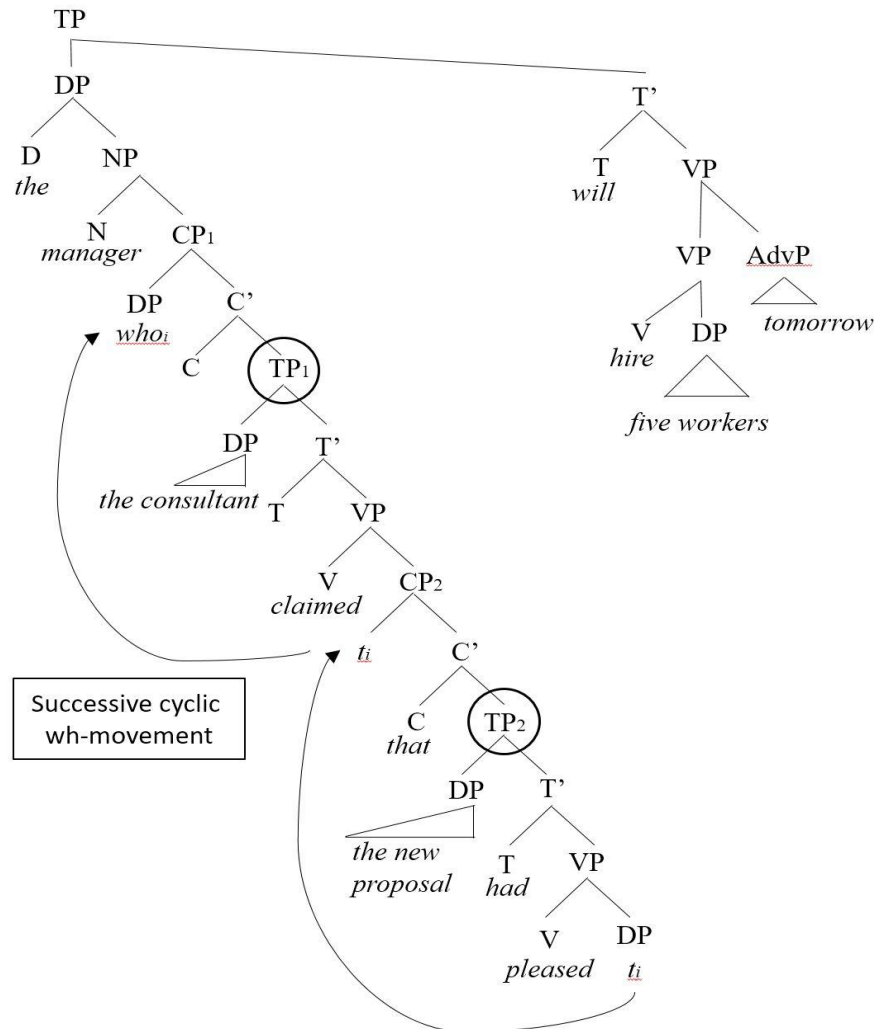
CHAPTER 2: REVIEW OF LITERATURE

2.1. Grammatical representations of wh-structures

2.1.1 Filler-gap dependency representations in English

According to most recent generative accounts (e.g., Adger, 2003; Chomsky, 1995), wh-questions or relative clauses in English are the product of movement, as shown in (3).

- (3) The manager who the consultant claimed that the new proposal had pleased will hire five workers tomorrow.



(The sentence was adapted from Gibson & Warren, 2004, p.61)

In (3), the wh-phrase *who*, the relative pronoun co-indexed with the antecedent *the manager*, is the object argument of *pleased* in its underlying position. It is fronted to the sentence initial specifier of the matrix CP, CP₁ in order to check and delete the strong uninterpretable wh-feature in English. The movement operation in this case is to be guided by the grammatical constraints generally known as subjacency (Chomsky, 1973; 1981), according to which a wh-constituent may not cross more than one bounding node² at a time thus restricting movement of the wh-phrase to be more local (for a review of more recent theoretical accounts and discussions on subjacency, see Belikova & White, 2009). Taking into account this movement constraint, the wh-phrase *who* in (3) needs to be moved up through two separate movement steps; first from the canonical position to the specifier of CP₂—a syntactic gap known as an empty category (*e*) or wh-trace (*t*)—and then to the specifier of CP₁ successively. This movement, which is referred to as successive cyclic movement, illustrates not only how the dislocated wh-phrase is syntactically associated with the other parts of the sentence—specifically with its subcategorizing verb—but also the role of the mediating site (i.e., the empty category CP₂) in the grammatical representation for establishing legal long-distance movements (cf. Pickering & Barry, 1991; Sag & Fodor, 1995).

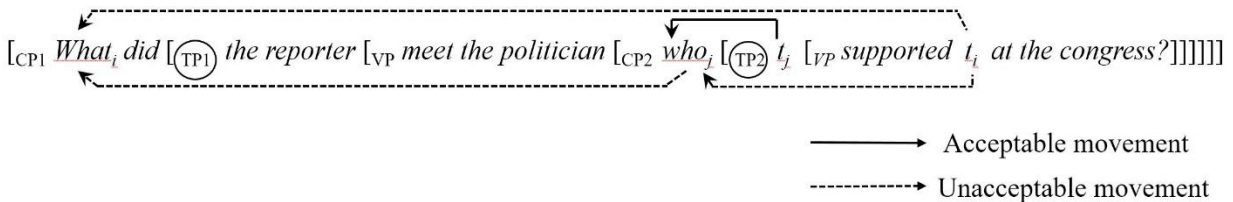
2.1.2. Island Constraints: Violation of the movement constraints

As shown, the movement of wh-phrases in wh-movement languages is strictly limited to be local, and it needs a mediating site in its representation to go out of more than one bounding

² While what constitutes a bounding node varies cross-linguistically, it generally refers to a NP/DP or IP/TP (S) in English, which is circled in the tree structure in (3) (Hawkins, 1999; Rahman, 2009).

node. The unavailability of a mediating site in the structural representation may result in movements that are not grammatically licit for long-distance movement, as illustrated in (4).

(4) What did the reporter meet the politician who supported at the congress?



What makes the sentence in (4) ungrammatical is the movement of *what*, the object argument of the relative clause verb *supported*. Considering there is more than one bounding node (i.e., TP1 and TP2 that are circled) between the surface (i.e., sentence initial) and original canonical position, the only way the *what* can move to the current place without violating the subadjacency principle is to move via the spec of the lower CP (i.e., CP2), just as is the case with *who* in (3). However, it is not possible to take this route because the site is already occupied by the relative pronoun *who*. As a result, the only option is moving directly to the spec of CP1, but this makes the movement illicit, resulting in the sentence being ungrammatical as it crosses two bounding nodes (TP1 and TP2) in a single movement. This phenomenon is generally known as the island constraints, specifically a relative clause island in the case of (4) (Ross, 1967). Ross identified that a to-be-raised wh-constituent cannot be placed within certain structure types (called *islands*, including relative clauses among others such as complex NP, subject NP, and adjunct island), because as shown in (4), there is no way for the constituent to be legally extracted out of those island structures without violating the locality constraint (for a review, see Belikova & White, 2009).

2.1.3. Grammatical representations of wh-structures in wh-in-situ languages

While English, a language with a strong wh-feature, involves syntactic movement operations guidance of the locality constraints as discussed above, some other languages such as Chinese, Japanese, and Korean do not require such overt movement operations at least at the level of syntax, because the wh-feature of those languages is weak, and therefore does not require any further steps for feature checking. See the Japanese wh-question in (5) and its equivalent in Korean in (6) below.

(5) Japanese

John·wa [CP Mary·ga kinou nani·o kat·ta· to] oboete imasu-ka?

John·TOP [CP Mary·NOM yesterday what·ACC buy·past-C] remember is-Q-Part?

What did John remember Mary bought yesterday? (Hawkins & Hattori, 2006, p.275)

(6) Korean

John·en [CP Mary·ka eoje mwsett·eul satta·ko] kiyeokkako isseumni·ka?

John·TOP [CP Mary·NOM yesterday what·ACC buy·past-C] remember is-Q-Part?

What did John remember Mary bought yesterday?

In these two sentences, the wh-phrase *nani* in (5) and *mwsett* in (6) is a wh-word equivalent to ‘what’ in English, which remain in their canonical position (i.e., in situ) as the object of *kata* and *satta* ‘bought’ respectively. Thus, neither overt syntactic movement operations for feature checking nor the subsequent locality movement constraint is instantiated in those languages³.

³ Some researchers have questioned whether languages such as Chinese really lack overt evidence of the operation of Subjacency. See Lardiere (2008) for discussion.

2.1.4. Summary

As shown above, there is a cross-linguistic variation in the strength of wh-feature and its consequences in the structural representations across different languages. This has been a topic of considerable interest within the Universal Grammar (UG)-based L2 acquisition research (e.g., Bley-Vroman, Felix, & Ioup, 1988; Johnson & Newport, 1991; Schachter, 1989, 1990; Schachter & Yip, 1990; White, 1992; White & Juffs, 1998). Specifically, the main research question these studies address is whether adult L2 learners whose L1s lack wh-movement can successfully acquire the abstract and subtle locality constraints instantiated in their target languages (e.g., English). Whereas most of these studies used offline grammaticality judgment tests (e.g., detecting sentences with subjacency violation), the results are mixed. As noted earlier in Chapter 1, the current study tested Chinese and Korean speakers (wh-in-situ) learning English (wh-movement) to examine how they process filler-gap dependency structures online in their L2 English. Instead of using grammaticality judgment tests, this study analyzed learners' eye-movement patterns during reading to provide insight into whether they have acquired the relevant L2 grammar that cannot be acquired from their L1 (i.e., locality constraint, specifically with relative clause islands), and if so, whether they can employ such knowledge of L2 grammar during online processing. The next section discusses the nature of language processing in general followed by a review of literature on L1 and L2 filler-gap processing research.

2.2. Incremental processing: Evidence from processing of garden-path sentences

Models of processing offer different accounts of when and how the different components of processing come into play during comprehension. Of those, the modular-based accounts (e.g., Frazier, 1987, 1998; Frazier & Rayner, 1982) have been predominant both in L1 and L2 sentence

processing research⁴. This model assumes that each operation is computed in its own module separately due to computational limitations; the syntactic operation occurs at an earlier stage of processing and builds structural representations so that other types of processing (e.g., semantic processing) can come into play at later stages of processing. One of the most consistent observations found across the sentence processing research under these modular accounts is that comprehension is formed through a series of incremental interpretations. That is, the parser organizes a representation of the sentence incrementally, word by word, in a bottom-up fashion, computing applicable syntactic/semantic information immediately as each word comes into the parse, which is well presented in the parsing principle proposed by Frazier and Rayner (1982)'s 'late closure' and Pritchett's (1992) 'generalized theta attachment' as in (7) and (8), respectively.

(7) Late Closure: When possible, attach incoming lexical items into the clause or phrase currently being processed. (Frazier & Rayner, 1982, p. 180)

(8) Generalized Theta Attachment (GTA): Every principle of the syntax attempts to be maximally satisfied at every point during processing. (Pritchett, 1992, pp. 155)

The incremental nature of language processing has been well attested in research on processing of garden-path type of sentences (e.g., Frazier, 1987), such as the one in (9):

(9) After Mary ate the pizza arrived from the local restaurant. (Juffs, 1998, p. 411)

What may lead readers down the so-called garden-path in reading (9) is the likely initial interpretation of *the pizza* as the direct object and theme of the preceding verb *ate*, by means of

⁴ Another model that competes with the modular-based model is the constraints-based interactive model (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus & Trueswell, 1995), which assumes that all possible sources of syntactic alternatives (e.g., semantics, context, and frequency of syntactic structure) are processed in parallel, and one information receiving most support from the analysis gets higher activation. See, van Gompel & Pickering (2007).

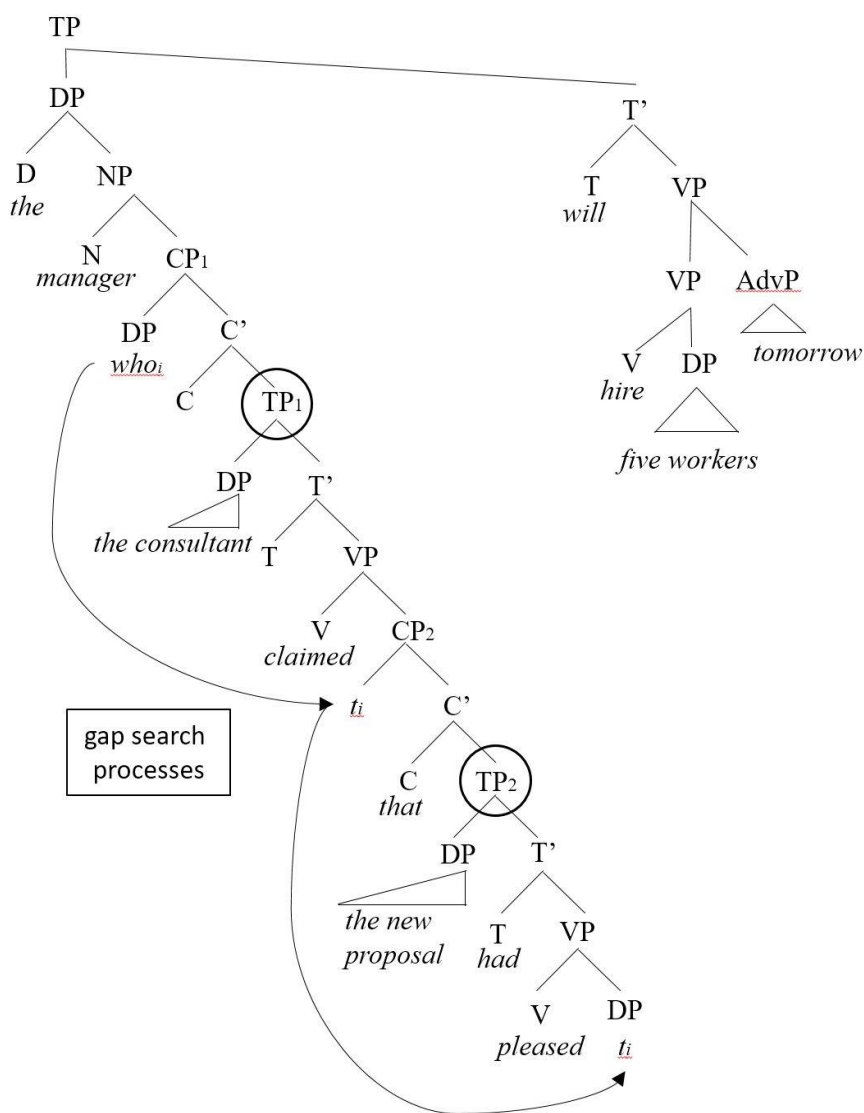
incremental VP integrations. This interpretation, however, must be rapidly revised as soon as the parser reaches *arrived*, in that the noun phrase (*the pizza*) should be integrated into the matrix clause receiving a new case and thematic role from the matrix verb *arrived*. The structural computations for such reanalysis are costly and may impose a momentary processing burden on the parser (e.g., theta reanalysis constraint⁵, see Pritchett, 1992), possibly yielding a slowdown at *arrived*. This phenomenon is generally referred to as a garden-path effect. As illustrated, incremental structure building may result in relatively complex computations at times because it integrates a word without knowing what will follow next, but it has been well attested that it is an essential design feature that eventually helps the processor to keep its (working) memory system manageable for efficient comprehension (e.g., Frazier & Fodor, 1978; Lewis, 1998; Staub & Clifton, 2006). The next section discusses how the parser carries out gap search processes for the filler while building the structural representation incrementally.

⁵ Theta Reanalysis Constraint (TRC): “Syntactic reanalysis which re-interprets a theta-marked constituent as outside of a current theta domain is costly.” (Pritchett, 1992, p.15)

2.3. Processing of filler-gap dependencies: Incremental gap search processes

When processing a sentence such as the one in (3) above, which is copied in (10) below with a slight modification for readers' convenience, the parser must search for the canonical position of the filler (i.e., the gap) and integrate it with a relevant component (e.g., the verb).

(10) The manager who the consultant claimed that the new proposal had pleased will hire five workers tomorrow.



This so-called filler-gap process must be completed as quickly as possible given that working memory, which is responsible for maintaining unanalyzed filler information, has limited capacities (e.g., Gibson, 1998; Wagers & Philips, 2014). Linguistic theories and related processing frameworks diverge as to how the filler-gap dependency relation is formed in the representation, and consequently what kind of linguistic resources the parser consults to construct a representation to link the dislocated filler with the canonical gap site in the most economical way (e.g., Gibson & Warren, 2004; Traxler & Pickering, 1996), but a broad consensus is that the gap-search process is essentially incremental as well. That is, once a displaced filler is identified, the parser actively searches for its canonical position by incrementally testing out syntactic and/or semantic fits of the filler at every grammatically possible gap position as it moves forward, whether it be of empty categories (e.g., trace-based active filler hypothesis, Clifton & Frazier, 1989; trace reactivation hypothesis, Swinney et al., 1988)—or subcategorizing verb⁶ (e.g., traceless-based direct/immediate association strategies, Pickering & Berry, 1991). In the current study, I adopt the generative-based parsing framework and assume that *wh*-constructions are formed by means of movement operations through syntactically postulated *wh*-traces (i.e., a silent copy of the filler), given that there is a good

⁶ Non-transformational syntactic frameworks such as General Phrase Structure Grammar (GPSG) or Head-driven Phrase Structure Grammar (HPSG) neither accept the concept of syntactic *wh*-movement operations nor the postulation of unpronounced hypothetical *wh*-traces (empty categories). Processing frameworks based on those syntactic accounts (e.g., direct association) assume that the filler is directly associated with the unresolved subcategorizing verb (i.e., filler as a missing obligatory argument of the verb). Thus, the crucial cue for filler-gap processing under this system is the verb, not empty categories (cf. Aoshima et al., 2004). Note that with head-initial languages such as English, it is empirically difficult to dissociate the two accounts (i.e., trace- vs. traceless-based) because the sites of the potential gaps and verb subcategorization overlap with each other (cf. Gibson & Warren, 2004; Lee, 2004). Nakano and colleagues used *wh*-scrambling and object-topicalization structures in Japanese to test trace-reactivation effects, a head-final language where a verb rigidly occurs before its arguments, and found evidence of trace reactivations that occurred even before the verb is processed.

amount of literature that shows evidence for the psychological reality of syntactic wh-traces in processing, as will be discussed below (e.g., Gibson & Warren, 2004; Lee, 2004; Nakano, Felser, & Clahsen, 2002, as cited in Marinis et al., 2005; see also, Featherston, 2001 for related discussions).

According to the trace-based parsing accounts, the incremental gap-search process is mediated by sets of wh-trace positions assigned by the grammar, such as potential argument positions or cyclic non-argument Spec of CP₂ positions at the clausal boundaries as in (8), through which the parser reactivates the filler information from the left-most possible extraction site (i.e., structurally defined gap) until it finally confirms the true canonical position of the filler. Retrieval of the filler in such a manner has been argued to not only reduce memory cost in the working memory system that otherwise may have been higher especially with increased linear filler-gap distance (Gibson, 1998), but also to facilitate the ultimate filler-gap integration (e.g., Gibson & Warren, 2004; Traxler & Pickering, 1996). In their self-paced reading experiment, Gibson and Warren (2004) tested a) whether native English speakers make use of the mediating wh-trace site (“intermediate structure” in the authors’ terms) when processing English long-distance wh-constructions such as the one in (10)—specifically a reactivation at the non-argument Spec CP₂ posited at the complementizer *that*—and if so, b) whether this facilitates the later filler integration at the final destination, *pleased* as a result of a decrease in linear distance between the filler and the gap. To observe the facilitation effect of the mediating gap site [a] in (11), the [- intermediate gap] counterpart of the sentence was added, as in (12).

- (11) The manager *who*_i the consultant claimed [A] *ti* that the new proposal had pleased [B] *ti*
will hire five workers tomorrow. (+extraction across VP, + intermediate gap)

(12) The manager *who_i* the consultant's claim about the new proposal had pleased [C] *ti*
will hire five workers tomorrow. (+extraction across NP, - intermediate gap)

Also included were non-extraction counterparts of the two extraction types, as exemplified in (13) and (14).

(13) The consultant claimed [D] that the new proposal had [E] pleased
the manager who will hire five workers tomorrow. (-extraction, VP)

(14) The consultant's claim about the new proposal had [F] pleased the manager who will
hire five workers tomorrow. (-extraction, NP)

Gibson and Warren found that reading times (RTs) at the gap sites (i.e., [A], [B], and [C]) in the extraction condition were significantly longer than RTs at the corresponding regions in the non-extraction condition (i.e., [D], [E], and [F] respectively), showing evidence for filler-retrievals in the gap sites. This suggests that in the extraction conditions, the parser spent extra processing time to postulate a gap for the filler retrieval and run analyses to evaluate its appropriateness as the potential landing site, whereas there is no need to do so for the sentences in the non-extraction condition. More crucially, reading profiles at the ultimate gap (i.e., *pleased*) were found to be significantly shorter in the [+ intermediate gap] in (11) than in the [- intermediate gap] condition in (12). Taken together, Gibson and Warren claimed that native English speakers incrementally postulate intermediate gaps in accordance with the syntactic representation in their mental grammar and utilize those gaps for the filler reactivation, which not only helps the parser to maintain the filler information with less memory burden, but also facilitates the integration of the filler with its subcategorizing verb (i.e., [b] *pleased* in (11)) as the linear distance between the reactivated filler and the gap decreases.

Another piece of evidence supporting incremental gap search comes from studies that investigated the processing of wh-structures manipulated for plausibility, as illustrated in (15). Plausibility in this case refers to the semantic relationship between the filler and the first verb that the parser encounters where an early gap-filling analysis can take place under the assumption of the active filler-gap creations. Thus, the integration of each of the antecedent NPs, *the book* and *the city*, with the verb *wrote* yields either a plausible (i.e., *wrote the book*) or an implausible (i.e., *wrote the city*) interpretation respectively.

(15) We liked the *book_i / city_i* that the author wrote *t_i* unceasingly and with great dedication about *t_i* while waiting for a contract.

Traxler and Pickering (1996) examined through their eye-tracking reading study whether L1 English speakers show any plausibility effect at the early gap site. It was observed that the participants displayed significantly longer RTs in the implausible condition than in the plausible condition at the verb, *write*. Such mismatched RTs between the two plausibility conditions suggest that the parser postulates an object gap as soon as it encounters the verb (*write*) rather than postponing the filler-integration until it identifies the ultimate gap at the preposition *about*. The integration yielding implausible interpretations should bother readers' processing, simply because the interpretation does not make sense up until the initial integration point at the least, and also partly because the parser must prepare for a reanalysis more immediately compared to the plausible condition (see also Aoshima, Philips, & Weinberg, 2004; Frazier & Clifton, 1989; Lee, 2004; Stow, 1986, for more review of L1 filler-gap processing).

2.4. Processing of filler-gap dependencies by nonnative speakers

2.4.1. A review of early research on L2 processing

One of the early studies that brought issues of L2 processing into focus was Juffs and Harrington (1995), which examined how advanced Chinese-speaking (wh-in-situ) learners of English process long-distance wh-constructions such as the ones in (16) and (17). While previous L2 acquisition research reported learners performed comparatively poorer on GJTs in the subject wh-extraction condition (e.g., Schacter & Yip, 1990; White & Juffs, 1998⁷), the authors investigated if the subject-object asymmetry phenomenon found in the past research was associated with learners' processing problems rather than a representational deficit in the L2.

(16) *Who_i did the police know t_i killed the pedestrian?* (subject wh-extraction)

(17) *Who_i did the police know t_i the pedestrian killed t_i ?* (object wh-extraction)

Juffs and Harrington focused on different levels of parsing complexities between the two extraction conditions based on Pritchett's (1992) principle-based parsing accounts (i.e., see, GTA provided in (8) and TRC provided in footnote 5 on p. 17 in this paper), and hypothesized that L2 learners would employ native-like active filler-gap strategies to fill the gap as early as possible. However, they reasoned that learners might have relatively heavier parsing difficulties in the subject wh-extraction condition, especially where the parser is expected to deal with momentarily more demanding linguistic analyses.

Assuming the operation of the active filler-gap strategy, the parser would initially postulate an object gap right after the verb *know* in both extraction conditions, assigning a case and thematic role to the filler (i.e., accusative and theme, respectively from *know*). The relative

⁷ For clarification, the experiment in White and Juffs (1998) preceded the ones carried out in Juffs and Harrington's (1995) study (Juffs, 2005, p.123).

difference in terms of processing complexity emerges in the next string, *killed* and *the* (*pedestrian*), respectively: In (16), the parser must cancel the initial analysis above as soon as it encounters the embedded verb, *killed* and postulate a subject gap concurrently with immediate case/theta reassignments (i.e., from object/theme of *know* to subject/agent of *killed*). The reassignment of the case and thematic role in this case occurs across the two different theta domains (i.e., the two verbs; *know* and *killed*), which according to Pritchett's (1992) theta reanalysis constraints is more costly for the parser. In contrast, however, such heavy reanalysis on the filler is not needed at this point in the object wh-extraction condition in (17), thus making the momentary parsing relatively easier. The results from the word-by-word self-paced reading confirmed the authors' hypothesis in that the Chinese ESL participants showed greater processing difficulties in the subject wh-extraction condition, as revealed by significantly longer RTs specifically at the second verb *killed* and significantly lower GJT performance. The authors argued that despite the cross-linguistic variation, the Chinese ESL participants processed long-distance wh-constructions qualitatively similarly to native English speakers, but that learners' processing ability to deal with moment-by-moment computations in real time might not be as fast and efficient as native speakers, especially when the loaded parsing assignments are heavy.

Williams et al. (2001) also argued for qualitative similarities between L1 and L2 processing, which included a stop-making-sense judgment task during self-paced reading to investigate how L2 learners deal with plausibility constraints, as in (18) and (19), during filler-gap processing.

(18) Which girl_i did the man push *ti* the bike into *ti* late last night? [plausible at V]

(19) Which river_i did the man push *ti* the bike into *ti* late last night? [implausible at V]

In both (18) and (19), the canonical position of the filler *which girl/which river* is after the preposition *into* in the adjunct phrase, and the plausibility manipulation is on the main verb *push*, the earliest possible extraction location as a landing site for the filler (i.e., *push the girl* and *push the bike*). Williams and colleagues found that all advanced ESL learner groups from various L1 backgrounds—[wh-]: L1 Chinese & L1 Korean; [wh+]: German—showed sensitivity to the plausibility constraints similarly to the English controls in that they made use of the left-most possible extraction site (i.e., *push*) as a possible landing site for the filler, yielding more stop-making-sense responses at this region in the implausible condition. However, the analysis of reading time patterns suggested that L2 learners' timing of reanalysis was not as immediate as the native speaker group and was found to be delayed especially in the implausible (i.e., *push the river*) than in the plausible condition (i.e., *push the girl*) despite the fact that implausibility information and an incoming determiner *the* in '*the bike*' signals a need for reanalysis. Furthermore, in a separate stop-making-sense judgment experiment, it was observed that the L2 learners had problems in canceling their initial analysis even in the offline task, particularly when the initial analysis was plausible (see also Williams, 2006). Taken together, Williams et al. concluded that the way L2 learners process filler-gap processing is qualitatively the same as native speakers, but they just may be slower in computing syntactic analysis and more prone to experience greater processing difficulties, especially when the parse has to withdraw its plausible misanalysis by the additional information that follows—reanalysis problems similar to the garden-path effect (e.g., Juffs & Harrington, 1996).

2.4.2. Shallow structure hypothesis

As briefly introduced in the previous chapter, the main argument of Clahsen and Felser's (2006a, 2006b, 2006c) shallow structure hypothesis (SSH) is that there are fundamental differences between L1 and adult L2 parsing. What distinguishes adult L2 processing from L1 processing under this hypothesis is the limited types of linguistic resources available to the L2 parser, particularly with respect to the availability of the parser's access to the full-fledged and hierarchically detailed grammatical representations during online processing. That is, the SSH assumes that the grammatical representations that feed the L2 parser during online processing entail structural information that is rather "rudimentary," "shallower," and lacking "hierarchical details," as compared to those deployed by native speakers (Clahsen & Felser, 2006a, p. 32).

Clahsen and Felser provided two possible reasons to account for learners' reduced ability to access the full parsing route. First the SSH assumes that adult L2 learners' interlanguage grammar system that feeds the parser is likely deficient and/or inadequate to process the target language input (e.g., Bley-Vroman, 1990). Another possibility is that even if learners' L2 grammar is fully detailed and appropriate for parsing, adult L2 learners may not have adequate parsing mechanisms and efficient processing abilities to compute sufficient information in real time. Shallower representations thus subsequently may prohibit learners from constructing a structural representation for the input in a native like manner because they simply don't have sufficient tools—for example, abstract features and grammatical constraints such as copies of movement traces and subjacency—for constructing hierarchically detailed syntactic representations during online processing. The SSH predicts that learners rely on the shallow parsing route predominantly instead, which is fed by pragmatic, simple verb-argument, and lexical information. Evidence that supports the SSH mostly comes from research that tested

learners' processing of either long distance filler-gap dependency constructions such as (10) above (e.g., Felser & Robert, 2007; Marinis, et al., 2005) or L2 processing of ambiguous relative clause (RC) constructions such as the one in (18) below (e.g., Felser et al., 2003; Papadopoulou & Clahsen, 2003). In the following, I will briefly review the literature that investigated learners' processing of ambiguous RC structures first, and provide more detailed review of research on processing of filler-gap dependency constructions.

2.4.3. A Review of the empirical research testing the SSH

2.4.3.1. L2 processing of ambiguous RC constructions

In (20) below, the noun phrase (NP) preceding the relative clause is complex, consisting of two NPs linked by genitive *of*.

(20) An armed robber shot [_{NP1} the sister of [_{NP2} the actor]] [_{RC} who was on the balcony].

The structural ambiguity arises when the parser must determine where to attach the RC. In other words, the RC can modify either the head of the NP phrase, NP1 (also referred to as high attachment), *the sister*, or NP2 (also referred to as low attachment), *the actor*, thereby inviting two possible interpretations regarding '*who was on the balcony*'. Languages differ as to how such ambiguity in (20) is resolved; speakers of some languages prefer to attach the RC to the NP1 (i.e., the sister was on the balcony; e.g., German, Greek, Spanish, Korean), and speakers of some other languages prefer the NP2 interpretation (i.e., the actor was on the balcony; e.g., English, Romanian, Swedish). Such different interpretation preferences can be explained by cross-linguistically different structure-based parsing strategies derived from different syntactic properties between languages. One explanation involves rigidity of word order; rigid word order languages such as English prefer a low-attachment interpretation (referred to as recency

preference), and languages with relatively free word order such as Spanish or Korean prefer a high-attachment interpretation (referred to as predicate proximity preference). See Gibson, Pearlmutter, Canseco-Gonzalez, and Hickok (1996) for more theoretical accounts). Felser, et al. (2003) investigated how advanced L1 Greek and L1 German (NP1 preference, both languages) learners of English (NP2 preference) process sentences like (18) during self-paced reading and also during offline reading. They found no clear NP attachment preference at all (i.e., no more than chance) for either L2 groups in both online and offline reading tests, whereas the English controls presented a clear NP2 preference as predicted. The authors suggested that no attachment preferences found in the performance of those L2 learners may be due to their lack of ability to apply any structure-based parsing strategies linked to the L2 (i.e., neither L1 transfer nor nativelylike), consequently making their attachment decisions random (see, Clahsen and Felser, 2006, and Papadopoulou & Clahsen, 2003, for similar findings; but see Dussias, 2003; Dussias & Sagarra, 2007; Frenck-Mestre, 1997, for counter-evidence).

2.4.3.2. L2 processing of long-distance filler-gap dependencies

Recall the studies discussed in section 2.4.1, which maintained that L2 learners, like native speakers of the target language, are able to apply relevant syntactic information and incrementally construct a representation that includes unpronounced syntactic gaps that are compatible with the grammar, but in a less efficient way when the parser is loaded with complex linguistic computations. In this regard, however, Clahsen and Felser (2006a) and Marinis et al. (2005) pointed out that with the types of test materials used in Juffs and Harrington (1995) and Williams et al. (2001), it is not possible to provide unequivocal evidence that L2 learners indeed made use of syntactically driven L2 information to postulate a gap. This is because in English, a

potential wh-extraction site is always adjacent to a verb, so that the potential gap positions and verb subcategorization and argument positions always overlap with each other. Consequently, it is unclear whether the incremental gap search by L2 learners is guided by syntactically driven trace information (active filler/trace reactivation hypothesis) or lexically driven verb-argument information (direct or immediate association hypothesis).

In an attempt to disambiguate such obscurity as to what types of knowledge resources are used by L2 learners, Marinis et al. (2005) adopted the test materials used in Gibson and Warren (2004) that observed the filler reactivation effect at the non-argument trace positions (specifier of the CP), as shown in (11) and (12), copied in (21) and (22) below. Slashes indicate how the sentences were segmented in their self-paced reading task.

- (21) The manager *who_i* / the consultant claimed / [A] *t_i* that / the new proposal /
had pleased [B] *t_i* / will hire five workers tomorrow. (+extraction, + intermediate gap)
- (22) The manager *who_i* / the consultant's claim / about / the new proposal /
had pleased [C] *t_i* / will hire five workers tomorrow. (+extraction, - intermediate gap)

L2 learners from both wh-movement (German & Greek) and wh-in-situ (Chinese & Japanese) L1 backgrounds, as well as native English speakers as controls performed a segment-by-segment self-paced reading task in English. Marinis et al. obtained RT profiles of the English controls that are similar to the native speaker data in Gibson and Warren's study; the native speakers displayed elevated RTs in the intermediate gap at [A] in (21), relative to its non-extraction counterpart (see an example in (13)), signaling a filler-activation effect at this non-argument trace position. The authors also found the reading patterns that were similar to those found in Gibson and Warren (2004). That is, the filler integration at [B] by the native English speakers was significantly faster than their filler integrations at [C], suggesting linearly decreased distance

between the gap and the filler as a consequence of the filler retrieval at the intermediate gap [A] eventually facilitated the filler integrations at the ultimate gap site where the filler is integrated with its subcategorizing verb. In contrast, neither the filler activation effect at [A] nor the facilitation effect at [C] were found in the L2 learners' RT profiles regardless of their L1 backgrounds, which according to the authors is suggestive of no syntactically driven gap postulations at the intermediate gap positions for learners, presumably due to unavailability of fully detailed syntactic information from their interlanguage grammar, at least during online processing where moment-by-moment rapid computations must take place. Based on those findings, Marinis and colleagues concluded that although L2 learners might search for the gap in an incremental manner like native speakers, they tend to rely much more on lexical-semantic information rather than syntactic information, thus attempting to associate the filler directly with incoming verbs.

However, in their reanalysis of Marinis et al.'s RT data later, Dekydtspotter, Schwartz, and Sprouse (2006) found that the German and Japanese groups displayed spillover effects at the region right after the intermediate gap region [A], *the new proposal* in the [+intermediate gap] condition. Dekydtspotter et al. argued that the delayed filler activation effects from learners might suggest that L2 learners might be slower and less efficient in integrating syntactic information during online parsing, but that does not necessarily indicate that learners' underlying L2 parsing mechanisms are qualitatively distinct from those of the native speakers.

Using the same test design and materials tested in Marinis et al. (2005), Pliatsikas and Marinis (2013) probed whether more exposure to naturalistic L2 input has an influence on adult L2 learners' processing. They compared reading profiles of native English speakers with the two advanced Greek-speaking ESL groups: the NE (naturalistic exposure) group consisting of ESL

learners who had been immersed into English-speaking environments (average LOR of 9.42 years) and those whose L2 experiences were limited to classroom exposure (CE) from their home countries. The authors found that while the CE group showed nonnative-like processing patterns (no evidence for filler reactivation on the non-argument syntactic gap and no facilitation effect at the final gap site in the +intermediate gap condition), the reading profiles from the NE group were found to converge with those from the native speaker group, showing evidence for the gap postulation at a site consistent with the grammar.

Another study by Felser and Roberts (2007) also explored adult L2 learners' use of intermediate gaps in real time processing. Felser and Roberts implemented a cross-modal picture priming task to examine whether advanced Greek-speaking [wh-movement, head-initial] ESL learners, divided by the two groups; low and high WM capacity, can demonstrate a picture priming effect (i.e., a picture that is identical to the antecedent) as a reference to the filler reactivation at structurally defined gap positions, as do the native speakers tested in Roberts, Marinis, Felser, & Clahsen (2007). Felser and Roberts found that performance of both the high and low WM L2 groups differed from the native speakers in Roberts et al.'s study in that the learner group showed a priming effect not only at the structurally positioned gap site, but also at the control gap (a structurally unrelated gap), whereas the native speakers, more precisely only those with the high WM spans⁸, showed the priming effect only at the syntactically relevant gap site. Based on these results, the authors concluded that whereas the learners in their study activated and maintained the filler information throughout their listening to the target sentences, they did not make use of grammatical details specifically for the filler reactivation at a gap, but

⁸ Note that in Roberts et al.'s study, the native speaker group with the low WM spans show priming effects neither in the control nor in the syntactic gap sites.

instead relied more on lexical and other non-structural cues to compensate for their limited grammatical processing.

2.4.3.3. L2 processing of island constraints

More recently, a few studies have begun to explore whether the parser respects syntactic island constraints in processing long-distance filler-gap dependency structures in the L2. Recall that the filler cannot be placed within certain structure types (islands) because the to-be-raised filler cannot move out of those island structures without violating locality constraints. From a processing perspective, this means that the parser would not postulate a gap when it encounters a syntactic island structure, as there is no syntactically licit gap in the grammatical representation. Omaki and Schulz (2011), for example, investigated how advanced Spanish-speaking ESL learners process wh-constructions such as (23) and (24), as compared to native English speakers (p. 575).

(23) [No island, \pm plausible]

The book/city that the author wrote____ regularly about____ was named for an explorer.

(24) [Island, \pm plausible]

The book/city that the author who wrote regularly saw____ was named for an explorer.

In their self-paced reading experiment, the authors used a plausibility manipulation as a diagnostic tool to examine whether learners avoid gap postulations inside the relative clause island structure. In the non-island condition, a plausibility effect is expected presupposing the parser' use of active filler strategy at *wrote* (i.e., longer reading times to read *wrote the city* than *wrote the book*). On the other hand, no such effect must be expected if the parser brings the

detailed syntactic representation of the island constraints into the parse, given that a plausible or implausible interpretation occurs only when the parser integrates the filler with the potential subcategorizer (*wrote*). The authors found that both the native English controls and the Spanish L2 group showed a plausibility effect only in the non-island condition, suggesting that the Spanish ESL learners did not attempt to create a gap in the island condition by utilizing the island constraints during reading. Based on their results, Omaki and Schulz (2011) have challenged the claims of the SSH, arguing that nonnative speakers, at least those who are advanced learners, can build abstract and detailed syntactic representations during filler-gap processing (see also, Aldwayan, Fiorentino, & Gabriele, 2010 [L1 Najdi Arabic <-wh>]; Cunnings et al., 2010 [L1 German <+wh> & L1 Chinese <-wh>]; Felser et al., 2012 [L1 German <+wh>] for similar findings).

In a more recent study by Kim, Baek, and Tremblay (2015), the authors investigated how L1 Korean [-wh] and L1 Spanish [+wh] learners of ESL process island constraints in English. While the test materials they used were similar to those in Omaki and Schulz (2011), Kim et al. employed a stop-making-sense task in the course of a segment-by-segment self-paced reading. The target sentences in their experiment are illustrated in (25) and (26).

(25) I wonder / *which book* | *which city* / the author / wrote passionately / about /
while / he / was travelling. [non-island, plausible | implausible]

(26) I wonder / *which book* | *which city* / the author / who wrote passionately / saw /
while / he / was travelling. [island, plausible | implausible]

In their results, the authors found in the stop-making-sense task that all group showed a plausibility effect only in the non-island condition. However, the L1 Korean group showed different response patterns from the L1 Spanish group and the English controls at *while* in the non-island condition. That is, for the native English and L1 Spanish ESL groups, their stop-making-sense rate increased as soon as they found the object argument of the preposition was missing in (25), reflecting a reanalysis effect (i.e., cancelling the initial plausible interpretation). The L1 Korean group, however, did not show such an effect in either island condition. In their reading time analysis, Kim et al. found somewhat interesting patterns. That is, the reading time profiles of the native English and L1 Spanish groups showed a significant interaction of plausibility and island, signaling that these participants did not postulate a gap in the island condition, and showed a plausibility effect only in the non-island condition. On the other hand, the L1 Korean group showed a similar reading pattern across the island conditions, with increased reading times in both island conditions (i.e., no interaction of island and plausibility). The statistical analysis revealed that unlike the other two groups, the L1 Korean ESL learners showed no significant interaction of the two factors. Kim et al. interpreted these results as suggesting that although L1 Korean participants knew that a gap was not allowed inside the island structure (from their stop-making-sense judgments and offline grammaticality judgments), their application of the relevant grammatical constraints might have been delayed at an early stage of processing (from their reading profiles) presumably due to crosslinguistically different ways of filler-gap formations in Korean. As a result, the authors claimed that unlike the Spanish ESL group that shares the same [wh] feature property and overt wh-movement characteristics with English, the Korean ESL learners whose L1 is distinct from English in this respect might

have more difficulties applying the grammatical representations immediately in real time (i.e., L1 effect).

2.5. The effect of age of immersion (or acquisition) and critical period hypothesis

As discussed earlier, the successful application of syntactic representations during online processing largely depends on the availability of adequate knowledge of target language grammars in the first place, without which the parser may not construct fully detailed syntactic representations, simply because it would not have the necessary tool to work with regardless of the availability of sufficient parsing strategies and abilities. The question of to what extent adult learners can acquire target language grammar has been one of the hotly debated topics over the years in second language research, especially in relation to the role of age of acquisition/immersion on adult learners' ultimate attainment (e.g., Birdsong, 2005; Bley-Vroman, 1990, 2009; Dekeyser, 2000, 2010; Johnson & Newport, 1989, 1991; Juffs & Harrington, 1995; Rothman, 2008; Schwarz & Sprouse, 1996; Weber-fox & Neville, 1996; White & Juffs, 1998). The discussion of age-related effects in adult L2 acquisition starts from the general observations that the adult L2 acquisition is not as reliable and stable as child L1 acquisition. Such differences between L1 and L2 acquisition often have been explained by the critical or sensitive period hypothesis (CPH) (Penfield & Roberts, 1959; Lenneberg, 1967). The CPH assumes that there is a limited developmental period that allows the acquisition of a language (L1, and perhaps L2) at a normal and native-like level. Once this period is over, the ability to learn a language declines, due to maturational changes in the neuro-biological system that is responsible for language learning (Birdsong, 1999; see also Singleton, 2005, for a review of different ranges of CP across different studies).

There are reasons to believe, however, that no such critical or sensitive period exists. As pointed out by Slabakova (2006), L2 acquisition differs from L1 child acquisition, given the fact L2 learners already have L1s, meaning that the language learning system in the brain has already been activated fairly early. She suggested that it would be more appropriate to consider age-related effects in L2 learning more in general. In the following, I review a few crucial empirical studies that directly investigated the effect of age of immersion in the acquisition of L2 grammars.

First, Johnson and Newport (1989) tested 46 L1 Chinese or Korean speakers learning English using an audio grammaticality judgment task that included various types of morpho-syntactic structures. The L2 participants varied in their ages of arrival (AOA), ranging from 3 to 26 years old, based on which they were divided into two groups: early arrivals (AOA:3-15) and late arrivals (AOA:17-35). The two groups were matched for length of residence. Johnson and Newport found that L2 learners with an AOA of seven and under showed a native-like GJT performance. For those early arrivals whose AOA fell between 7-15, there was a linear decline in the GJT scores. The late arrivals performed generally more poorly than the early arrivals, but there were no further gradual declines between their performance and increasing AOA, and they showed greater degrees of performance variability regardless of their AOA. Johnson and Newport suggested that there is a critical period for second language acquisition (see also Johnson & Newport, 1991, for similar findings from their test of L2 subadjacency with L1 Chinese-speaking learners of English).

Using the same type of an audio grammaticality judgment task developed from Johnson and Newport (1989), Dekeyser (2000) tested 57 L1 Hungarian-speaking ESL learners whose ages ranged from 16 and 81, with a minimum of 10 years of length of residence. The participants

were divided into two groups based on their ages of arrival: early learners (AOA between 1-16) and late learners (AOA between 17-40). In addition to the GJT, Dekeyser also measured L2 learners' verbal aptitude using the Hungarian version of the MLAT (Modern Language Aptitude Test). Dekeyser found a strong negative correlation between learners' AOA and their GJT performance. However, differently from Johnson and Newport (1989), neither late learners nor early learners showed a linear decline when the two subgroups were analyzed separately. Dekeyser also found that there was a positive correlation between the GJT performance and the verbal aptitude scores for the late learners, whereas no such correlation was found among the early learners. Dekeyser used this finding to support Bley-Vroman's (1990) fundamental difference hypothesis, claiming that whereas early learners reached native-like levels of proficiency independently from their language aptitude, late learners cannot acquire native-like L2 competence unless they had above average language aptitude that signals more explicit and analytic language analysis and general problem solving skills.

Birdsong (2014) later conducted an additional correlation analysis with the data in Dekeyser (2000, see appendix A in Dekeyser's paper). Interestingly, he found that for all AOAs together, learners' years of schooling was significantly correlated with their grammatical proficiency. Birdsong also found that learners' levels of education was positively correlated with their GJT scores, not only for the late learners, but also for the early learners in Dekeyser's study, indicating that the "education effect is systemic: significant correlations are not restricted to certain AOA spans or certain aptitude levels" (p. 48). See also, Hakuta, Bialystok, & Wiley, 2003, for a similar finding regarding the role of the amount of formal education.

As shown above, it appears that there is a role of age of acquisition and/or age of immersion in the acquisition of L2 grammar at least to a certain degree. However, the question of

how strictly certain critical periods constrain the degree to which adult learners can develop their target language grammar remains to be seen.

2.6. The role of working memory on L2 parsing

The role of individual differences in working memory (WM) has been receiving more attention in L2 processing research recently (e.g., Dussias & Piñar, 2010; Felser & Roberts, 2007; Juffs, 2004, 2005; Juffs & Harrington, 2011; Sagarra & Herschensohn, 2010). WM is a “multicomponent system responsible for active maintenance of information in the face of ongoing processing and/or distraction” (Conway, et al., 2005, p. 770). According to Baddeley’s (2003) most recent WM model, WM is made up of 3 sub-components, namely, the central executive, the short-term storage system (subdivided into the visuospatial sketchpad and phonological loop), and the episodic buffer. Under this system, the central executive supervises processing of perceived information (e.g., auditory/reading input) and controls the flow of this information to the other subcomponents. These subcomponents are a) the episodic buffer for linking to the long-term memory system, and b) the phonological loop for temporarily storing information (specifically, auditory input) in the phonological short-term store⁹ and maintaining it through the rehearsal process while other information is processed. In sum, WM involves a storage that can maintain a limited amount of information (e.g., trying to retain a filler) in the face of simultaneous processing (e.g., concurrent processing of incoming input for integrating them into meaningful units).

⁹ WM differs from phonological short term memory (PSTM). One such example for PSTM is the Non-Word Repetition (NWR) test, which have been often used in SLA research (e.g., Hummel, 2010) especially in relation to L2 vocabulary development. Since PSTM catches the memory capacity that occurs only within the phonological loop without simultaneous processing consideration, it is different from WM.

In regard to the role of working memory in L2 domain, a few studies reported a positive correlation between WM span measures (particularly, reading-span tests) and L2 reading skills (e.g., the grammar and reading sections in the TOEFL test; Harrington & Sawyer, 1992), and grammaticality judgment tests (e.g., Robinson, 2002). However, how WM has an influence on L2 online processing, especially processing of filler-gap dependencies and ambiguous relative clause constructions, has been investigated only in a limited number of studies thus far, and the results are mixed. For example, in his replication study of Juffs and Harrington (1995), Juffs (2005) used a series of working memory capacity measures to see if individual differences in WM affect learners' reading patterns during filler-gap processes. Specifically, Juffs wondered whether individual differences in WM have an influence on L2 learners' processing, especially when they were under greater processing pressures for computing a filler-gap reanalysis at *killed* in reading *Who_i did the police know t_i killed the pedestrian?* (see section 2.4.1). The measures included a L2 English reading-span test (Daneman & Carpenter, 1980), an L1 Spanish/Chinese/Japanese reading test (e.g., Osaka & Osaka, 1992), and a word-span test in L1 and in L2 (Baddeley et al., 1998). Juffs found no correlations between individual learners' reading patterns at the critical region and any of the WM measures. However, Juffs and Rodriguez (2015) later noted that non-significant associations between processing and WM in that study might have been due to the older methods employed in their tests (e.g., manually with cards).

Felser and Roberts's (2007) study that used a reading span test (Harrington & Sawyer, 1992) in learners' L2 also found no WM effect for L2 learners in that participants with both high and low working memory failed to present a position specific antecedent reactivation effect. In other words, although the presentation of the antecedent (filler)-matched picture is supposed to

facilitate participants' reaction (i.e., priming effect) only at a structurally possible gap position, those L2 learners showed a reactivation effect not only at the gap site, but also at the non-gap site, making it difficult to interpret the reactivation effect as a result of use of structural information. Note however, the low WM native English group in their study did not show any reactivation effect either, even in the no-gap control condition, making it difficult to interpret the differences between the L2 group and the NS English group with low WM. Thus, it could be that the cross modal (listening and visual) picture priming task they implemented might have been too difficult even for some of the native English speakers, which might have masked a potential role of individual differences in WMC (see also Nakano, Felser, & Clahsen, 2002, for similar native speaker results from the same experiment type).

Another study that found no WM effect on L2 processing is Felser et al. (2012). Although they did not include the WM results in their article, Felser et al. noted that they implemented a reading span test (L1- Daneman & Carpenter, 1980, and L2- Harrington & Sawyer, 1992) in their eye-tracking research on L2 processing of island constraints with proficient German-speaking learners of English. However, they dropped the WM results as no WM-related effects were found in their analysis.

On the other hand, Dussias and Piñar (2010) found some reliable effects of WM in their L2 English long distance filler-gap processing study with proficient L1 Chinese learners of English. They used a reading span test adopted from Waters et al. (1987) and Waters and Caplan (1996)'s reading span WM measure, which has a plausibility judgment of sentences as a processing component and recall of the last word of each sentence as a memory component. The test was given in the L2 (i.e., English). In the analysis, Dussias and Piñar divided each L1 group into two subgroups (high and low WM) by using the median WM score as a splitting point. They

found that the high WM learners, but not the low WM learners, showed evidence of filler-gap reanalysis in a similar way to the native English controls (see O'Rourke, 2013; Sturt et al, 1999 for similar findings in L1 processing research on processing of garden-path and filler-gap dependency).

In a more recent study by Hopp (2014), the author investigated the effect of individual differences in working memory and lexical decoding skills on processing of globally (offline) and (temporarily) ambiguous relative clause constructions in L2 English by German speakers. The author used a reading span test developed by Ariji, Omaki, and Tatsusa (2003), in that participants were asked to perform a segment-by-segment self-paced reading followed by an acceptability judgment about each sentence. The target to recall was one of the words in each sentence that was printed in capitals. Along with other results, Hopp found that higher WM L2 learners tended to prefer to attach the relative clause to a lower NP during their offline judgment test, suggesting that they employed phrase-structure-based parsing strategies more, whereas the lower WM L2 learners adopted chunking strategies and preferably attach the relative clause to a more discourse prominent higher NP. However, in his online eye-tracking reading experiment, Hopp found that learners' lexical decoding skills (as measured by a lexical decision task) were a better predictor for L2 learners' behaviors.

2.7. Research Questions

The main goal of the present study is to investigate how advanced early and adult ESL learners process structurally complex filler-gap dependency constructions in the L2 during online processing. By implementing an eye-tracking method, the focus of the study is primarily on 1) whether these learners are sensitive to the structural cues and making use of relevant

grammatical knowledge of island constraints in an appropriate way, and 2) whether individual differences in working memory capacity have influence on learners' application of the grammatical information in real time. This dissertation is guided by the following research questions:

1. The effect of age of acquisition/immersion on L2 processing

Do early ESL learners, adult ESL learners, and native English speakers show any different processing behaviors across the experimental conditions while processing filler-gap constructions in English?

1.1. Early gap: Use of active filler strategy and online application of island constraints

At the earliest possible gap site (Region1), do native speakers, advanced early and adult ESL learners show evidence for active gap creation in the non-island condition, but not in the island condition, thereby showing a reliable interaction of plausibility and island constraints? Specifically,

(A) Do all three groups attempt to postulate a gap and integrate the filler, thereby presenting a sensitivity to the plausibility manipulation in the non-island condition, as measured by longer reading times and more regressions in the implausible than in the plausible condition?

(B) Do all three groups avoid postulating a gap when encountering the verb inside the embedded relative clause island, thereby presenting no plausibility effect, thus showing evidence that they integrate detailed grammatical information of the relative clause island constraint into the parse?

1.2. Ultimate gap: The effect of filler-gap (re)analysis

At the ultimate gap site (Region3), do native speakers, advanced early and adult ESL

learners show evidence for filler-gap reanalysis in the non-island condition, but not in the island condition? Specifically,

(A) Do all three groups show a reanalysis effect in the non-island condition, displaying more difficulties in cancelling and revising their misanalysis in the plausible than in the implausible counterpart, as measured by longer RTs and more regressions in the plausible than in the implausible condition?

(B) Do all three groups show no reanalysis effect in the island condition, in consequence of no gap-postulation inside the relative clause island?

2. The effect of individual differences in WMC on L2 processing

How do individual differences in working memory capacity (WMC) influence the way native English speakers, early and adult ESL learners process filler-gap dependencies in L2 English?

2.1. Do differences in WMC of the native English speakers, early and adult ESL learners influence the way they respond to the plausibility manipulation in the non-island condition? Specifically, do lower WMC readers show any evidence that they are less sensitive to the plausibility manipulation in the non-island condition?

2.2. Do differences in WMC of the native English speakers, early and adult ESL learners influence the way they integrate the knowledge of island constraints into the parse? That is, do lower WMC readers show any evidence that they are more likely to postulate an illicit gap inside the island structure?

2.3. Do differences in WMC of the native English speakers, early and adult ESL learners influence the way they perform a reanalysis in the non-island condition at Region3? Specifically, do lower WMC readers show any evidence that they are less sensitive to the need for a reanalysis in the non-island condition?

CHAPTER 3: METHOD

3.1. Participants

A total of 52 advanced learners of ESL took part in the current study. They varied in terms of their age of arrival to the United States from two to thirty-one years. A group of 25 native English speakers also participated as controls. Data from three ESL learners had to be excluded from the analyses due to their overall lack of comprehension of the target sentences in the eye-tracking experiment, details of which are provided later in this chapter. In addition, data from one native English speaker displayed a high percentage of track loss consistently across the trials during eye-movement recording, and his/her data were removed from the analyses. As a result of these exclusions, the sample size was adjusted to 49 for the L2 learners and 24 for the English controls.

Most of the native English speaker controls ($N = 24$: 14 female & 10 male, mean age: 23.42, $SD = 7.79$, range: 18 - 51) were either undergraduate ($n = 15$) or graduate ($n = 7$) students studying at Michigan State University. The remaining two participants were recent MA graduates who were working as ESL instructors at the time of testing. The L2 participants had either L1 Chinese ($n = 13$) or L1 Korean background ($n = 36$), but the two L1s were collapsed to represent the L2 learners in this study. Crucially, Chinese and Korean are both wh-in-situ languages, thus providing a good testing ground to evaluate whether these ESL learners have acquired relevant L2 grammatical knowledge of wh-movement constraints that is not instantiated in their L1 syntactic representations, and if so, whether they can make use of the knowledge in order to construct detailed English filler-gap dependency constructions in real time.

To explore whether there is an age-related effect on L2 processing, the ESL learners were assigned into one of two groups based on their age of arrival in an English-speaking

environment, an early ESL group ($N = 21$) and an adult ESL group ($N = 28$). As discussed earlier, adult ESL learners in the current study were operationalized as those whose ages of arrival were after the age of 16 (i.e., $AOA > 17$; e.g., Johnson & Newport, 1989, 1991). The early ESL learners were operationalized as those who were immersed into an English-speaking environment from before the age of 12. The ESL learners' biodata and English learning background are provided in Table 1.

Table 1. *Biodata and English learning background of the ESL learners*

	Early ESL ($N = 21$)	Adult ESL ($N = 28$)
Age	23.67 (5.75)	31.11 (4.98)
Gender	18 female & 10 male	7 female & 14 male
L1 background	16 Korean & 5 Chinese	20 Korean & 8 Chinese
Age of Arrival	7.43 (2.25)	26.29 (3.05)
Length of Residence (yrs)	16.44 (5.39)	4.76 (4.65)

The adult ESL learners varied in their ages (range: 25 – 46 years old, $M = 31.11$, $SD = 4.98$), AOA (range: 18 – 31 years old, $M = 26.29$, $SD = 3.05$), and length of residence (LOR) (range: 4 months – 20 years, $M = 4.76$ years, $SD = 4.65$). They were mostly graduate students from a variety of majors enrolled in either master's ($n = 6$) and doctoral ($n = 18$) degree programs at Michigan State University, with the exception of four participants: Three participants were academic faculty teaching at the same institution, and one participant was a recent MA graduate working at an American corporation at the time of her participation. They had started learning English between ages seven and thirteen either as part of formal education at school or in the form of tutoring at private institutes in their home countries ($M = 11.61$, $SD =$

1.59). However, none of them had extensive English immersion experiences prior to their current residence in the United States. All the adult learners received their primary and secondary education in their home countries. As one way to estimate their overall level of English proficiency, the adult learners were asked through a language background questionnaire to report any type of standardized English proficiency test scores if available (See Appendix A). Twenty-four of the 28 participants responded, providing their TOEFL scores. With individual scores ranging from 94 to 116, the mean self-reported iBT TOEFL score of the adult learners was 104.54 ($SD = 7.05$), indicating that the adult ESL learners had, by and large, high levels of English proficiency.

The early ESL learners' ages of arrival ranged from two to eleven years old ($M = 7.43$, $SD = 2.25$), presenting a significant difference from the adult ESL learners in this respect, $t(47) = 23.84$, $p < .001$, $d = 7.033$. In contrast to the adult ESL learners, the early learners all received their primary and secondary education in the United States. Their ages ranged from 18 and 39 ($M = 23.67$, $SD = 5.75$). Fourteen participants were undergraduate students, and five participants were graduate students enrolled in either master's ($n = 4$) or doctoral ($n = 1$) programs at MSU. The remaining two participants were college graduates who were working at American corporations at the time of testing.

3.2. Materials

3.2.1. English proficiency measures

Two different types of proficiency measures were used to evaluate levels of ESL learners' L2 proficiency, self-rated English proficiency ratings obtained from individual learners and a web-based measure called LexTALE (Lexical Test for Advanced Learners of English, available from <http://www.lextale.com>). The details of each measure are discussed below.

Self-rated English proficiency. The L2 participants were asked to self-assess and indicate their levels of English proficiency for each language skill, on a scale from zero (not proficient at all) to 10 (near native-like) in the background questionnaire. The results of the two groups' self-rated proficiency are summarized in Table 2.

Table 2. *Self-rated English proficiency of the ESL learners for each language skill*

	Early ESL (<i>N</i> = 21)	Adult ESL (<i>N</i> = 28)	<i>df</i>	<i>t</i>	<i>p</i>	<i>d</i>
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)				
Listening	9.14 (.66)	7.89 (.69)	47	6.440	<i>p</i> < .001	1.865
Speaking	8.95 (.59)	7.43 (.88)	46.514	7.523	<i>p</i> < .001	2.031
Reading	8.76 (.83)	8.46 (.57)	47	1.481	<i>p</i> = .145	.421
Writing	8.52 (.60)	8.21 (.79)	47	1.502	<i>p</i> = .140	.442
Grammar	8.24 (.70)	8.46 (.58)	47	1.240	<i>p</i> = .221	.342
Overall	8.76 (.77)	8.14 (.71)	47	2.926	<i>p</i> = .005	.832

Overall, both groups showed fairly high proficiency ratings across different areas of English skills. The early ESL learners tended to assess their proficiency higher than the adult ESL learners in all language skills with an exception of the grammar part. Sets of independent-

samples *t*-test were carried out for each language skill and the overall proficiency to examine whether there is a reliable difference between the two groups. As shown in Table 2, there were significant differences between the two groups in Listening ($p < .001$), Speaking ($p < .001$), as well as in overall proficiency ratings ($p = .005$), in that the early ESL learners' proficiency ratings were significantly higher than the adult ESL learners. However, the two groups did not differ in reading ($p = .145$), writing ($p = .140$), and grammar ($p = .221$).

LexTALE measure. In addition to obtaining learners' self-rated English proficiencies discussed above, individual participants' general English proficiency was also measured independently using the LexTALE measure (Lemhöfer & Broersma, 2012). The LexTALE is an untimed lexical decision task designed primarily to evaluate vocabulary knowledge of highly advanced ESL learners. However, the test has also been found to be a good predictor of learners' general English proficiency as well (Lemhöfer & Broersma, 2012), thus allowing researchers to use the test result as an indication of learners' general L2 proficiency (e.g., Declerck, Lemhöfer, & Grainger, 2016; Mirdamadi & De Jong, 2015; Zufferey, Mak, & Degand, 2015). The test consisted of 3 practice and 60 vocabulary items adapted from Meara (1996): 40 items were low-frequency English words and 20 items were non-words (see Appendix B for the list of the items). The participants were instructed to indicate, using the computer mouse, whether each word on the screen is an existing word in English (by clicking a 'yes' button) or not a word in English (by clicking a 'no' button). See Figure 1.

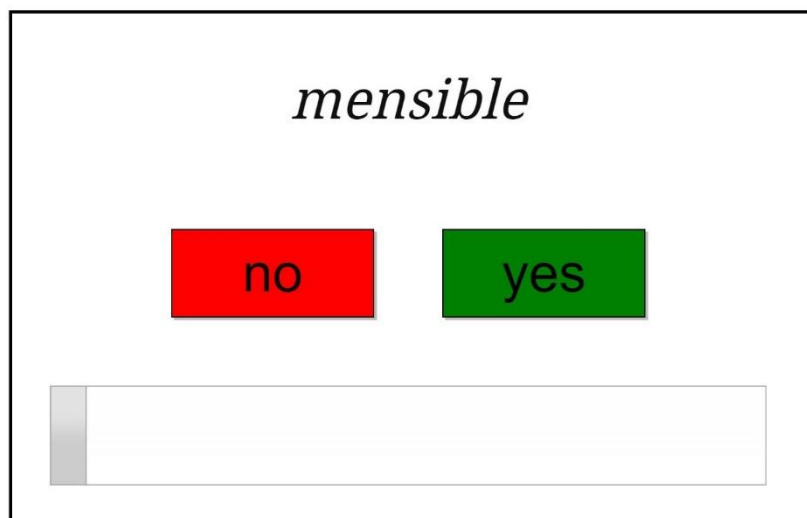


Figure 1. A screenshot of the LexTALE Test

The NS English group also took this test in order to estimate how close the levels of ESL learners' English proficiency are compared to native English speakers. The summary of the LexTALE results of the three groups is presented in Table 3.

Table 3. *LexTALE scores (in percent) of the native speakers and the ESL learners*

	<i>M</i>	<i>SD</i>	Range
NS English (<i>N</i> = 24)	91.95	5.80	78.75 - 100
Early ESL (<i>N</i> = 21)	88.69	6.79	75 - 100
Adult ESL (<i>N</i> = 28)	83.90	7.18	70 – 97.5

Note. Score in % = [(No. of English words correct /40*100)] + (No. of nonwords correct/20*100)]/2 (see, Lemhöfer & Broersma, 2012, for the scoring method)

According to the test developers, the average LexTALE score of a large group of advanced ESL learners in Lemhöfer & Broersma's (2012) study was 70.7 (in percent).¹⁰ Given this information, the two ESL groups in the current study showed fairly high LexTALE scores, presenting the mean scores of 88.69 and 83.90, for the early and adult ESL groups, respectively. A one-way analysis of variance (ANOVA) was performed to examine if there were any proficiency differences between the three groups including the NS English controls. The result showed a reliable difference among the groups, $F(2, 72) = 9.700, p < .001, \eta^2 = .217$. The follow-up Bonferroni post-hoc comparisons revealed that the scores of the adult ESL learner group were significantly lower than both the English control group ($p < .001$) and the early ESL group ($p = .044$). The scores of the early ESL groups were slightly lower than the NS English, but the difference was not significant ($p = .313$).

3.2.2. Working memory capacity measures

Participants' working memory capacity (WMC) was measured using two subsets of a battery of the automated complex WM span tests developed by Oswald, McAbee, Redick, & Hambrick (2015). Of the three different processing modalities in the test set—that is, operation span, symmetry span, and reading span—the participants took the operation span (O-Span) and the symmetry span (S-Span) tests. As discussed in the previous chapter, it has been a common practice to implement a reading span test (e.g., Daneman & Carpenter, 1980) both in L1 and L2 sentence processing research, mainly because the reading span test shares the same type of processing component (e.g., reading plausible or implausible sentences) with tasks in the

¹⁰ The current study set the LexTALE score of 70 percent as the prerequisite for the participation to ensure learners' high L2 proficiency. Scores of four adult L2 learners did not meet this requirement. Those participants received a portion of small payment and did not participate in the rest of the tasks.

reading-based language processing research. The current study, however, used the two non-verbal WM span tests above for the following reasons: First, although the L2 participants in this study were arguably high advanced ESL learners, administering a reading span test in their L2 could still present a proficiency confound at least for some learners (see Gass & Lee, 2011, for related discussion). In other words, a lower proficiency level in English will add additional burden in the reading, consequently affecting the size of the memory span. In this case, the observed WM span size could be the result not only of participants' WMC, but also of their lower English proficiency. Alternatively, the reading span test may be administered in participants' L1s. An attempt to create word-for-word translations into different languages, however, might inevitably cause some divergence among the three different language versions. For example, it would be difficult to match the length of each sentence or location of the critical area of the sentence that is directly related to the given task¹¹. Consequently, it may be difficult to maintain the test reliability across the different language formats. In addition, a confound with language proficiency could feasibly work in the other direction when giving a reading span test in L1; that is, for some early ESL learners the proficiency level of their L1 may be too low to take the reading span test in the L1.

Taking into account the concerns discussed above, the current study used the non-verbal O-span and S-span tests. Crucially, these tests were reported to be not only highly compatible with the reading span test in the same test set, but also reliable and valid as a measure of WMC (Oswald et al., 2015; see also Conway et al., 2005; Redick et al., 2012, for discussion on reliability and validity of the WMC measures). In this respect, Conway et al. (2005) noted that

¹¹ Note that depending on where the critical region that determines “plausible or implausible” or “making sense or not” is located in a sentence, the amount of pressures during reading may be different. Matching the location of this spot in two different languages will be very challenging especially considering the different word orders between English/Chinese and Korean.

although different measures of WMC are assumed to measure the same underlying construct (i.e., WMC) reliably well, it may be dangerous to rely only on a single WM measure, given the fact that different measures (with different processing modalities) would unavoidably tap into different areas of test-takers' abilities (e.g., mathematical ability for the O-span test). To overcome such shortcomings, Conway and colleagues suggested to run more than one WM span test and then use the composite scores on all the tasks as the measure of WMC. With this in mind, this study used these two tests and calculated the average scores of the two to obtain more reliable WMC measures (see also, Barrouillet & Lepine, 2005; Leiser, 2007, for empirical studies that used the composite scores as a WMC measure).

As discussed earlier, WMC is the magnitude of the memory storage that maintains a limited amount of information in the face of ongoing processing. Thus, the WM test consists of two parts; the processing component, and the storage component. The details of each span test are provided in turn.

O-span test In the O-span test, the processing component was judging whether a given arithmetic operation is correct or incorrect, and the storage component (i.e., target to recall) was remembering an English letter. Figure 2 illustrates a sequence of the O-span test. Participants first received a simple math problem (left), and then they were asked to judge within a limited amount of time whether the given answer was true or false (middle). Upon their response, they were presented with a to-be-recalled item, which remained on the screen for 800 milliseconds until the screen advanced to the next math problem. The test included a total of 30 operation-storage pairs, divided into 6 trials with two trials for each set size of four, five, and six. At the end of each set, participants were presented with a response screen, in that they were asked to

provide the recalled items in the same order they were presented. For example, when the set size was four, then they had to report 4 English letters in the same presentation order.

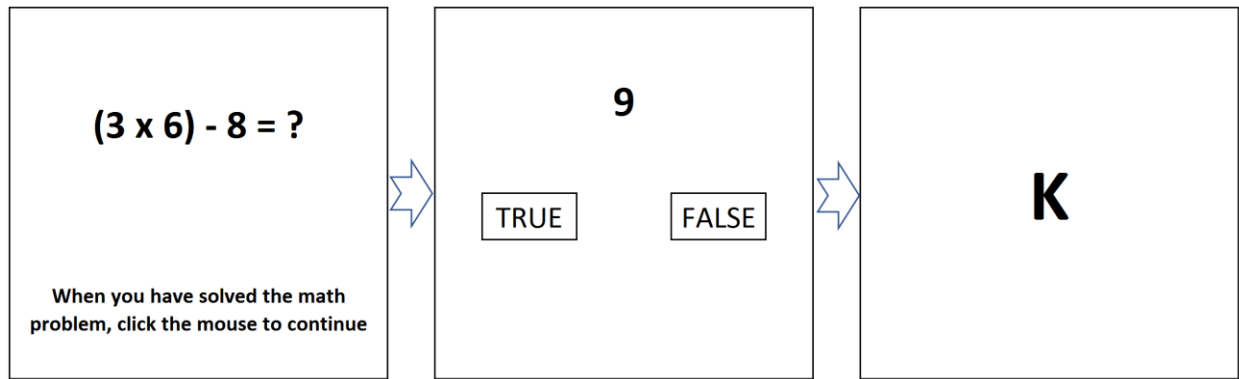


Figure 2. Processing and storage component of the operation span test

S-span test In the S-span test, the processing component was determining whether the two sides (left and right) of a given picture is symmetrical or asymmetrical, and the storage component was recalling a location of the red square in a 4 x 4 matrix, as illustrated in Figure 3.

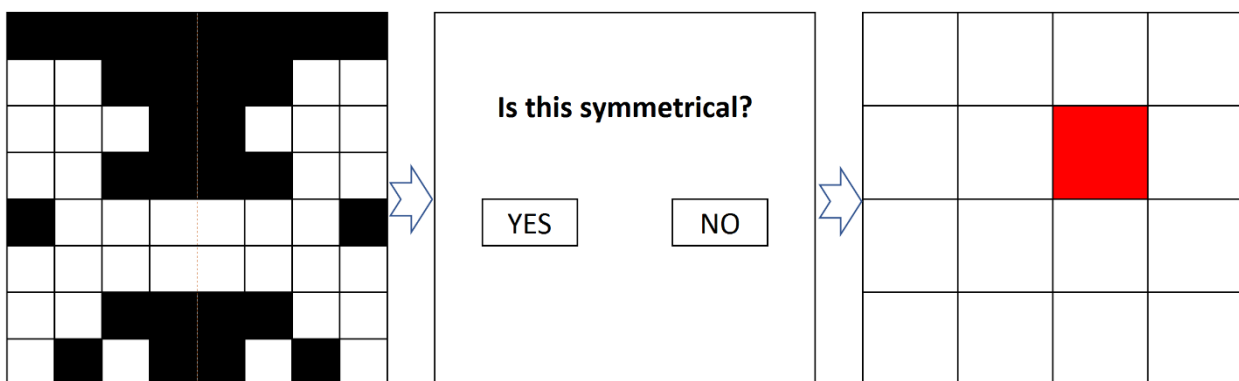


Figure 3. Processing and storage component of the symmetry span test

The test procedure was same as the O-span test. The S-span test includes a total of 24 symmetry-storage pairs, divided into six trials with two trials for each set size of three, four, and five.

The order of the two tests was counterbalanced so as to avoid any potential test fatigue and/or familiarity confounds. Half of the participants took the operation span test first, and the other half took the symmetry span test first.

For both tests, participants were informed during the practice session that in order for the researcher to be able to use their WM span scores (i.e., the number of recalled items), it is important that they score at least 85 percent on the processing part (i.e., math problems and symmetry judgments, respectively). This was to ensure they are indeed engaged in both processing and recall parts, rather than rehearsing the to-be recalled items (i.e., phonological short-term memory) without much effort on processing. By the test design, participants were provided with their current processing score on the screen after each set so that they could keep balance between the two parts as the test progresses. All participants showed acceptable processing performance. The processing scores ranged from 80 to 100 percent accuracy in the O-span test, 75 to 100 percent in the S-span test¹².

¹² The test development team addressed through their FAQ section on their website (<http://englelab.gatech.edu/faq.html>) that they generally remove participant data when accuracy on the processing part (e.g., math problem) is below 85 percent, although they also acknowledged that the 85 percent threshold was an “arbitrary rule of thumb”. They commented, “it is not so much the actual accuracy that matters but more so if the participant was attending to the processing trials,” In this study, there were eight participants whose math accuracies ranged between 80.00 and 83.33 percent in the O-span test, and there were six participants whose symmetry accuracy ranged between 75 to 83.33 percent in the S-span tests. The observation of those participants’ WM span scores showed relatively much lower than the averages of the group to which they belong, suggesting that they did not take much benefit of scoring high WM span scores at the cost of less attention to the processing part. Rather, it would be more likely that their processing abilities (e.g., math ability) might have been slightly lower thus making the test(s) more demanding for both parts. For this reason, I decided not to exclude their data.

3.2.3. Main experiment: Eye-tracking reading

3.2.3.1. Reading materials

The eye-tracking reading materials consisted of 7 practice, 28 target, and 54 filler items. The target sentences were developed based on the materials used in Omaki and Schulz's (2011) self-paced reading experiment, primarily to investigate whether the early and adult ESL learners with wh-in-situ L1 backgrounds (i.e., Chinese and Korean) in this study can make use of relevant syntactic island constraints in L2 English, thus avoid postulating a gap within the relative clause island structures. Each target sentence had four experimental conditions in a 2 x 2 Latin square design, with plausibility (i.e., plausible and implausible) and island (non-island and island) manipulation (see Appendix C for the target sentences). Four different experiment subsets were then created with each subset including only one of the four versions of each trial, such that individual participants received only one version of each trial. The two plausibility conditions and the two island conditions were counterbalanced across the items in each subset, including seven target trials for each experimental condition. Examples of the four experimental conditions are illustrated in (27) through (30), in which the two regions in bold indicate the two critical regions (Region1 and Region3), and the two underlined regions are the two spillover regions (Region2 and Region4).

(27) [plausible, non-island]

The book that the journalist **wrote *ti*** fairly regularly **about *ti*** **was** named for an explorer.

(28) [implausible, non-island]

The city that the journalist **wrote *ti*** fairly regularly **about *ti*** **was** named for an explorer.

(29) [plausible, island]

The book that the journalist who **wrote** fairly regularly **mentioned** *ti* **was** named for
an explorer.

(30) [implausible, island]

The city that the journalist who **wrote** fairly regularly **mentioned** *ti* **was** named for
an explorer.

As shown in the examples, the sentences in the two plausibility conditions were all identical except in the filler nouns (e.g., *the book* and *the city*) that differed for the plausibility manipulation at the first verb *wrote* (i.e., *wrote the book* and *wrote the city*). The length and word frequency of the filler nouns were matched between the plausible and implausible condition. The length of those nouns ranged from 4 (e.g., *book*) to 9 (e.g., *crocodile*) characters, and the mean length was matched to 6.214 characters for both plausibility conditions ($SD_{\text{plausible}} = 1.49$, $SD_{\text{implausible}} = 1.52$). The word form frequency was checked using the American English Subtitles (SUBTLEXus) corpus data (Brysbaert & New, 2009): the mean frequency of the plausible nouns was 76.58 per million ($SD = 107.44$), and the mean frequency of the implausible nouns was 73.76 ($SD = 106.08$) per million, which were not statistically different from one another, $t(54) = .099$, $p = .922$. The sentences in the island condition, as in (29) and (30), included an additional relative clause embedded in another relative clause, which is preceded by the relative pronoun *who* (e.g., the book [_{RC} that the journalist [_{RC} who wrote...]]). The relative pronoun *who* in this case forms a relative clause island because the position that the relative pronoun occupies is the only place that a to-be-raised constituent can make a legal movement. The presence of *who*, however, blocks such movement, as illustrated in (4) in Chapter 2. As a result, the syntactic representations of the sentences in the island condition do not have a silent copy of the wh-trace

or empty category within the embedded relative clause. In other words, the parser must not postulate a gap at the verb *wrote* in (27) and (30).

3.2.3.2. Areas of Interest for analyses

First critical region (Region1) The first critical region (Region1) includes the first verb that the parser encounters in the sentence in all four experimental conditions, which, according to the active filler hypothesis, is the earliest structurally possible gap site where the parser can (temporarily) retrieve the filler from the WM and make syntactic and semantic analyses of its goodness of fit in the non-island condition in (23) and (24). In this respect, Region1 is the site where a plausibility effect can take place by virtue of integrating the filler into the grammatical gap and analyzing it as the object of the verb *wrote* at the moment of processing: One yields a plausible interpretation (i.e., *the journalist wrote the book*), and the other yields an implausible interpretation (i.e., *the journalist wrote the city*). The implausible interpretation obtained in (24) may challenge the interpretive processes of the parser at the moment, likely resulting in a plausibility effect with elevated RTs and/or more regressive eye-movement patterns at the verb *wrote* in (24), compared to its counterpart in (23).

On the other hand, no such plausibility effect should be expected in the island condition in (25) and (26), under the assumption that the parsing is guided by fully detailed syntactic information. The reason is that the verb (i.e., *wrote*) is located inside another relative clause (i.e., [*the journalist [who **wrote** fairly regularly]...*]) in the island condition, and the filler cannot be moved out of the relative clause island by the grammar, as shown earlier in (4). In other words, postulating an object gap on the verb *wrote* inside the relative clause island is not possible because there is no empty category or object trace posited within the island in its structural

representations. For this very reason, the verbs located in Region1 were always optionally transitive verbs (e.g., *wrote*, *read*, *advise*, *perform*), so that the target sentences in the island condition can eventually be grammatical. The absence of a gap postulation at the verb *wrote* in (25) and (26) would consequently contribute to a no plausibility effect, with relatively comparable reading patterns between the plausible and implausible sentences. In contrast, if the ESL learners' knowledge of the movement constraint in English is deficient, and/or if they are not capable of deploying the syntactic information during in real time for some reasons (e.g., lower WMC), thus relying on lexical-thematic and semantic information instead as the SSH would predict, then they may display the plausibility effect in the island condition as well, forming filler-gap dependencies between the filler and the verb inside the relative clause island. Taken together, a significant 2-way interaction (i.e., plausibility x island) may be observed at Region1 only if the participants utilize the relevant movement constraints of English at the right moment during reading.

Second critical region The second critical region (Region2) is where 1) the parser is supposed to cancel its previous analysis computed in the first critical region (Region1), followed by a filler-gap reanalysis in the case of non-island sentences, and 2) the parser creates a gap for the first time—assuming no gap postulations inside the relative clause island—on the verb *mentioned* in the case of island sentences. As shown above, Region2 consists of two words, the preposition *about* and the adjacent auxiliary verb *was* in the non-island sentences, and the verb *mentioned* and the same auxiliary verb *was* in the island sentences, in that the fillers (*the book & the city*) are the object of the preposition *about* in the non-island condition, and is the object of the second verb *mentioned* in the island condition. In the non-island condition, the parser must initiate the reanalysis as soon as it recognizes no presence of an argument of the preposition

about (i.e., *about t_i was*). In doing so, the parser first needs to cancel its previous analysis of the filler as the object of *wrote* at Region1, and then take the filler as the object of *about*. In this regard, Pickering and Traxler (1998) claimed, based on their L1 eye-tracking studies on processing of garden-path sentences, that the parser may be more taxed when it has to withdraw its earlier analysis that is more plausible, because the level of (L1) readers' commitment to a semantically plausible interpretation is relatively much deeper. With this in mind, the participants may display a type of garden-path effect at Region2 *about was* in the plausible condition in (23) with longer RTs and/or more regressive eye-movement patterns, compared to its counterpart in (24). On the other hand, no such different reading pattern should be found between the plausible and implausible sentences on *mentioned was* in the island condition, when taking into account the assumption that there is no gap postulation up until Region2 (i.e., no reanalysis).

Spillover Regions (Region2 and Region4) Previous research has shown that L2 learners may display qualitatively similar processing patterns that are comparable to those of native speakers of the target language, but that some expected effects such as the plausibility effect or filler integration effect may occur with some delays in L2 processing, possibly due to learners' slower and less efficient processing abilities (e.g., Dekydtspotter et al., 2006; Williams et al., 2001). In addition, it may be possible that the participants, especially the ESL learners with lower WMCs, may be slower or less efficient in releasing the filler information on the gap sites, displaying somewhat delayed plausibility effect at a later region. Taking these into account, the two regions that come right after the two critical regions (Region1 and Region3) were also analyzed to examine any potential spillover effects. The spillover region (Region2) that comes immediately after the the first critical region (Region1) always included two adverbs (e.g., *fairly*

regularly) in all four experimental conditions. Another spillover region (Region4) next to Region3 was always [passive participle + preposition, e.g., *named for*] for all four experimental conditions.

3.2.3.3. Eye-tracking reading task design and procedures

The eye-tracking reading task was programmed using Experiment Builder (version 1. 10. 1630), an experiment programming software for the EyeLink 1000 Desktop-mounted system (SR Research Ltd. <http://www.sr-research.com>) used in the current study. The experiment consisted of four pseudorandomized blocks with each block including seven to twelve sentences intermixed with the target and filler trials. The participants were able to take a short break between the blocks, so as to minimize task fatigue. At the outset of each block, the participants went through the calibration and validation process with the researcher to setup the camera. In addition, drift correct was implemented before each trial appeared on the screen to ensure the accuracy of the eye-movement recording. The participants read each sentence on a 19-inch computer screen while the eye-tracker in front of the screen collected their eye movements on the sentence. All sentences fit on a single line. The font type of the text was Serif and the size of the text was 19 for all trials. The participants all had normal or corrected to normal vision at the time of participation.

The sentences (e.g., *The wall that the soldier throw quite forcefully toward was covered with moss.*) were followed by a comprehension check in the form of a true or false question, in which the participants indicated by pressing one of the two designated buttons whether a statement on the question screen (e.g., *The wall was covered with moss*) was true [green button] or not [red button] based on their comprehension of the sentence. The inclusion of the

comprehension questions was to ensure that the participants paid attention to the reading (for meaning), and also to monitor whether participants comprehended the complex target sentences well. As mentioned earlier, the data from three participants were excluded from the analyses due to their lack of comprehension on the target sentences. A comprehension accuracy score of 70 percent on the target sentences was set as a cutoff. Two adult ESL participants who scored 57.14 and 60.71 percent respectively were thus removed from the analyses. In addition to checking the comprehension scores, the researcher conducted a brief interview with each participant after they completed all required tasks. The participants were presented with four target sentences they had read during the reading task, with one sentence for each experimental condition. They were asked to paraphrase those sentences verbally and explain to the researcher how they comprehended those sentences during the task. Two adult ESL participants (one was the same participant whose comprehension score was below 70%) reported that they could not understand the sentences well most of the time, especially those in the island condition. Both participants mentioned that they considered them as not grammatical or typos and just tried to somehow catch the meaning to answer the questions. The data from those participants were also excluded from the analyses.

3.2.3.4. Eye-tracking dependent variables

The eye-movement measures examined in the current study are as follows: first fixation duration, first-pass reading time, and first-pass regression, regression path duration, and total reading times. An illustration of eye-movements during reading is provided in Figure 4.

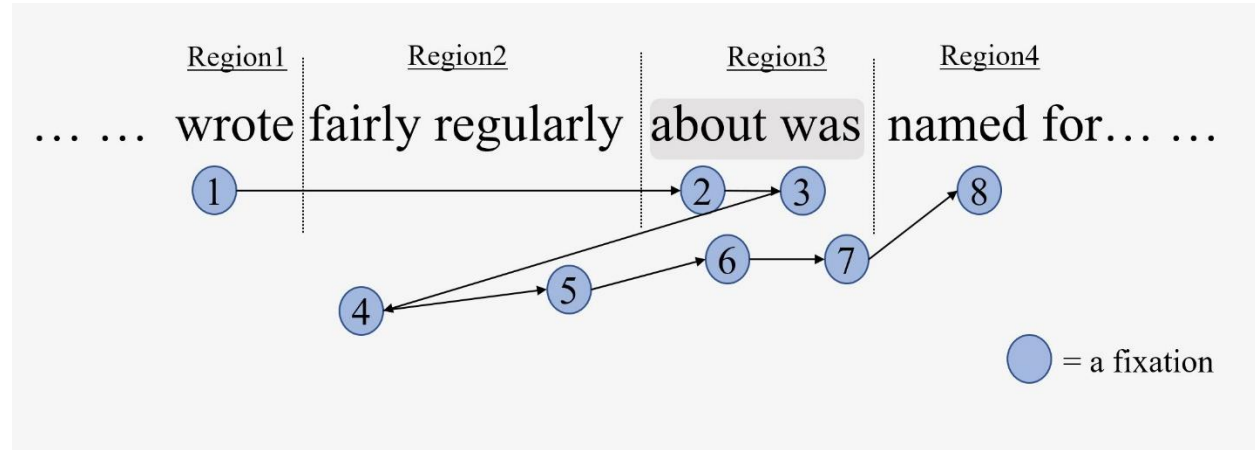


Figure 4. An illustration of eye-movements during reading

Each circle presents a fixation and the numbers inside indicate the order of the fixation occurrences. Note that in an actual recording, the fixations are marked on the text in normal circumstances. The fixations in the example above were placed below the text intentionally for demonstration purposes.

First fixation duration First fixation duration refers to the duration of the first entered fixation in an interest area (or word), provided that there is no fixation in later regions marked prior to the current first fixation (i.e., the first fixation during first-pass). For example, the first fixation at Region3 in Figure 4 is ②, but the first fixation duration at Region2 is not the duration of ④, but zero (also referred to as ‘skip’), because there are fixations in a later region before ④

is fixated¹³ (i.e., ② & ③ at Region3).

First-pass reading time (First-pass RT) First-pass reading time (also referred to as gaze duration, especially when an interest area consists of a single word; Roberts & Siyanova-Chanturia, 2013) is the sum of all eye fixations in an interest area, from its first entrance until the eye leaves the interest area in any direction either to the left (regressive) or right (progressive) boundary of the area, provided that, like first fixation duration, there is no fixation in later regions recorded before the first entering fixation in the current interest area (i.e., the first fixation during first-pass). Thus, the first-pass reading time at Region3 includes the fixations of ② and ③.

First-pass regression (probability) First-pass regression is defined as the percentage of regressive eye movements from the interest area to a preceding area that occur during the first-pass reading. Unlike the other eye-tracking measures that provide processing time course measures (in milliseconds), the first-pass regression provides binary data, in that it gets “1” if there was a regressive movement out of the area during first-pass, and it receives “0” for no regression during first-pass. At Region3, there is a backward movement from ③ to the ④ at Region2, consequently, the first-pass regression probability will get one point in this area. Regressive eye-movements during reading often indicate some processing difficulties at the moment, for example, in associating currently processed unit with previous parts of the sentence (Clifton et al., 2007; Vasishth & Drenhaus, 2011).

Regression path duration Regression path duration refers to the sum of all fixations recorded from its first entrance to an interest area up until the eye exits the right boundary of the interest

¹³ In this case, the first fixation duration at Region2 is zero, and the sum of ④ and ⑤ will be recorded as a second pass reading time, not first-pass reading time at this region.

area (i.e., progressive eye-movements passing the interest area). When there is a regressive eye movement out of the current interest area during the first-pass reading (i.e., first-pass regression = 1), regression path duration also includes the time spent at earlier regions on the left side (for left to right reading as in English) after the regression is initiated. Consequently, the regression path duration at Region3 is the sum of the fixations of ②, ③, ④, ⑤, ⑥, and ⑦. When the region involves no regression, the regression path duration is as same as the first-pass reading time.

Total reading time (Total RT) Total reading time (also referred to as total duration) is the sum of all fixations recorded within an interest area, indicating how much total time a reader spent at the region during the entire course of reading (e.g., ② + ③ + ⑥ + ⑦ at Region3). which is generally considered as a late measure that may reflect readers' later processes related to text comprehension and information reanalysis during later stages of processing, and recovery from misanalysis and/or reanalysis (Clifton, et al., 2007; Roberts, et al., 2012).

Of the five eye-tracking measures discussed above, the first three measures, namely first fixation duration, first-pass reading time, and first-pass regression, are generally considered to reflect readers' early stage of processing, likely at the level of morphology (e.g., lexical access) and syntax (e.g., integration of words into phrases). The regression path duration and total reading time are known to index readers' processes at later stages of processing, related to text comprehension and information reanalysis, and a recovery from misanalysis/reanalysis.

3.3. Overall procedures

All data were collected in a laboratory equipped with the EyeLink 1000 Desktop-mounted system. Individual participants attended a single 40 - 60 minute session. Upon arrival, participants completed the following tasks in this order: [1] LexTALE proficiency test, [2] eye-tracking reading task, [3] one set of the WM span test, [4] background questionnaire, [4] the remaining WM span test set, and [5] a brief oral interview with the researcher for a comprehension check. All participants were paid 20 US dollars for their participation.

3.4. Data Analysis

3.4.1. Preparation of the data for analyses

Eye-tracking data trimming In preparation of the online reading data for analyses, any fixations that were shorter than 80 milliseconds were automatically filtered out before extracting the data set from the eye-tracking system¹⁴. The data where the participants skipped the area either during the first-pass reading or during the entire reading process (i.e., no fixation recorded on an area of interest) were replaced with missing values. Additionally, for each measure (except the first-pass regression), RTs that were beyond the 2.5 SDs from individual participants' mean RTs on the same region were also replaced with missing values, which all together affected 2.44 percent of the entire data (approximately 3.37% of the control group, 2.09% of the early ESL,

¹⁴ Note that it is a common practice in eye-tracking reading research to eliminate extremely short (generally < 80ms) or long (generally > 800ms) fixations as they are considered as noise rather than a reflection of readers' cognitive processes (see Rayner and Pollatsek, 1989). These thresholds have been applied in many L1 reading research, and recently carried over to L2 reading research. However, applying the same thresholds used in L1-based research to L2 reading study may be potentially problematic, especially the threshold for the longer fixation, considering L2 learners' generally slower reading speed. Furthermore, the target sentences used in this study were all structurally complex even for native speakers. For these reasons, the removal of fixations larger than 800ms was not performed in this study.

and 1.90% of the adult ESL group data). After those data trimming processes, individual participants' mean RTs and the first-pass regression probability ratios (for by-subjects analysis: F_1), and mean RTs and first-pass regression ratios on each target sentence (for by-items analysis: F_2) were calculated for each measure across the interest areas for the main analyses.

An initial inspection of the raw eye-tracking reading data showed that they were not normally distributed for the most measures across the interest areas, displaying a range of (mostly positive) skewness across the experimental conditions and groups. Therefore, a log transformation was performed on each measure to correct this issue. The transformed data largely met the normality and the homogeneity of variance assumptions for ANOVA¹⁵. However, the first-pass regression data (both F_1 and F_2) showed a wide range of violations across the regions. The regression path duration and total reading time data at Region3 were also found to violate both the normality and equal variances assumptions across the experimental conditions and groups. To resolve this problem, the following analyses were used alternatively: For the analysis of the first-pass regression data, a logistic random effects regression model with the option of the robust covariance matrix estimation was performed using the raw binary regression data (1 = regression, 0 = no regression). In this model, both subjects and items were taken as random effects. For the analysis of the regression path duration and total reading time data at Region3, a set of nonparametric Wilcoxon signed ranks tests were used instead for each group separately, for both by-subjects and by-item analyses. Lastly, the Greenhouse-Geisser correction was applied in case when the sphericity test showed a violation (Field, 2009).

¹⁵ Normality was tested using the Shapiro-Wilk goodness-of-fit test (Larson-Hall, 2010) supplemented with the normal Q-Q plots. Homogeneity of variance was checked using Levene's test.

Working memory span scores For the scoring of the WM capacity of individual participants, the current study adopted a partial-credit loading scoring method, which is one of the most widely used scoring methods for span measures (Conway et al., 2005; see also Juffs & Harrington, 2011 for some other scoring methods used in L2 research). A partial-credit loading score calculates the sum of all correctly recalled items in the right order across the trials. Consequently, the maximum raw score was 30 for the O-span and 24 for the S-span test. An initial correlation analysis was performed to examine how reliable participants' performance was in the two tests. The result showed a significant moderate to strong correlation between the two tests, $r(71) = .558, p < .01$. Individual participants' two span test scores were inspected before calculating the composite scores for the main analyses. There were two participants (one early ESL and one adult ESL) who showed extremely contrasting results between the two tests—100% in the O-span, but 33.33 % in the S-span, and similarly 100 % in the O-span, and 54.17% in the S-span—thus making the results rather unreliable. The WM data of those learners were replaced with missing values. A subsequent correlation analysis was performed again after excluding the two participants. The results revealed a strong positive correlation between the two tests, $r(69) = .727, p < .01$, reflecting a high reliability between the two tests (cronbach's alpha = .840). Such high reliability also provided a reasonable basis for using the composite WM span scores. The summary of the results of the two WM-span tests are provided in Table 4.

Table 4. *Summary of the WM span test results in percent*

	O-span		S-span	
	<i>M (SD)</i>	Range	<i>M (SD)</i>	Range
NS English (<i>N</i> = 24)	75.97 (15.42)	43.33 - 100	71.18 (14.32)	41.67 – 100
Early ESL (<i>N</i> = 20)	74.83 (13.40)	50 - 93.33	73.75 (12.32)	54.17 – 91.67
Adult ESL (<i>N</i> = 27)	78.27 (14.86)	50 - 100	77.47(12.62)	41.67 - 100

Overall, the adult ESL group presented slightly higher WM span scores than the other two groups on both tests. The comparison of the three groups with a pair of ANOVAs for each measure, however, showed no significant differences among the groups, $F(2, 70) = .342$, $p = .712$, $\eta_p^2 = .010$, in the O-span, and $F(2, 70) = 1.479$, $p = .235$, $\eta_p^2 = .042$, in the S-span test, indicating that the three groups did not statistically differ from one another with respect to their level of working memory capacities (WMCs). To minimize multicollinearity (Marquardt, 1980), each span score of individual participants was standardized (i.e., Z-score), and then the two Z-scores were averaged to obtain the composite WM span scores (hereafter, WM span scores) for each participant (e.g., Barrouillet & Lepine, 2005; Leiser, 2007).

3.4.2. Main Statistical analyses

Effect of age of immersion To investigate the extent to which the reading behaviors by the early and adult ESL groups converge on or diverge from those by the native English speakers, statistical analyses of participants' eye-movements were conducted on the four regions of interest, namely the earliest gap (Region1) and the following spillover region (Region2), and the ultimate gap (Region3) and the following spillover region (Region4), respectively. The

analyses included both by-subjects (F_1) and by-items (F_2) analyses. As a preliminary step, a 3-way ($3 \times 2 \times 2$) mixed design ANOVA was carried out for each time-course eye-tracking dependent measure (first fixation duration, first-pass RT, regression path duration, and Total RT) at each interest area, with group as the between-subject variable and the two item conditions—plausibility and island constraints—as the within-subject variables. As addressed above, participants' mean first-pass regression data were not appropriate for ANOVA analysis. Thus, a mixed effects logistic regression analysis was used to model the raw binary outcome variables (1= regression, 0 = no regression), with group, plausibility, and island constraint as fixed effects, and subjects and items as random effects. When this preliminary analysis presented any significant group related interactions, then a 2 (plausibility) \times 2 (island constraints) repeated measures ANOVAs for reading time measures, and a mixed effects logistic regression model for the first-pass regression measure, were carried out separately for each group, in order to better interpret the significant interactions and get a clearer picture of different reading patterns among the three groups. Lastly, when the follow-up analysis of each group displayed a significant interaction between the two factors (i.e., island constraints and plausibility), subsequent planned paired sample *t*-tests were performed for each island condition to further examine how plausibility functioned differently across the two island conditions.

WMC effect To examine the effect of different working memory capacities of individuals on their processing of filler-gap dependency constructions, a series of repeated measures analysis of covariance (ANCOVAs) was performed for each time course eye-tracking dependent variable, separately for each group; with the two item conditions—plausibility and island constraints—as the within-subject variables, and the WM span scores as a continuous covariate. For the analysis of first-pass regression probability, the logistic random effects

regression model was applied as before, with the two item conditions—plausibility and island constraints—and the WM span scores as fixed effects, and subjects as a random effect.

The results were reported when there was a significant WM main effect and/or a significant WM-related interaction. In interpreting the results of the reading time data analysis, the parameter estimates (beta coefficients, $\hat{\beta}$) for the WM span scores were examined across the four experimental conditions to identify the directions of the relationship between the WM span scores and the dependent measures. When a value of the beta coefficient is positive (i.e., $\hat{\beta} > 0$), this means that the outcome variable increases by the amount of the $\hat{\beta}$ as the standardized composite WM span score increases by one unit. On the other hand, a negative coefficient values (i.e., $\hat{\beta} < 0$) indicates that the reading time decreases by the amount of the $\hat{\beta}$ as a function of one unit increase in the WM span score. Generally, the beta coefficient indicates an approximate amount of a predicted change in the unit of the dependent measure (e.g., milliseconds in case of reading times). However, because all the dependent measures in the current study were log-transformed as discussed earlier, the obtained beta coefficients presented the approximate amount of changes in percentage ($10^{\hat{\beta}} \times 100\%$). As a result, a beta coefficient value that is larger than 1 ($\hat{\beta} > 1$) indicates a positive relationship and a beta coefficient value that is less than 1 ($\hat{\beta} < 1$) indicates a negative relationship.

In interpreting the results of the first-pass regression analysis, an odds ratio (OR) was calculated to diagnose the relationship between WM span and first-pass regression. An OR is an “indicator of the change in odds resulting from a unit change in the predictor” Field (2009, pp. 270). Thus, it provides the probability of the occurrence of making a first-pass regression when the WM span score decreases or increases by one unit. Similar to the beta coefficient, a positive OR (i.e., $OR > 1$) signals an increase in probability of making first-pass regressions by the

amount of the OR value as the WM span scores are increased by one unit, and a negative OR (i.e., $OR < 1$) indicates a decrease in probability of making first-pass regressions as the WM span score decreases. Note, however, the information that the beta coefficients and the ORs provide (i.e., the relationships between the dependent measures and the WM span as the predictor) is limited only to each experimental condition, thus making it rather difficult to interpret interactions of the WM span scores and the other factors in some cases. For example, if the beta coefficient for the WM span scores in the [non-island, implausible] condition is positive (e.g., $\hat{\beta} = 1.5$), and if it is larger than the positive $\hat{\beta}$ value found in the [non-island, plausible] condition (e.g., $\hat{\beta} = 1.25$), it does not necessarily entail that higher WM participants' reading time in the [non-island, implausible] condition was longer than their own reading time in the plausible counterpart, because what $\hat{\beta}$ and OR indicate is relative amount of changes between higher WM participants and lower WM participants in the same experimental condition. Consequently, depending on what the lower WM participant's reading times in the two experimental conditions were, reading patterns of the higher WM participants in this example could go either direction. Taking into account this shortcoming in interpreting interactions among the factors, in cases when the analysis of the parameter estimates and ORs for the WM span scores did not provide enough information to interpret a significant WM associated interactions, the group was divided into two subgroups based on their WM span scores, namely the higher WM and lower WM subgroups, and the descriptive statistics for those two subgroups were created with their raw data to supplement the analysis.

CHAPTER 4. RESULTS

4.1. Comprehension Accuracy

The comprehension accuracy of the NS of English and the ESL learners are summarized in Table 5. The total comprehension score includes participants' responses on both the filler and target sentences, and the target comprehension score includes only the responses on the target sentences.

Table 5. *Mean comprehension accuracy in percent in the reading task*

	Total comprehension		Target comprehension	
	<i>M (SD)</i>	Range	<i>M (SD)</i>	Range
NS English (<i>N</i> = 24)	92.21 (3.80)	84.21 – 98.68	90.62 (6.39)	75.75 – 100.00
Early ESL (<i>N</i> = 21)	90.65 (3.85)	80.00 – 96.05	88.77 (5.09)	82.14 – 100.00
Adult ESL (<i>N</i> = 28)	87.04 (5.39)	73.08 – 96.05	86.35 (7.65)	71.43 – 100.00

Note. The accuracy scores were all rounded off to two decimal digits.

Overall, both the NS English group and the two ESL groups showed high rates of accuracy on both accuracy measures, with the mean accuracy scores of the three groups ranging from 87.04 (Adult ESL) to 92.21 (NS English) percent in the total comprehension, and from 86.35 (Adult ESL) to 90.62 (NS English) percent in the target comprehension. This indicated that the participants paid attention to the reading, and they were able to understand the structurally complex target sentences correctly. A one-way ANOVA was run on each comprehension measure to examine if there was any significant difference among the groups. The analysis of the total comprehension scores showed a reliable difference among the three groups, $F(2, 70) = 9.133, p < .001, \eta^2 = .208$. The following Bonferroni post hoc analysis revealed that total comprehension score of the adult ESL group was significantly lower than both the early ESL

group ($p < .021$), and the NS English control group ($p < .001$). The early ESL learners' overall accuracy was slightly lower than the native English speakers, but the two groups were not statistically different from one another ($p = .745$). The target comprehension scores of the adult ESL learners were also slightly lower than the other two groups, but the result showed no significant difference among the groups, $F(2, 70) = 2.752$, $p = .071$, $\eta^2 = .073$.

4.2. Overview of reading profiles

Prior to the main analysis, a series of fixation duration-based heatmaps were created with the raw fixation data¹⁶ for each target structure and for each group, in order to review overall reading profiles of the three groups on the target sentences. Figure 5, Figure 6, and Figure 7 provide a set of heatmaps of the NS English group, early ESL group, and the adult ESL group, respectively. Each map reflects participants' aggregated fixations recorded on the same target structure type across the trials (e.g., all trials in the non-island, plausible condition), and the text included in the map is one of the trials selected from a group of the same structure. In order to make comparisons possible across the experimental conditions and across the groups, the same maximum fixation value of 1200 milliseconds was applied across the maps (see the legend on the right bottom), so that the same color schemes could be applied to reflect the same fixation durations¹⁷. In the heatmaps, a more reddish color represents more aggregated fixation durations on a spot, which may indicate more processing burden during reading¹⁸.

¹⁶ The trimmed reading time data, specifically the participants' raw reading times that were over 2.5 SDs from their mean RTs, are not included in the maps.

¹⁷ When the peak fixation values (i.e., a largest single fixation) between maps differ, different colors are used to present the fixation duration, making it difficult to make a direct comparison between maps across the experimental conditions (see EyeLink Data Viewer User's manual for more information).

¹⁸ Note that the number of participants differed between the groups (24 NS English, 21 early ESL, and 28 adult ESL), meaning that the number of trials reflected in the maps was different

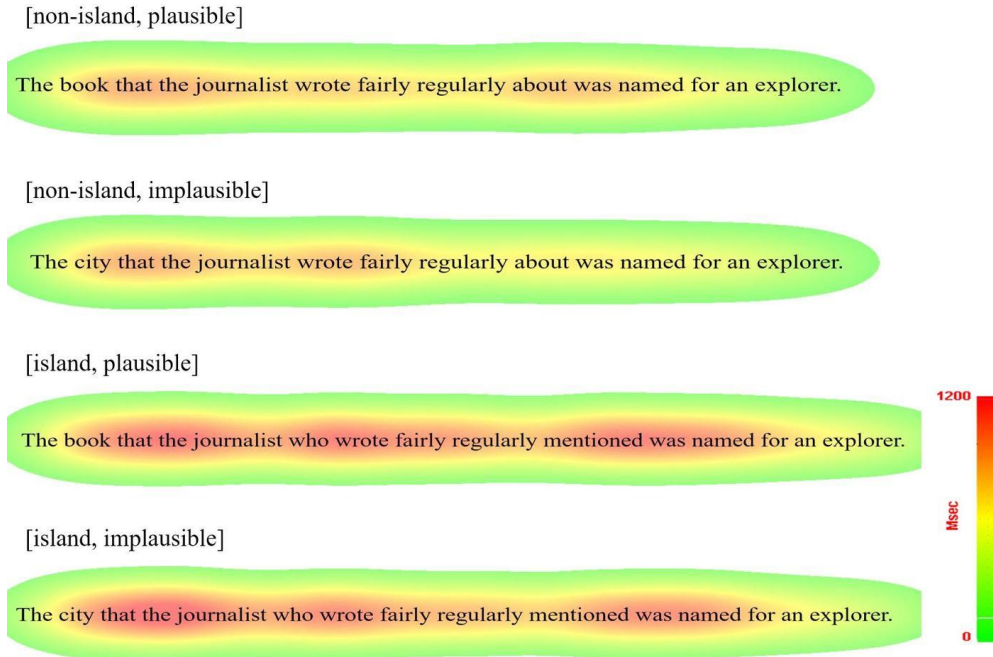


Figure 5. Fixation map: Reading profiles of the NS English speakers

As shown in Figure 5, the NS English group displayed slightly more aggregated fixations at the early gap Region1, the verb *wrote* in the [non-island, implausible] condition, compared to its counterpart in the [non-island, plausible] condition. Such reading pattern was more clearly shown in the two ESL groups in Figure 6 and Figure 7, in that both the early and adult ESL learners tended to spend more time when encountering implausible interpretations (i.e., *wrote the city*), indicating greater processing difficulties at this point over the course of reading.

among the groups, with the adult ESL group including the most ($n = 196$ trials per structure) and the early ESL group including the least trials ($n = 148$ trials per structure). Different sample size of the groups therefore must be taken into account when comparing the maps between the groups, although this should not be the case when comparing the maps of different structures within the group.

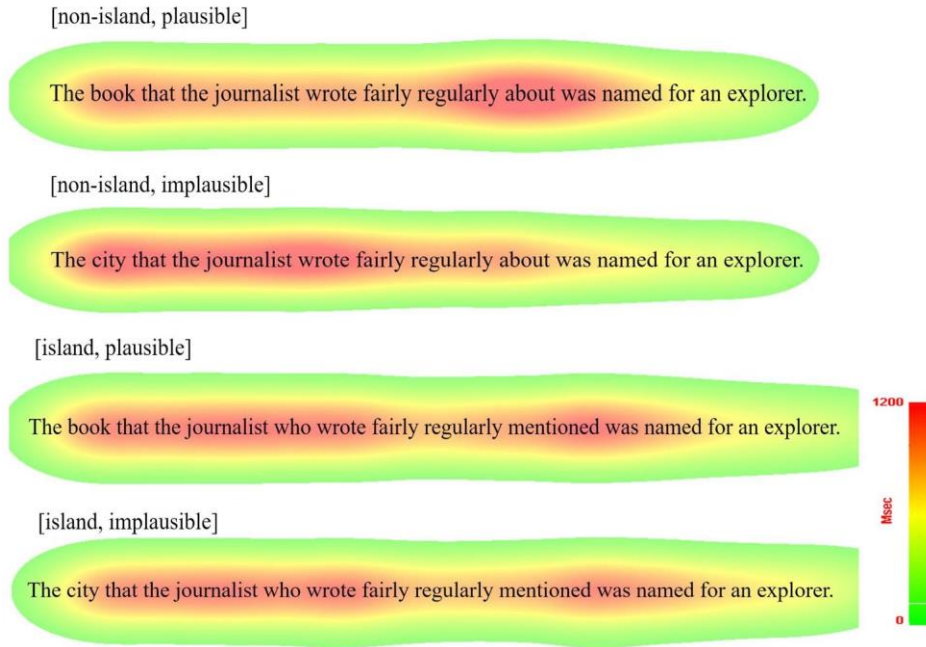


Figure 6. Fixation map: Reading profiles of the early ESL learners

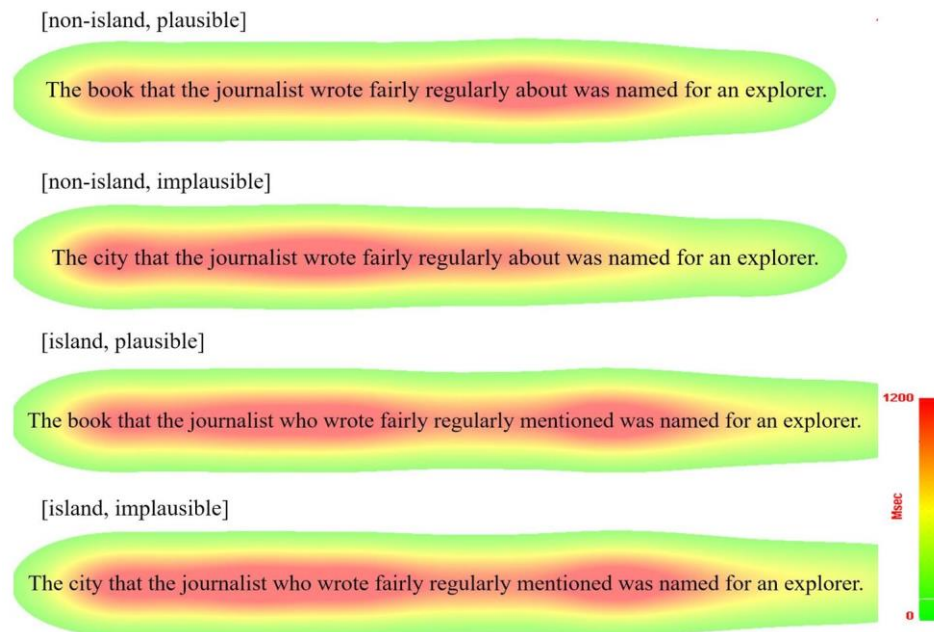


Figure 7. Fixation map: Reading profiles of the adult ESL learners

At Region3, *about was*, a reverse reading pattern was observed for all groups in the non-island condition. The native English speakers seemed to have spent slightly, but visibly more time at this region in the plausible than in the implausible sentences. The two ESL groups showed a similar pattern, but the differences between the two plausibility conditions were shown to be much greater for both learner groups. As discussed, Region3 is the ultimate gap position where the parser must withdraw its initially established dependency analysis at Region1 (i.e., *wrote the book* for plausible, and *wrote the city* for implausible), followed by an immediate reanalysis of relocating the filler as the object of the preposition *about* when reading sentences in the non-island condition. This reanalysis can be more taxing, especially when the initial analysis bears a more plausible interpretation (i.e., *wrote the book* as opposed to *wrote the city*). Taking this into account, the overall reading profiles of the three groups at Region3 appeared to be in line with this account. In the island condition, more reddish colors in both plausible and implausible conditions suggest that all groups tended to have more difficulties to digest structurally more complex sentences, as shown by more aggregated fixation durations across the regions, compared to their reading profiles in the non-island condition. However, the comparison of the two fixation maps in the island condition seemed to suggest that plausibility effects observed at Region1 and Region3 in the non-island condition did not seem to be present, or relatively weaker for all groups at the least. Note that the plausibility effects in the non-island condition are the byproducts of the filler's attempt to fill the gap as soon as possible. However, at the same verb, *wrote*, the parser may not postulate a gap in the island condition (only if it respects the island constraints), thereby yielding no plausible or implausible interpretations. In the same vein, there should be no reanalysis effect at Region3 as well in the island condition, because the verb *mentioned* is the structurally earliest gap available for the parser in this case.

Consequently, Region3 should be the place where an initial gap postulation should occur, not a reanalysis. Given that an integration of the filler (either the book or the city) as the object of *mentioned* does not render any semantically diverging manipulation at this region (i.e., no plausibility effect, and both should be plausible), reading profiles between the two plausibility conditions at this region should be comparable to one another in the case of the island condition.

The observation of the overall reading profiles across the experimental conditions indicated that the reading patterns of the advanced early and adult ESL learners in the current study were similar to those of the NS of English. However, it does not necessarily mean that the learners utilized the same types of linguistic information as the native English speakers for comprehension. Note that the maps included all fixations recorded over the course of the entire reading. For that reason, although the heatmaps could help identify the sources of the spots that make differences across the experimental conditions, it is still not clear whether some effects (e.g., plausibility effect) observed in the maps are lingering effects that started from an early stage of processing, or effects that occurred at later stages of processing. As discussed earlier, operations of syntactic information such as the island constraints are generally considered to occur at an earlier stage of processing. Therefore, it is necessary to examine more detailed time course reading processes, from an earlier point to later stages of processing, to better understand what kinds of parsing mechanisms and processing strategies learners computed to comprehend the L2 input. The next sections attempted to look into this aspect more in detail by analyzing multiple fine-grained eye-movement data, specifically at the aforementioned two critical regions. Region1 and Region3), and the two spillover regions (Region2 and Region4).

4.3. The effect of age of immersion on L2 processing of filler-gap dependencies

4.3.1. Active filler strategy and application of island constraints: Initial gap

4.3.1.1. Analysis of reading patterns at the first critical region (Region1)

Descriptive statistics of the three groups' RTs and first-pass regression probability at Region1 across the four experimental conditions are summarized in Table 6.

Table 6. *Descriptive statistics for RTs in and first-pass regressions in percent at Region1*

	island factor	plausibility factor	FFD	F-pass RT	RPD	Total RT	REGR
			<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
NS English	non-island	plausible	240 (42)	300 (90)	337 (105)	684 (235)	.08 (.09)
		implausible	281 (56)	379 (143)	487 (215)	822 (344)	.21 (.12)
	island	plausible	240 (50)	309 (138)	475 (247)	888 (257)	.21 (.15)
		implausible	231 (43)	300 (108)	442 (230)	844 (260)	.20 (.18)
Early ESL	non-island	plausible	258 (44)	443 (199)	593 (229)	944 (389)	.15 (.14)
		implausible	307 (67)	554 (248)	816 (331)	1268 (582)	.30 (.16)
	island	plausible	275 (46)	420 (142)	829 (391)	1122 (319)	.33 (.18)
		implausible	273 (51)	423 (161)	784 (344)	1095 (270)	.31 (.16)
Adult ESL	non-island	plausible	285 (52)	513 (196)	680 (288)	1064 (458)	.15 (.12)
		implausible	314 (78)	629 (306)	1160 (716)	1601 (664)	.30 (.26)
	island	plausible	280 (41)	481 (174)	1079 (709)	1326 (488)	.33 (.24)
		implausible	294 (50)	514 (217)	1145 (602)	1385 (432)	.37 (.26)

Note. FFD = first fixation duration, F-pass RT = first-pass RT, REGR = First-pass regression, RPD = regression path duration.

As shown, all groups showed increased RTs in the [non-island, implausible] condition for all eye-tracking measures, compared to its plausible counterpart. This means that the participants attempted to fill the gap at this early region, thereby experiencing difficulties in dealing with implausible interpretations. In this case, the adult ESL group showed a largest plausibility effect, as measured by regression path duration (plausible: 680ms; implausible: 1160ms) and Total RTs (plausible: 1064ms; implausible: 1601ms). The early ESL learners also exhibited strong plausibility effects on those same measures, although the differences between the plausibility conditions were not as much as those of the adult ESL group. The first-pass regression in the non-island condition also patterned similarly between the plausible and implausible condition when there was no island structure (i.e., non-island condition), in that all three groups made more regressions as soon as they encountered implausible interpretations. Somewhat different patterns emerged between the adult ESL group and the other two groups in the island condition; the NS English and the early ESL groups presented slightly longer RTs and more regressions in the plausible condition. On the other hand, the adult ESL group showed, albeit marginal, increased RTs and more regressions in the implausible condition, similarly to their reading patterns in the non-island condition. Preliminary 3 (group) x 2 (island) x 2 (plausibility) mixed design ANOVAs were carried out separately for each dependent measure to examine if there were any differences in reading patterns among the groups across the experimental conditions. A summary of the inferential statistics is provided in Table 7.

Table 7. *Summary of the results of preliminary analyses at Region1*

		by-subject (F_1)				by-item (F_2)			
		df	f	p	η_p^2	df	f	p	η_p^2
first fixation duration	I	1, 70	5.877	.018	.078	1, 27	8.510	.017	.238
	P	1, 70	25.119	.001	.264	1, 27	9.827	.004	.267
	G	2, 70	10.160	.001*	.225	1.6, 44.5	22.106	.001*	.450
	I x G	2, 70	1.595	.210	.044	2, 54	.996	.376	.036
	P x G	2, 70	.191	.826	.005	2, 54	.215	.807	.008
	I x P	1, 70	14.905	.001*	.176	1, 27	14.872	.001	.355
	I x P x G	2, 70	3.744	.029	.097	2, 54	2.432	.097	.083
first-pass RT	I	1, 70	19.665	.001*	.219	1, 27	14.257	.001	.346
	P	1, 70	16.834	.001*	.194	1, 27	13.256	.001	.329
	G	2, 70	13.879	.001*	.284	2, 54	65.768	.001*	.709
	I x G	2, 70	.875	.888	.024	2, 54	1.881	.162	.065
	P x G	2, 70	.003	.997	.001	2, 54	.615	.544	.022
	I x P	1, 70	14.837	.001*	.175	1, 27	13.185	.001	.328
	I x P x G	2, 70	.980	.380	.027	2, 54	.567	.571	.021
regression path duration	I	1, 70	9.087	.004	.115	1, 27	12.092	.002	.309
	P	1, 70	27.431	.001*	.282	1, 27	27.034	.001*	.500
	G	2, 70	17.076	.001*	.328	2, 54	188.302	.001*	.875
	I x G	2, 70	.717	.492	.024	2, 54	1.248	.295	.044
	P x G	2, 70	1.602	.209	.044	2, 54	1.836	.169	.064
	I x P	1, 70	31.716	.001*	.312	1, 27	33.658	.001*	.555
	I x P x G	2, 70	.709	.495	.020	2, 54	.012	.404	.069
Total RT	I	1, 70	10.409	.002	.129	1, 27	9.058	.006	.251
	P	1, 70	30.188	.000*	.301	1, 27	19.196	.001*	.416
	G	2, 70	12.687	.000*	.266	2, 54	130.396	.001*	.828
	I x G	2, 70	.997	.374	.028	2, 54	2.066	.075	.102
	P x G	2, 70	2.906	.061	.077	2, 54	3.027	.057	.101
	I x P	1, 70	34.063	.000*	.327	1, 27	23.557	.001*	.466
	I x P x G	2, 70	1.433	.245	.039	2, 54	.817	.447	.029
		df	f	p					
first-pass regression	I	1, 2032	29.077	.001*					
	P	1, 2032	28.077	.001*					
	G	2, 2032	6.323	.002					
	I x G	2, 2032	.255	.775					
	P x G	2, 2032	.320	.726					
	I x P	1, 2032	33.764	.001*					
	I x P x G	2, 2032	.629	.533					

Notes. 1. I = island constraints factor, P = plausibility factor, G = group. .001* = $p < .001$.

2. The first-pass regression analyses took into account both the subjects & items as random factors.

The results of the preliminary analyses showed that for all dependent measures in both by-subjects (F_1) and by-items (F_2) analyses, there was a significant main effect of island condition, likely due to elevated RTs (first fixation duration & first-pass RT) and increased regression ratios reading in the non-island condition, and because of the generally longer RTs in the island condition (both plausible and implausible) for regression path duration and Total RT. There was also a significant main effect of plausibility, presumably because of longer RTs in the [implausible, non-island] condition, and a significant group effect, arguably due to significantly faster reading speed of the NS English group than the ESL learners¹⁹. A significant plausibility by island interaction was also observed in both F_1 and F_2 analyses across the measures, reflecting a strong plausibility effect (i.e., longer RTs in the implausible than in the plausible condition) that was restricted to the non-island condition. However, no group related interaction was found, except in the first fixation duration, and marginally in the Total RTs ($p_1 = .061$, $p_2 = .076$). For the first fixation duration, a significant 3-way interaction was found in the F_1 analysis; $F_1(2, 70) = 3.744$, $p = .029$, $\eta_p^2 = .097$, and marginally in the F_2 analysis; $F_2(2, 54) = 2.432$, $p = .097$, $\eta_p^2 = .083$. The reading patterns of the three groups at Region 1 are plotted in Figure 8 (first fixation duration & first-pass RT) and Figure 9 (regression path duration & Total RT).

¹⁹ Because native speakers' reading speed is generally much faster than nonnative speakers (see e.g., Juffs, 2005), an observation of a significant main group effect is not surprising and less informative. Therefore, a main group effect will not be further addressed in the discussion of the results, but all the results are provided in the summary of the preliminary analyses in the tables.

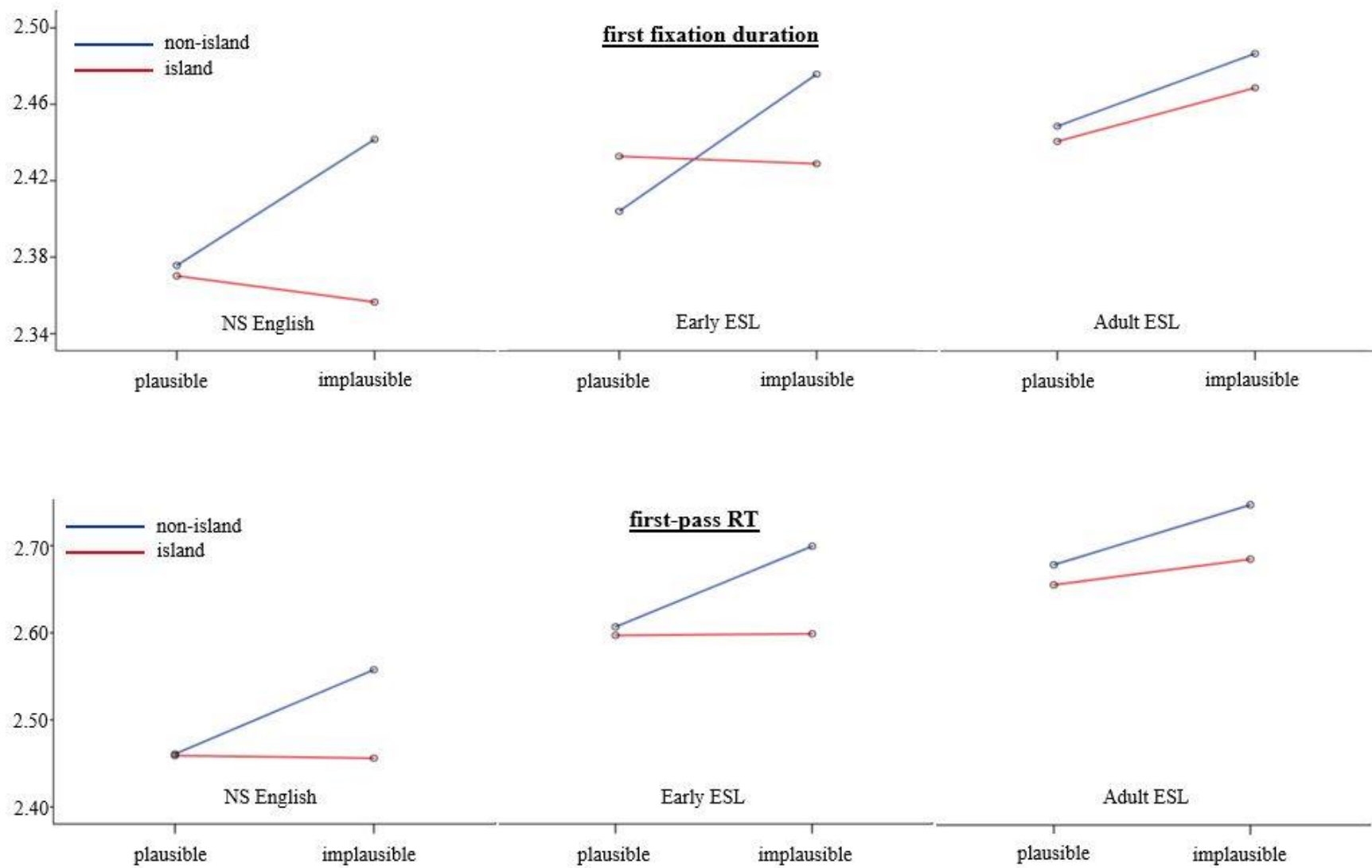


Figure 8. Reading patterns of the three groups during early stages of processing at Region1

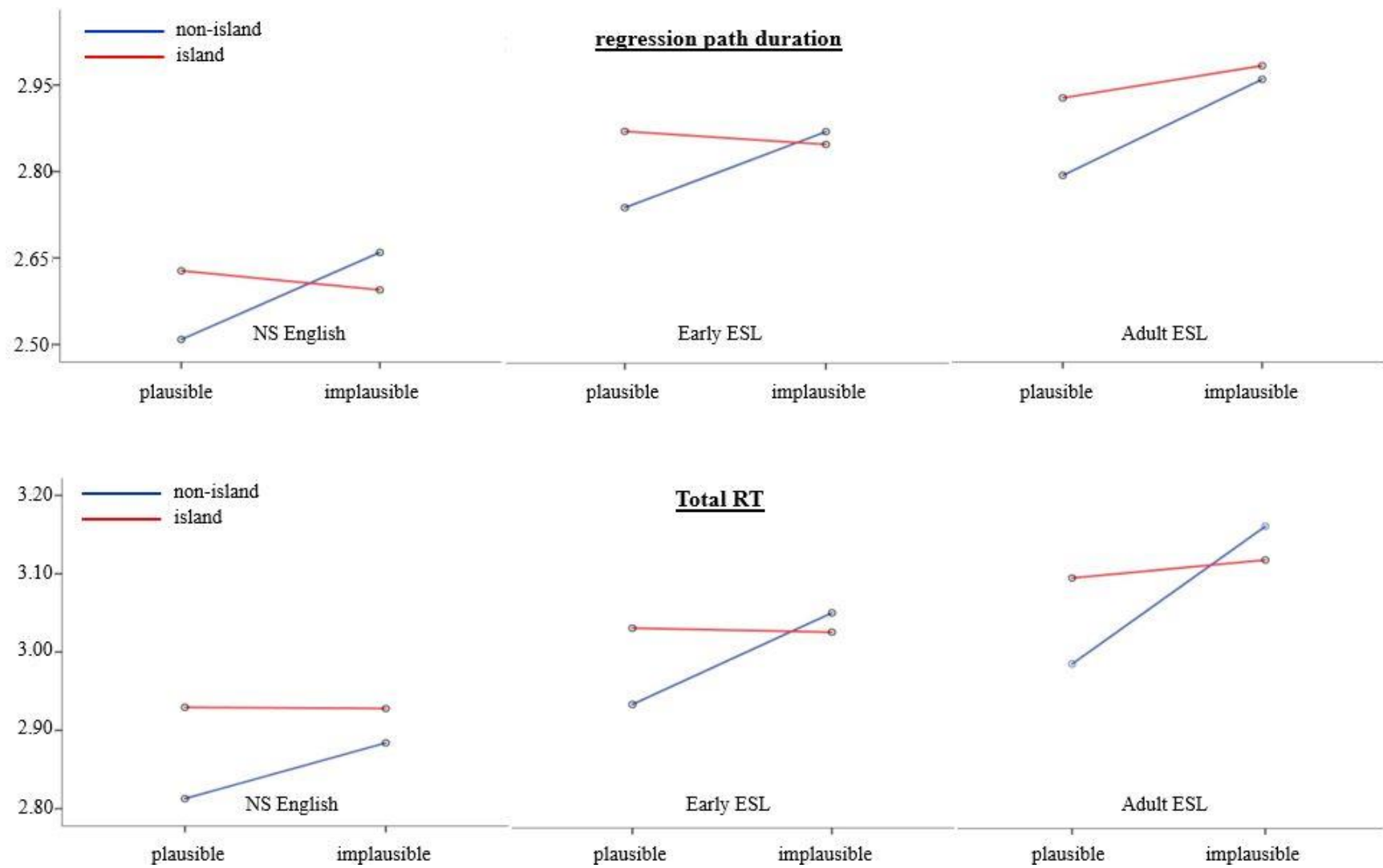


Figure 9. Reading patterns of the three groups during late stages of processing at Region1

As the preliminary analyses revealed a group-related interaction on the first fixation duration, a series of follow-up 2 x 2 repeated measures ANOVAs were performed for each group separately to better identify the source of the significant interaction. First, the NS English group showed a significant island effect, $F_1(1, 23) = 7.753, p = .011, \eta_p^2 = .252$; $F_2(1, 27) = 10.146, p = .004, \eta_p^2 = .273$, a significant plausibility effect in the F_1 analysis, $F_1(1, 23) = 7.784, p = .010, \eta_p^2 = .253$, and marginally in the F_2 analysis, $F_2(1, 27) = 3.698, p = .065, \eta_p^2 = .120$, and, crucially, a reliable island by plausibility interaction, $F_1(1, 23) = 11.631, p = .002, \eta_p^2 = .336$, power = .904; $F_2(1, 27) = 15.572, p = .001, \eta_p^2 = .366$. The source of such significant main effects and interaction for the NS English group should be attributed to the plausibility effect (i.e., RT discrepancies between the plausible and implausible condition) that was restricted only to the island condition. Subsequent planned paired sample t -tests confirmed this account. The mean RTs of the NS English group in the implausible condition were significantly longer than those in the plausible condition, only within the non-island condition: non-island: [$t_1(23) = 5.484, p < .001, d = .845$; $t_2(27) = 4.663, p < .001, d = 1.872$]; island: [$t_1(23) = .785, p = .440, d = .161$; $t_2(27) = .796, p = .433, d = .229$].

The early ESL group showed no reliable RT differences across the island conditions, $F_1(1, 20) = .243, p = .627, \eta_p^2 = .012$; $F_2(1, 27) = .941, p = .341, \eta_p^2 = .034$; although they also showed a plausibility effect in the non-island condition, their RTs in the island condition were overall high, offsetting the longer RTs in the [non-island, implausible] condition. A main plausibility effect was found only in the F_1 analysis, $F_1(1, 20) = 10.526, p = .004, \eta_p^2 = .345$; $F_2(1, 27) = 2.776, p = .107, \eta_p^2 = .093$. Importantly, however, the early ESL group also displayed a significant plausibility by island interaction in both F_1 & F_2 analyses, $F_1(1, 20) = 11.194, p = .003, \eta_p^2 = .359$; $F_2(1, 27) = 5.743, p = .024, \eta_p^2 = .175$. The following planned paired t -tests

presented, like the NS English group, a clear plausibility effect only in the non-island condition; non-island: [t_1 (20) = 4.264, $p < .001$, $d = .866$; t_2 (27) = 2.478, $p = .020$, $d = .807$]; island: [t_1 (23) = .280, $p = .782$, $d = .030$; t_2 (27) = .269, $p = .790$, $d = .069$].

The adult ESL group, like the early ESL group, also showed no significant main effect of island constraints, F_1 (1, 23) = .975, $p = .332$, $\eta_p^2 = .035$; F_2 (1, 27) = 2.935, $p = .098$, $\eta_p^2 = .098$. A significant plausibility effect was found in both F_1 and F_2 analyses, F_1 (1, 23) = 7.964, $p = .009$, $\eta_p^2 = .235$; F_2 (1, 27) = 8.610, $p = .007$, $\eta_p^2 = .242$, presumably due to increased RTs on the implausible sentences not only in the non-island condition, but also marginally in the island condition. This rendered no significant interaction of the two factors, F_1 (1, 23) = .231, $p = .634$, $\eta_p^2 = .008$; F_2 (1, 27) = .039, $p = .846$, $\eta_p^2 = .001$, in contrast to the other two groups (see Figure 8).

4.3.1.2. Analysis of reading patterns at the spillover region (Region2)

Descriptive statistics of the three groups' RTs and first-pass regression probabilities across the four experimental conditions at the spillover region (Region2) are given in Table 8.

Table 8. *Descriptive statistics for RTs and first-pass regressions at Region2*

	Island Cond.	Plausibility Cond.	FFD	F-pass RT	RPD	Total RT	REGR
			<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
NS English	non-island	plausible	253 (48)	520 (159)	646 (221)	1117 (418)	.15 (.14)
		implausible	277 (52)	495 (122)	714 (179)	1126 (385)	.21 (.12)
	island	plausible	248 (43)	473 (146)	729 (204)	1589 (541)	.23 (.18)
		implausible	255 (51)	442 (152)	694 (236)	1445 (516)	.20 (.16)
Early ESL	non-island	plausible	266 (49)	589 (131)	808 (286)	1703 (832)	.14 (.14)
		implausible	289 (46)	657 (145)	1198 (425)	1610 (633)	.29 (.16)
	island	plausible	280 (46)	668 (218)	983 (365)	1576 (472)	.22 (.19)
		implausible	263 (36)	625 (184)	966 (316)	1563 (389)	.24 (.13)
Adult ESL	non-island	plausible	272 (39)	712 (147)	908 (292)	1839 (690)	.13 (.15)
		implausible	300 (65)	802 (222)	1519 (631)	1934 (661)	.42 (.25)
	island	plausible	279 (51)	753 (199)	1115 (407)	1850 (500)	.21 (.17)
		implausible	277 (44)	764 (215)	1156 (405)	1827 (390)	.24 (.17)

Overall, the RT patterns of the three groups across the four experimental conditions at Region2 were somewhat similar to the critical region (Region1), but some differences emerged among the groups, especially in the non-island condition. For the NS English group, they tended to spend slightly more time reading implausible sentences than in reading plausible sentences, as measured by first fixation duration and regression path duration, indicating a plausibility effect at least to a certain degree. However, the size of the effect looked to be slightly decreased across those measures (i.e., smaller reading time differences between the two plausibility conditions), compared to the effects they exhibited at the previous region. The first-pass RT, first-pass

regression, and Total RT of the NS group showed neither a plausibility effect nor a reliable interaction of plausibility and island constraints, there was no plausibility effect in either island condition; the reading patterns of the NS English group in the island condition remained the same as those at Region1, without much difference between the two plausibility conditions.

The early ESL group, by and large, patterned similar to the NS English group across the measures, but their RT patterns in the non-island condition displayed relatively clearer spillover effects for regression path duration and first-pass regression as shown by longer RTs and more regressions in the implausible than in the plausible condition. This trend was also shown in the adult ESL group for those measures (i.e., regression path duration & first-pass regression), in that the plausibility effect they displayed in the non-island condition even appeared to be greater than the effects they had at the critical region, with slowdowns in RT and more regressions to greater degrees in dealing with implausible interpretations. However, both ESL groups, like the NS English group, did not show such plausibility effect in the island condition across the measures. Recall that the adult ESL group showed similar reading patterns between the two island conditions at Region1, with slightly longer RTs in reading of implausible sentences in the island conditions. At Region2, however, the RT differences the adult ESL learners showed in the island condition looked to be much smaller, compared to their RT differences at the previous region.

The preliminary analyses on each measure revealed the following significant group associated 2-way interactions: group by plausibility [regression-path duration; $p_1 = .003$, $p_2 = .009$, and first-pass regression; $p = .002$], and group by island [Total RT; $p_1 = .018$, $p_2 < .001$]. See Table 9 for a complete summary of the preliminary analyses at Region2. Also, the reading patterns of the three groups at Region 2 are plotted in Figure 10 (first fixation duration & first-pass RT) and Figure 11 (regression path duration & Total RT).

Table 9. *Summary of the results of preliminary analyses at Region2*

		by-subject (F_1)				by-item (F_2)			
		df	f	p	η_p^2	df	f	p	η_p^2
first fixation duration	I	1, 70	2.492	.119	.034	1, 27	2.091	.160	.072
	P	1, 70	6.514	.016	.081	1, 27	5.895	.022	.179
	G	2, 70	3.007	.056	.079	1.6, 44.5	5.741	.005	.175
	I x G	2, 70	.192	.826	.005	2, 54	.305	.731	.011
	P x G	2, 70	.657	.522	.018	2, 54	.733	.485	.026
	I x P	1, 70	14.318	.001*	.170	1, 27	4.994	.035	.155
	I x P x G	2, 70	.659	.521	.018	2, 54	.178	.831	.007
first-pass RT	I	1, 70	2.068	.155	.029	1, 27	.100	.754	.004
	P	1, 70	.372	.544	.005	1, 27	.015	.903	.001
	G	2, 70	31.56	.001*	.474	2, 54	120.77	.001*	.817
	I x G	2, 70	2.718	.073	.072	2, 54	3.158	.050	.105
	P x G	2, 70	3.416	.049	.082	2, 54	1.228	.301	.043
	I x P	1, 70	3.488	.068	.047	1, 27	1.970	.172	.068
	I x P x G	2, 70	.456	.636	.013	2, 54	.221	.802	.008
regression path duration	I	1, 70	.005	.944	.001	1, 27	.736	.399	.027
	P	1, 70	31.869	.001*	.313	1, 27	17.695	.001*	.396
	G	2, 70	23.287	.001*	.400	2, 54	99.266	.001*	.786
	I x G	2, 70	.321	.726	.009	1.4, 38.0	.781	.424	.028
	P x G	2, 70	6.277	.003	.152	2, 54	5.112	.009	.159
	I x P	1, 70	32.903	.001*	.320	1, 27	18.867	.001	.411
	I x P x G	2, 70	1.464	.238	.040	2, 54	2.346	.106	.080
Total RT	I	1, 70	5.835	.018	.077	1, 27	6.057	.021	.183
	P	1, 70	.453	.504	.006	1, 27	.652	.427	.024
	G	2, 70	10.911	.001*	.238	2, 54	133.594	.001*	.832
	I x G	2, 70	4.246	.018	.108	2, 54	15.433	.001*	.364
	P x G	2, 70	1.263	.289	.035	2, 54	2.646	.061	.098
	I x P	1, 70	.604	.440	.009	1, 27	.042	.840	.002
	I x P x G	2, 70	.042	.840	.002	2, 54	.929	.401	.033
		df	f	p					
first-pass regression	I	1, 2032	.466	.495					
	P	1, 2032	29.291	.001*					
	G	2, 2032	.782	.458					
	I x G	2, 2032	.602	.548					
	P x G	2, 2032	6.436	.002					
	I x P	1, 2032	20.756	.001*					
	I x P x G	2, 2032	2.521	.081					

Note. I = island constraints factor, P = plausibility factor, G = group, // .001* = $p < .001$.

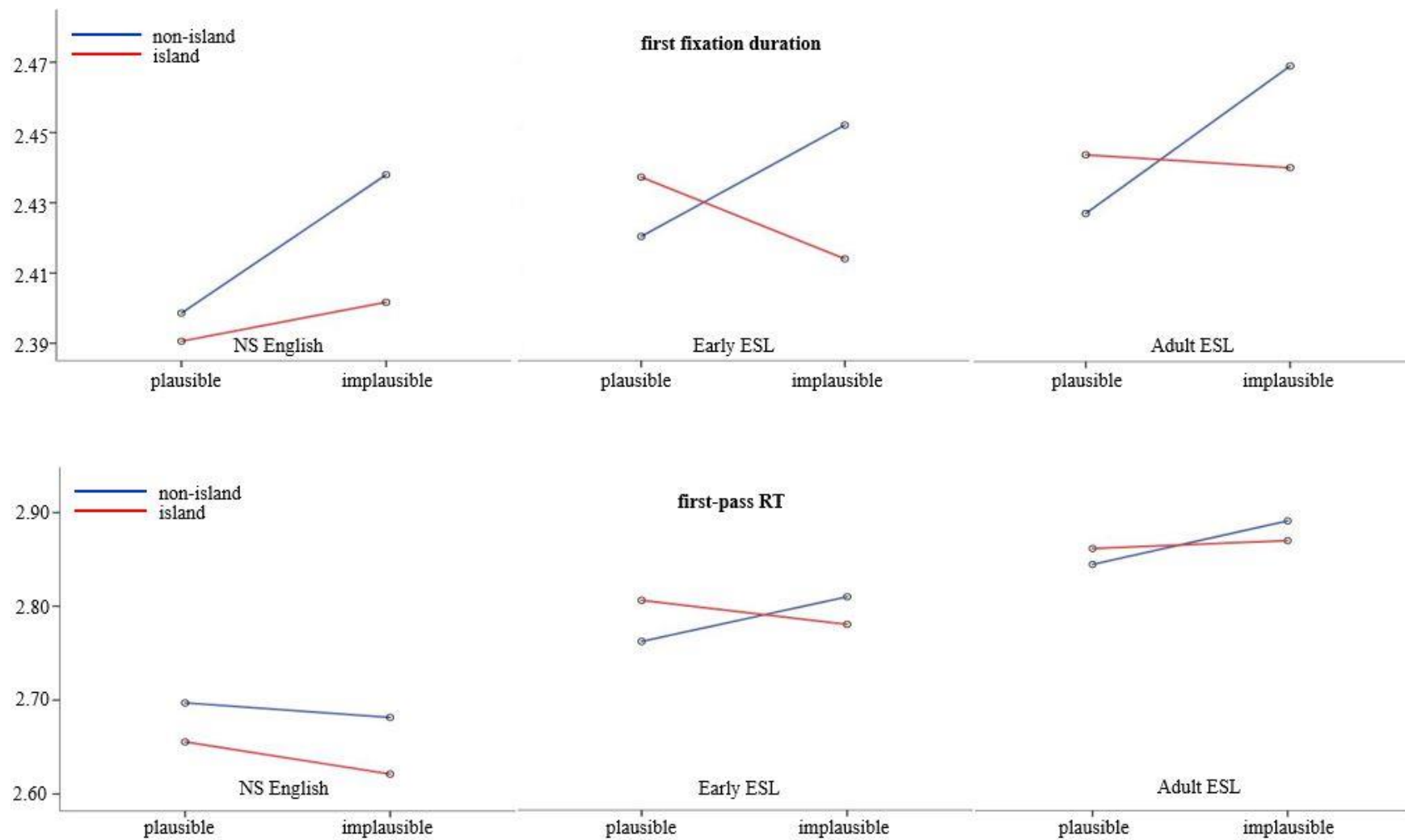


Figure 10. Reading patterns of the three groups during early stages of processing at Region2

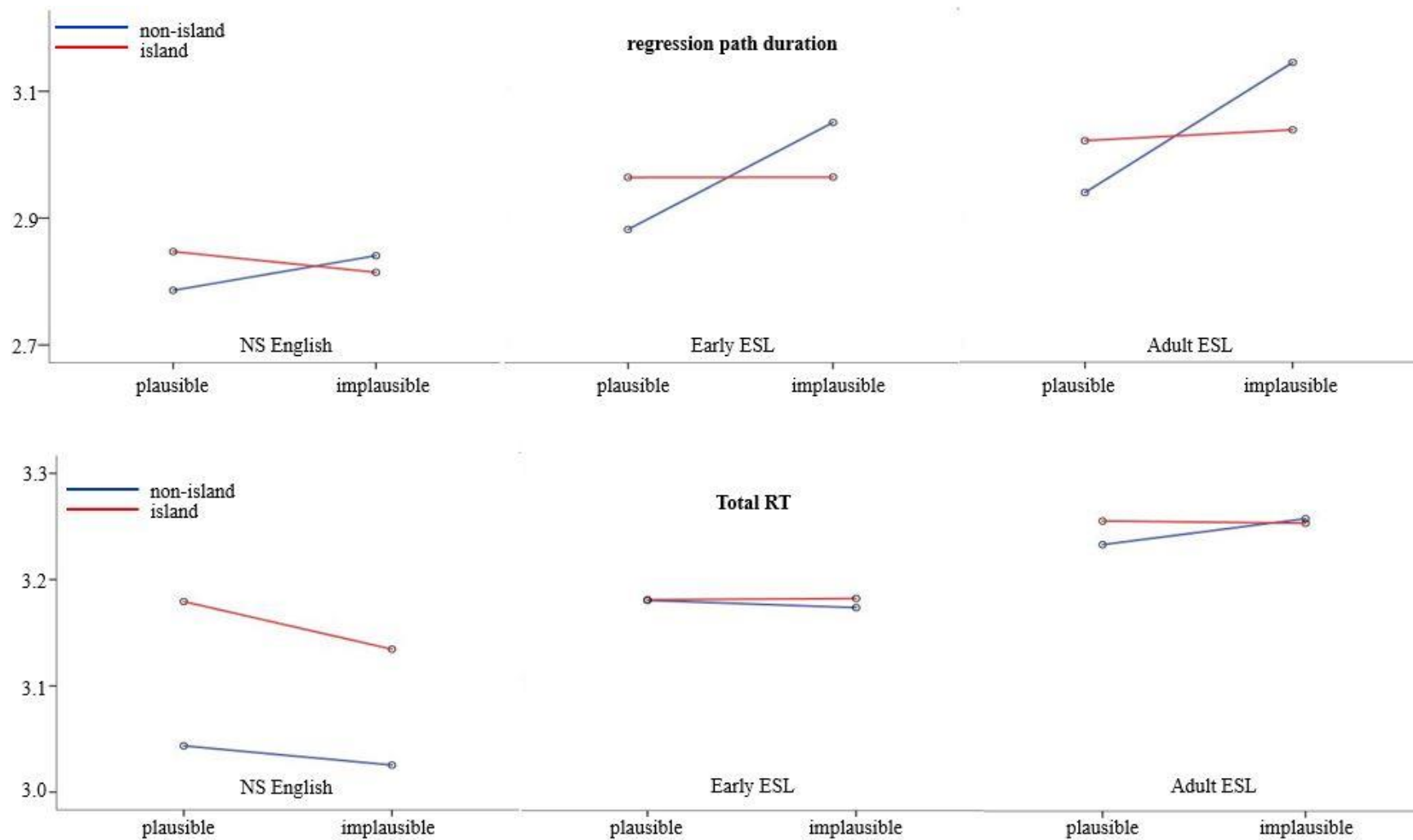


Figure 11. Reading patterns of the three groups during late stages of processing at Region2

Regression path duration It appeared that the significant group by plausibility interaction on regression path duration was on the non-significant main effect of plausibility for the NS English group, $F_1(1, 23) = .222, p = .642, \eta_p^2 = .010$; $F_2(1, 27) = .098, p = .757, \eta_p^2 = .004$], in contrast to the two ESL groups that showed significant plausibility effects, apparently because of significantly longer RTs in the [non-island, implausible condition] for both learner groups; the early ESL, $F_1(1, 23) = 14.880, p = .001, \eta_p^2 = .427$; $F_2(1, 27) = 4.943, p = .035, \eta_p^2 = .155$, and the adult ESL, $F_1(1, 23) = 44.733, p < .001, \eta_p^2 = .624$; $F_2(1, 27) = 45.189, p < .001, \eta_p^2 = .626$. The longer regression path durations in the [implausible, non-island condition] by the two ESL groups also appeared to contribute to a significant interaction of island and plausibility, $F_1(1, 23) = 10.739, p = .004, \eta_p^2 = .349$; $F_2(1, 27) = 13.666, p = .001, \eta_p^2 = .336$, for the early ESL group, and $F_1(1, 23) = 14.648, p < .001, \eta_p^2 = .352$; $F_2(1, 27) = 15.391, p = .001, \eta_p^2 = .363$, for the adult ESL group. Subsequent planned paired t -tests confirmed that a significant plausibility effect was present only in the non-island condition, for both learner groups; early ESL group: [non-island: $t_1(20) = 3.974, p = .001, d = 1.072$; $t_2(27) = 3.865, p = .001, d = .1.061$; and island: $t_1(20) = .307, p = .762, d = .003$; $t_2(27) = .173, p = .864, d = .047$], and adult ESL group: [non-island: $t_1(27) = 8.092, p < .001, d = 1.234$; $t_2(27) = 8.262, p < .001, d = 2.358$; and island: $t_1(27) = .582, p = .565, d = .108$; $t_2(27) = 1.032, p = .311, d = .290$]. The NS English group showed a significant island by plausibility interaction only in the F_1 analysis, $F_1(1, 23) = 4.433, p = .046, \eta_p^2 = .162$; $F_2(1, 27) = 1.112, p = .301, \eta_p^2 = .040$. However, the following t -tests showed only a marginal plausibility effect in the by-subjects analysis; [non-island: $t_1(23) = 1.957, p = .063, d = .419$; $t_2(27) = 1.125, p = .270, d = .268$; and island: $t_1(23) = .953, p = .351, d = .231$; $t_2(27) = .477, p = .637, d = .130$. Of the three groups, only the adult ESL group showed a main island effect, $F_1(1, 23) = 6.853, p = .014, \eta_p^2 = .202$; $F_2(1, 27) = 4.435, p = .045, \eta_p^2 = .162$.

= .141, likely due to their much slower reading in the [non-island, implausible] condition. For the other two groups, their generally slower RTs in the island condition tended to approximate the sum of relatively faster RTs in the plausible sentences and relatively slower RTs in the implausible sentences in the non-island condition, leading to no island effect; The NS English, $F_1(1, 23) = .543, p = .469, \eta_p^2 = .023$; $F_2(1, 27) = .001, p = .975, \eta_p^2 = .001$; the early ESL, $F_1(1, 23) = .005, p = .945, \eta_p^2 = .001$; $F_2(1, 27) = .020, p = .890, \eta_p^2 = .001$.

First-pass regression The results of the follow-up analyses on first-pass regression patterned very similarly to those of regression path duration. First, the adult ESL group showed a drastic increase in their regression ratio in the [non-island, implausible] condition (approximately 42%), compared to their regression in the plausible counterpart (approx. 13%), indicating a greater plausibility effect than they previously had at the critical region (Region1, approx. 15% and 30% respectively in reading plausible and implausible sentences). This resulted in a significant main effect for plausibility, $F(1, 780) = 56.882, p < .001$, as well as significant interaction of island and plausibility, $F(1, 780) = 27.221, p < .001$. The early ESL group also showed a significant effect of plausibility, $F(1, 584) = 10.491, p = 0.001$, with increased regression ratios in the [non-island, implausible] condition. However, neither a significant plausibility effect, $F(1, 668) = .231, p = .631$, nor a significant interaction, $F(1, 668) = .231, p = .099$, was found for the NS English group. Lastly no group showed a main effect of island: The NS English speakers, $F(1, 668) = 1.426, p = .233$, early ESL learners, $F(1, 584) = .592, p = .442$, and the adult ESL learners, $F(1, 780) = .226, p = .636$, presumably because their lower regression ratios in the [non-island, plausible] and higher regression ratio in the [non-island, implausible] condition offset their generally higher regression ratios in the island condition.

Total RT As reported above, the preliminary analysis on Total RT showed a significant group by island interaction in both F_1 ($p_1 = .018$) and F_2 ($p_2 < .001$) analyses. The following analyses for each group could identify that the source of the significant interaction was on the relatively longer reading times of the early and adult ESL groups spent in the non-island condition than in the island condition. In contrast, the NS English group showed the opposite pattern, spending more times in the island than in the non-island condition. This rendered a significant island effect for the NS English group, $F_1(1, 23) = 24.920, p < .001, \eta_p^2 = .520$; $F_2(1, 27) = 27.764, p < .001, \eta_p^2 = .507$, but not for the early ESL group, $F_1(1, 23) = .011, p = .917, \eta_p^2 = .001$; $F_2(1, 27) = 1.569, p = .221, \eta_p^2 = .055$, and the adult ESL group, $F_1(1, 23) = .089, p = .767, \eta_p^2 = .003$; $F_2(1, 27) = .991, p = .328, \eta_p^2 = .035$. No group showed a significant plausibility effect; NS English, $F_1(1, 23) = 2.019, p = .169, \eta_p^2 = .081$; $F_2(1, 27) = 1.701, p = .203, \eta_p^2 = .059$; early ESL, $F_1(1, 23) = .015, p = .904, \eta_p^2 = .001$; $F_2(1, 27) = 1.520, p = .228, \eta_p^2 = .053$; and the adult ESL group, $F_1(1, 23) = .515, p = .478, \eta_p^2 = .019$; $F_2(1, 27) = 3.405, p = .076, \eta_p^2 = .112$. Finally, no group showed a significant island by plausibility interaction at Region2m as measured by Total RT; NS English, $F_1(1, 23) = .618, p = .440, \eta_p^2 = .026$; $F_2(1, 27) = .831, p = .370, \eta_p^2 = .030$; early ESL, $F_1(1, 23) = .055, p = .816, \eta_p^2 = .003$; $F_2(1, 27) = .523, p = .476, \eta_p^2 = .019$; and the adult ESL group, $F_1(1, 23) = .693, p = .412, \eta_p^2 = .025$; $F_2(1, 27) = .238, p = .630, \eta_p^2 = .009$.

4.3.1.3. Interim summary of the results— Initial gap

Table 10. *Summary of the major findings at the initial gap*

Significant island x plausibility interaction?			Major implications
	(Critical) Region 1	(Spillover) Region 2	
NS English	<u>YES</u> ✓ FFD ✓ first-pass RT ✓ REGR ✓ RPD ✓ Total RT	<u>YES</u> ✓ FFD	<ul style="list-style-type: none"> All three groups appeared to have employed the active filler gap strategy in the non-island condition, demonstrating plausibility effect with longer RTs and more regressions in reading implausible sentences. All three groups seemed to have applied relevant relative clause island constraint from early stages of processing, avoiding gap postulations in the island environment.
		<u>NO</u> ✓ First-pass RT ✓ REGR ✓ RPD ✓ Total RT	
Early ESL	<u>YES</u> ✓ FFD ✓ first-pass RT ✓ REGR ✓ RPD ✓ Total RT	<u>YES</u> ✓ FFD ✓ REGR ✓ RPD	<ul style="list-style-type: none"> Adult ESL group showed slightly delayed applications of island constraints, compared to the other two groups, showing the similar reading patterns in both island conditions (i.e., longer RTs in reading implausible sentences), as measured by FFD.
		<u>NO</u> ✓ First-pass RT ✓ Total RT	
Adult ESL	<u>YES</u> ✓ first-pass RT ✓ REGR ✓ RPD ✓ Total RT	<u>YES</u> ✓ FFD ✓ REGR ✓ RPD	<ul style="list-style-type: none"> Both ESL groups displayed clearer spillover effects until the later stages of processing, compared to the NSs of English controls.
	<u>NO</u> ✓ FFD	<u>NO</u> ✓ First-pass RT ✓ Total RT	

Note. FFD = first fixation duration, REGR = first-pass regression, RPD = regression path duration

4.3.2. Filler-gap reanalysis: Ultimate gap

4.3.2.1. Analysis of reading patterns at the second critical region (Region3)

Table 11 provides descriptive statistics of the three groups' RTs and regression probabilities at the second critical region (Region3), the region that includes the ultimate gap for the filler for all experimental conditions. In the non-island condition, Region3 serves as a spot where the parser needs to withdraw its initial analysis at Region1, and performs an immediate reanalysis by reassigning the filler as the object of the preposition *about*.

Table 11. *Descriptive statistics for RTs and first-pass regression at Region3*

	Island Cond.	Plausibility Cond.	FFD	F-pass RT	REGR	RPD	Total RT
			<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
NS	non-island	plausible	293 (49)	442 (186)	.23 (.14)	697 (331)	980 (398)
		implausible	243 (48)	366 (126)	.13 (.10)	521 (243)	779 (315)
English	island	plausible	285 (63)	507 (233)	.28 (.23)	961 (519)	1802 (855)
		implausible	298 (80)	530 (212)	.33 (.19)	1086 (560)	1718 (800)
Early	non-island	plausible	325 (72)	652 (299)	.35 (.27)	1267 (857)	1685 (911)
		implausible	277 (56)	575 (218)	.23 (.22)	958 (559)	1057 (410)
ESL	island	plausible	316 (57)	814 (314)	.31 (.20)	1413 (712)	1774 (494)
		implausible	312 (64)	850 (378)	.29 (.22)	1374 (644)	1785 (421)
Adult	non-island	plausible	314 (88)	691 (295)	.35 (.25)	1535 (976)	1833 (1057)
		implausible	288 (64)	660 (295)	.37 (.25)	1301 (790)	1233 (457)
ESL	island	plausible	345 (65)	902 (333)	.44 (.26)	2174 (1131)	2284 (551)
		implausible	325 (73)	896 (370)	.46 (.27)	2300 (1287)	2187 (669)

As a result, Region3 is the place where a reanalysis effect is expected, in the form of a plausibility effect—but in a reverse direction compared to the plausibility effect found at the previous regions—as signaled by increased RTs and regression probabilities while reading sentences in the [non-island, plausible] than in the [non-island, implausible] condition. In the island condition, on the other hand, Region3 serves as an initial gap for the filler, given that there is no grammatical licit gap in the island structures. The parser therefore should be free from the plausibility manipulation (i.e., no plausibility effect) not only at the first critical region, but also at this region. As a result, the integration of the filler into a verb such as *mentioned* should yield about the same amount of processing load between the two plausibility conditions.

Bearing that in mind, the reading profiles of the three groups tended to show the expected patterns in the non-island condition, in that both the NS English and the two ESL groups exhibited longer reading times and made more first-pass regressions in the plausible than in the implausible condition across the measures. An exception was first-pass regression of the adult ESL groups, which had slightly more regressions in the implausible (approx. 37%) than in the plausible condition (approx. 35%). In addition, compared to the NS English and early ESL groups, the reading time differences of the adult ESL group between the two plausibility conditions looked to be slightly smaller, especially on first fixation and first-pass RT. On the other hand, the Total RT of the adult learners seemed to reflect the largest processing difficulties in the plausible condition. Interestingly, regression path durations of the adult ESL group were longer in the plausible condition, despite their slightly less frequent regression ratios (approx. 2% less). This might suggest that although the adult learners made more regressions when reading implausible sentences at this region (Region3), their recovery from the initial misanalysis made at Region1 took longer in the plausible condition. In the island condition, the reading patterns

between the plausibility conditions were somewhat mixed across the groups and measures, but, crucially, the differences between the two plausible conditions were generally smaller than the differences in the non-island condition regardless of the direction of them. The reading patterns of the three groups at the second critical region (Region 3) are plotted in Figure 12 (first fixation duration & first-pass RT) and Figure 13 (regression path duration & Total RT).

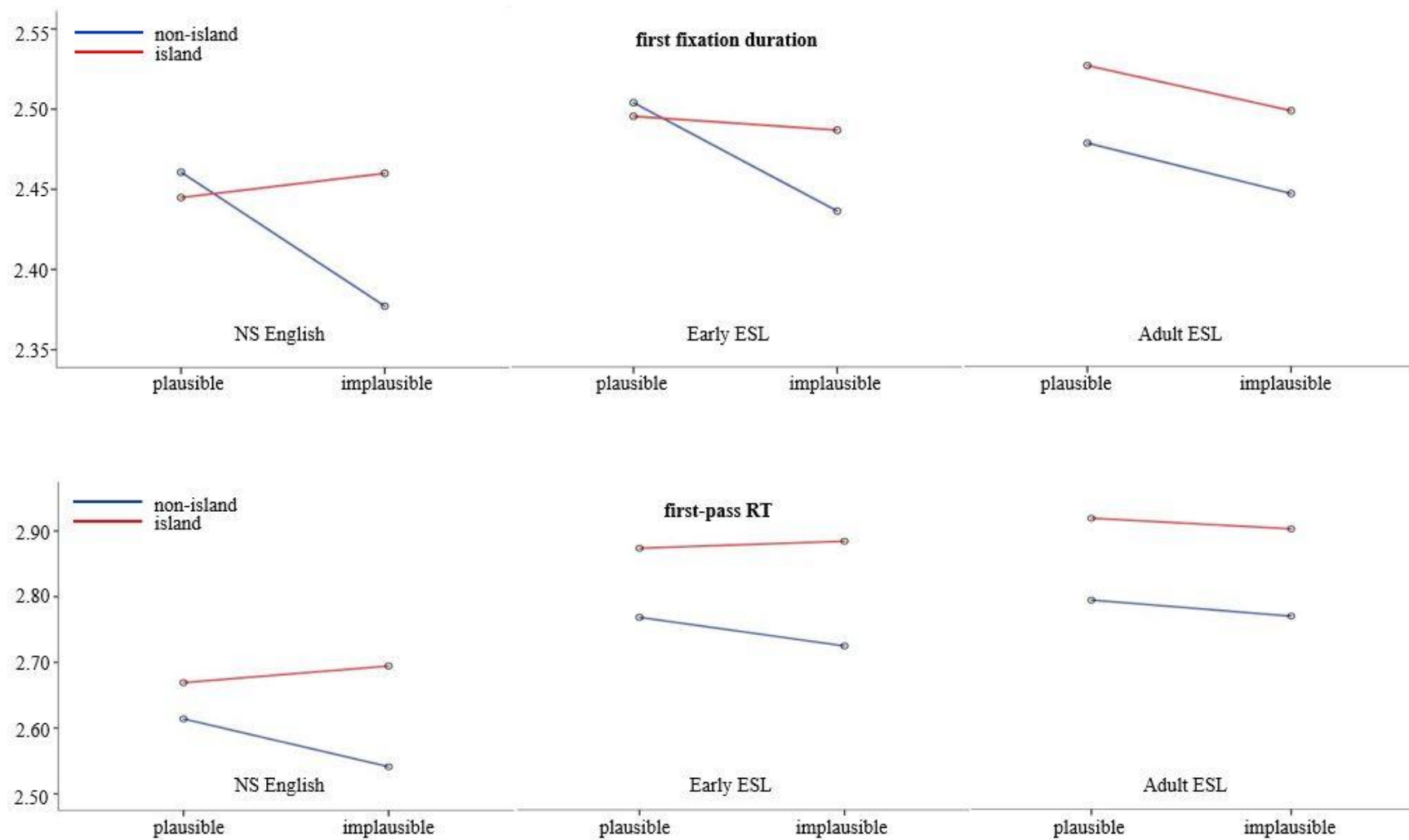


Figure 12. Reading patterns of the three groups during early stages of processing at Region3

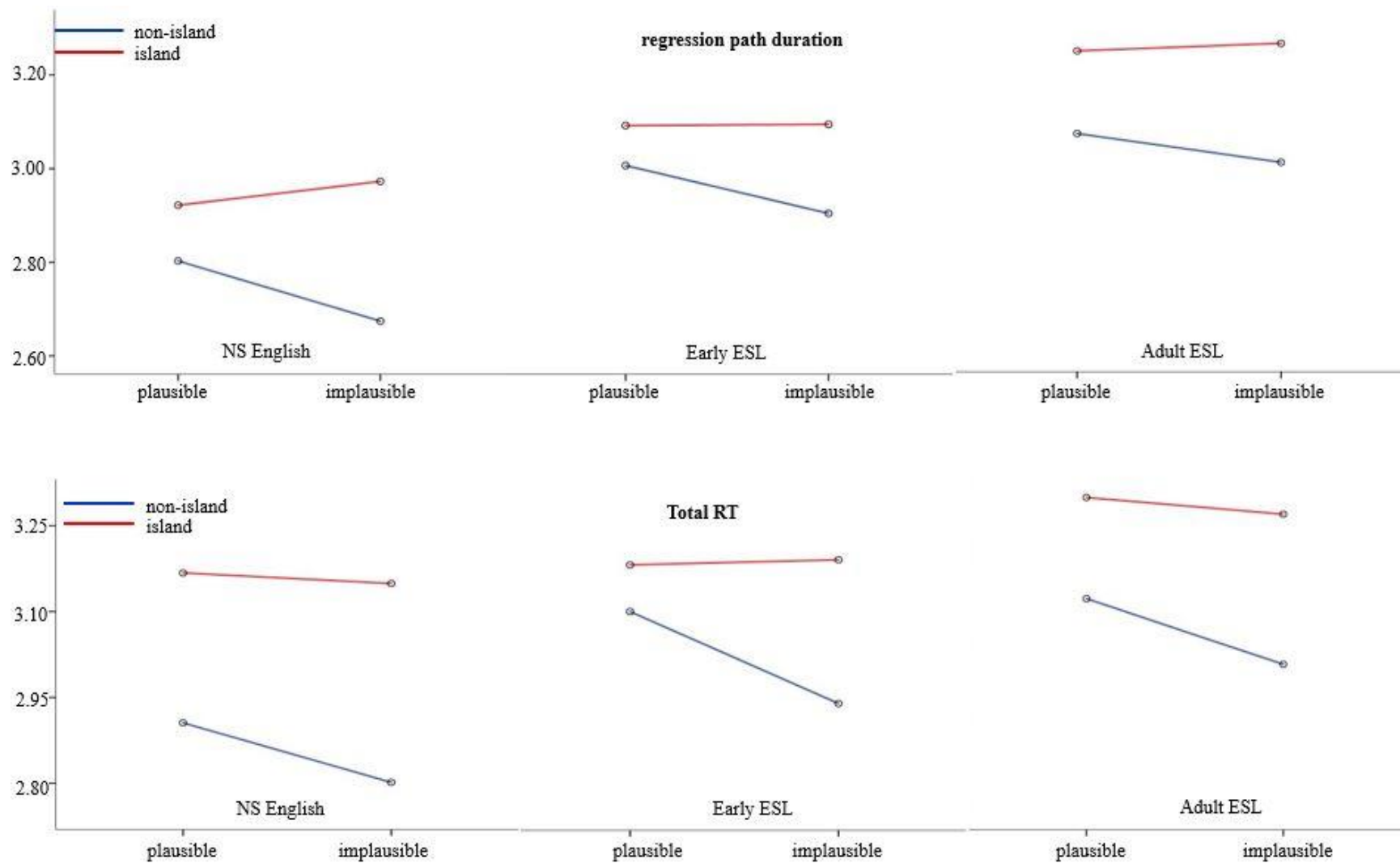


Figure 13. Reading patterns of the three groups during late stages of processing at Region3

A summary of the preliminary analyses at Region3 is given in Table 12. The preliminary analyses showed a main effect of island constraints for all measures, with significantly longer RTs and more regressions in the island condition, indicating generally more processing burden in the island condition.

Table 12. *Summary of the results of preliminary analyses at Region3*

Reading Times		by-subject (F_1)				by-item (F_2)			
		df	f	p	η_p^2	df	f	p	η_p^2
first fixation duration	I	1, 70	14.620	.001*	.173	1, 27	20.011	.001*	.426
	P	1, 70	22.932	.001*	.247	1, 27	11.217	.002	.294
	G	2, 70	3.743	.029	.097	2, 54	9.433	.001*	.259
	I x G	2, 70	.927	.400	.026	2, 54	1.050	.357	.037
	P x G	2, 70	.090	.914	.003	1.4, 38.8	.624	.490	.023
	I x P	1, 70	11.469	.001	.141	1, 27	10.093	.004	.272
	I x P x G	2, 70	3.311	.042	.086	2, 54	3.472	.038	.114
first- pass RT	I	1, 70	106.213	.001*	.603	1, 27	40.504	.001*	.600
	P	1, 70	3.318	.073	.045	1, 27	1.540	.225	.054
	G	2, 70	12.954	.001*	.270	2, 54	136.208	.001*	.835
	I x G	2, 70	.610	.546	.017	2, 54	.028	.972	.001
	P x G	2, 70	.031	.969	.001	2, 54	.253	.777	.009
	I x P	1, 70	8.234	.005	.105	1, 27	5.524	.026	.170
	I x P x G	2, 70	2.071	.134	.056	2, 54	.841	.437	.030
First-pass regression probability		df	f	p					
firs-pass regression	I	1, 2032	19.325	.001*					
	P	1, 2032	5.010	.025					
	G	2, 2032	4.823	.008					
	I x G	2, 2032	4.440	.012					
	P x G	2, 2032	2.922	.054					
	I x P	1, 2032	9.034	.003					
	I x P x G	2, 2032	3.339	.036					

Notes. 1. I = island constraints factor, P = plausibility factor, G = group. // .001* = $p < .001$.

2. As noted earlier, preliminary mixed ANOVA analyses on regression path duration and Total RT were not performed because the data did not meet the assumptions for ANOVA.

As shown, a significant main effect of plausibility was found on first fixation duration, first-pass regression, and marginally on first-pass RT in the F_1 analysis, which were most likely due to

increased reading times and regression ratios in the [non-island, plausible] condition for all groups. This pattern also contributed to a significant island and plausibility interaction for all measures, reflecting plausibility effects that are reverse to the plausibility effects observed at Region1. In regard to group related interactions, the following results were found. First, a significant 2-way interaction of group and island was found on first-pass regression ($p = .012$). Second, there was a marginally significant interaction of group and plausibility on first-pass regression ($p = .054$). Lastly, there was a significant 3-way (group x island x plausibility) interaction on first fixation duration ($p_1 = .042, p_2 = .038$) and first-pass regression ($p = .036$). The results of the follow-up analyses on each of those measures, and the result of the non-parametric Wilcoxon signed ranks test on regression path duration and Total RTs are reported below:

First fixation duration The NS English group showed a main effect of island, $F_1(1, 23) = 4.360, p = .048, \eta_p^2 = .159$; $F_2(1, 27) = 10.493, p = .003, \eta_p^2 = .280$, and a main effect of plausibility, $F_1(1, 23) = 10.584, p = .004, \eta_p^2 = .315$; $F_2(1, 27) = 4.760, p = .038, \eta_p^2 = .150$. Crucially, there was also a significant island by plausibility interaction for the NS English group, $F_1(1, 23) = 24.626, p < .001, \eta_p^2 = .517$; $F_2(1, 27) = 19.124, p < .001, \eta_p^2 = .415$, suggesting a reliable plausibility effect only in the non-island condition. The following planned paired t -tests confirmed this trend, in that longer RTs in the [non-island, plausible] condition were found to be significantly longer than RTs in the [non-island implausible] condition, $t_1(23) = 6.095, p < .001, d = 1.050$; $t_2(27) = 5.134, p < .001, d = 1.300$, whereas there was no significant plausibility effect in the island condition, $t_1(23) = .993, p = .331, d = .115$; $t_2(27) = 1.487, p = .149, d = .442$. The results of the early ESL group were also similar to those of the NS English group, in that they showed a significant plausibility effect, $F_1(1, 23) = 7.006, p = .015, \eta_p^2 = .259$; $F_2(1, 27) =$

6.493, $p = .017$, $\eta_p^2 = .194$, which was modulated by the island constraints factor, leading a significant 2-way interaction, $F_1(1, 23) = 5.855$, $p = .025$, $\eta_p^2 = .226$; $F_2(1, 27) = 2.278$, $p = .143$, $\eta_p^2 = .078$. There was no reliable difference between the two island conditions when the reading times in the two plausibility conditions in each island group were combined, $F_1(1, 23) = 2.450$, $p = .133$, $\eta_p^2 = .109$, $F_2(1, 27) = 2.205$, $p = .149$, $\eta_p^2 = .076$. The follow-up t -tests for each island condition confirmed that, like the NS English group, the early ESL group also displayed a significant plausibility effect only in the non-island condition (i.e., $RT_{\text{plausible}} > RT_{\text{implausible}}$); non-island condition: $t_1(20) = 3.163$, $p = .005$, $d = .752$; $t_2(27) = 2.901$, $p = .007$, $d = .744$; and island condition, $t_1(20) = .538$, $p = .596$, $d = .097$; $t_2(27) = .593$, $p = .558$, $d = .163$. For the adult ESL group, they showed a significant main effect of island with significantly longer first fixation durations in the island condition, $F_1(1, 23) = 7.657$, $p = .010$, $\eta_p^2 = .221$; $F_2(1, 27) = 17.358$, $p < .001$, $\eta_p^2 = .391$. A significant plausibility effect was found only in the F_1 analysis, $F_1(1, 23) = 6.877$, $p = .014$, $\eta_p^2 = .203$; $F_2(1, 27) = 1.540$, $p = .225$, $\eta_p^2 = .054$, as their reading time in the plausible condition were slightly higher in both island conditions. In contrast to the NS English and early ESL groups, the adult ESL group did not display a significant island by plausibility interaction, when measured by first fixation duration, $F_1(1, 23) = .077$, $p = .783$, $\eta_p^2 = .003$; $F_2(1, 27) = .267$, $p = .610$, $\eta_p^2 = .010$. Although there was no significant interaction of the two factors for the adult ESL group, a planned paired t -test was carried for each island condition for the sake of further identifying whether the non-significance was due to the lack of plausibility effect or due to a significant plausibility effect in both island conditions. The result showed that the source of the non-significant result was due to the lack of plausibility effects in both island conditions: non-island condition: $t_1(27) = 1.540$, $p = .135$, $d = .289$; $t_2(27) = 1.466$, $p = .154$, $d = .361$; and island condition: $t_1(27) = 1.467$, $p = .154$, $d = .311$; $t_2(27) = .429$, $p = .672$, $d = .125$.

First-pass regression The source of a significant interaction of group and island constraints was on the non-significant island effect of the early ESL group, $F(1, 584) = .040$, $p = .841$, whereas the other two groups showed a significant island effect with relatively higher overall regression ratios in the island condition for both the NS English, $F(1, 668) = 22.265$, $p < .001$, and the adult ESL group, $F(1, 780) = 9.732$, $p = .002$. On the other hand, a significant island by plausibility effect was found only in the NS English group, $F(1, 688) = 13.326$, $p < .001$, in that the NS English speakers made significantly more frequent regressions (approx. 10% more) in the [non-island, plausible] condition than in the [non-island, implausible], whereas they showed, albeit to a lesser degree, a reverse pattern in the island condition with approximately 5 percent more regressions in the implausible condition. The early ESL group, similarly to the NS English controls, showed about 10 percent more regressions in the [non-island, plausible] than in the [non-island, implausible] condition. This pattern was the same in the island condition, although the difference between the two plausibility conditions was much smaller (approx. 2%, compared to 10% difference in the non-island condition). However, the results showed no significant interaction of the two factors, $F(1, 584) = 1.931$, $p = .165$. The adult ESL group did not seem to have any significant plausibility effect in either island condition, as they showed only about 2% difference between plausible and implausible sentences for both island conditions, which resulted in no main effect for plausibility, $F(1, 780) = .431$, $p = .512$, and no significant interaction of island and plausibility, $F(1, 780) = .095$, $p = .758$.

Regression path duration & Total RT As reported earlier, the regression path duration and Total RT data at Region3 widely violated assumptions for ANOVA. Therefore, sets of nonparametric Wilcoxon signed ranks test were performed alternatively on each island condition for each group. Overall, the results were in line with the prediction of the garden-path effect in

the [non-island, plausible] condition for both measures. That is, both the regression path durations and Total RTs in the [non-island, plausible] condition were significantly longer than the RTs in the plausible counterpart for all groups; NS English: [regression path duration, $Z_1 = 3.457, p = .001, d = 1.152$; $Z_2 = 2.207, p = .043, d = .563$; Total RT, $Z_1 = 2.857, p = .004, d = .905$; $Z_2 = 2.619, p = .009, d = .747$]; the early ESL group: [regression path duration, $Z_1 = 3.901, p < .001, d = 1.508$; $Z_2 = 2.163, p = .031, d = .604$; Total RT, $Z_1 = 2.902, p = .004, d = 1.002$; $Z_2 = 4.349, p < .001, d = 1.474$]; and the adult ESL group: [regression path duration, $Z_1 = 3.256, p = .001, d = .966$; $Z_2 = 2.095, p = .036, d = .584$; Total RT, $Z_1 = 4.509, p < .001, d = 1.510$; $Z_2 = 4.440, p < .001, d = 1.474$]. In contrast to such reversed plausibility effects that were significant in the non-island condition, no such significance was found within the island condition across the groups; NS English: [regression path duration, $Z_1 = 1.400, p = .162, d = .413$; $Z_2 = 1.275, p = .202, d = .346$; Total RT, $Z_1 = 1.057, p = .290, d = .309$; $Z_2 = .387, p = .699, d = .104$]; early ESL: [regression path duration, $Z_1 = .174, p = .862, d = .054$; $Z_2 = .979, p = .327, d = .264$; Total RT, $Z_1 = .330, p = .741, d = .102$; $Z_2 = .023, p = .982, d = .006$]; and adult ESL group: [regression path duration, $Z_1 = 1.116, p = .256, d = .302$; $Z_2 = .979, p = .327, d = .264$; Total RT, $Z_1 = .638, p = .524, d = .171$; $Z_2 = .911, p = .362, d = .245$].

4.3.2.2. Analysis of reading patterns at the spillover region (Region4)

Table 13 provides descriptive statistics for RTs and first-pass regression (%) across the conditions at Region4. Some different reading patterns were found between the NS English and

Table 13. *Descriptive statistics for RTs and first-pass regressions at Region4*

	Island Cond.	Plausibility Cond.	FFD	F-pass RT	REGR	RPD	Total RT
			<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
NS English	non-island	plausible	238 (57)	356 (153)	.35 (.18)	965 (430)	905 (276)
		implausible	263 (45)	390 (121)	.23 (.17)	678 (338)	886 (292)
	island	plausible	250 (56)	440 (135)	.37 (.13)	976 (267)	1131 (338)
		implausible	249 (51)	401 (137)	.36 (.18)	1023 (497)	1064 (363)
Early ESL	non-island	plausible	291 (52)	534 (172)	.41 (.20)	1425 (1044)	1163 (411)
		implausible	271 (44)	519 (174)	.22 (.17)	743 (245)	900 (305)
	island	plausible	287 (43)	515 (119)	.35 (.24)	1080 (502)	1044 (444)
		implausible	277 (45)	530 (130)	.32 (.21)	1035 (441)	1041 (305)
Adult ESL	non-island	plausible	310 (49)	580 (122)	.46 (.27)	1761 (1123)	1393 (424)
		implausible	286 (41)	556 (131)	.27 (.18)	1030 (501)	944 (322)
	island	plausible	259 (36)	536 (177)	.34 (.21)	1253 (624)	1058 (292)
		implausible	298 (54)	591 (156)	.30 (20)	1168 (536)	1120 (385)

the two ESL groups in the non-island condition. Most notably, for first fixation duration and first-pass RT, the NS English group spent more times in the [non-island implausible] than in the [non-island, plausible] condition, contrary to the pattern they showed at the critical region (Region3) that had significantly longer reading times in the plausible condition. They, however, showed the same patterns for the other measures, with longer reading times and more regressions in the plausible condition. On the other hand, the early and adult ESL groups continued to exhibit

similar reading patterns they previously showed at Region3 for all measures, with longer RTs and more regressions in the [non-island, plausible] condition. The patterns of first-pass regression by the adult ESL group had somewhat remarkable changes across the experimental conditions, compared to the patterns they showed at the previous region.

Recall that the adult ESL learners had slightly higher regression ratios in the [non-island, implausible] condition at Region3, compared to its plausible pair (i.e., approx. 35% in the plausible and 37% in the implausible condition). At Region4, however, they made nearly 20 percent more regressions in the plausible condition (approx. 46% as opposed to 27%), possibly reflecting relatively delayed reanalysis, which were also reflected in their largely increased regression path durations and Total RTs in the non-island condition. In the island condition, the patterns between the plausible and implausible conditions were somewhat mixed across the measures. However, the differences between the two plausibility conditions were generally not large across the measures. The results of the preliminary analyses at Region4 are summarized in Table 14, followed by Figure 14 (first fixation duration & first-pass RT) and Figure 15 (regression path duration & Total RT) that show the reading patterns of the three groups at this region.

Table 14. *Summary of the results of preliminary analyses at Region4*

		by-subject (F_1)				by-item (F_2)			
		df	f	p	η_p^2	df	f	p	η_p^2
first fixation duration	I	1, 70	1.906	.172	.027	1, 27	.895	.352	.032
	P	1, 70	.436	.511	.006	1, 27	.060	.808	.002
	G	2, 70	9.583	.001*	.215	2, 54	31.025	.001*	.535
	I x G	2, 70	2.300	.108	.062	2, 54	2.336	.106	.080
	P x G	2, 70	3.780	.028	.097	2, 54	4.353	.018	.139
	I x P	1, 70	1.645	.204	.023	1, 27	1.962	.173	.068
	I x P x G	2, 70	7.879	.001	.184	2, 54	5.161	.009	.160
first-pass RT	I	1, 70	1.365	.247	.019	1, 27	.431	.517	.016
	P	1, 70	.178	.674	.003	1, 27	.466	.500	.017
	G	2, 70	20.319	.001*	.367	2, 54	76.771	.001*	.740
	I x G	2, 70	2.081	.132	.056	2, 54	2.569	.086	.087
	P x G	2, 70	.164	.849	.005	2, 54	.301	.741	.011
	I x P	1, 70	.004	.948	.001	1, 27	.237	.631	.009
	I x P x G	2, 70	6.427	.003	.155	2, 54	3.825	.028	.124
regression path duration	I	1, 70	1.585	.212	.022	1, 27	.961	.336	.034
	P	1, 70	30.472	.001*	.303	1, 27	19.543	.001*	.420
	G	2, 70	9.401	.001*	.212	2, 54	19.077	.001*	.414
	I x G	2, 70	1.852	.165	.050	2, 54	6.572	.016	.196
	P x G	2, 70	.246	.783	.067	2, 54	.662	.445	.022
	I x P	1, 70	15.978	.001*	.186	1, 27	11.118	.002	.292
	I x P x G	2, 70	.119	.888	.003	2, 54	.624	.539	.023
Total RT	I	1, 70	1.932	.169	.027	1, 27	3.974	.056	.128
	P	1, 70	20.475	.001*	.226	1, 27	19.823	.001*	.423
	G	2, 70	1.762	.179	.048	2, 54	8.607	.001	.242
	I x G	2, 70	3.853	.026	.099	2, 54	9.111	.001*	.252
	P x G	2, 70	2.449	.094	.065	2, 54	2.713	.075	.091
	I x P	1, 70	13.123	.001	.158	1, 27	9.042	.006	.251
	I x P x G	2, 70	5.528	.006	.136	2, 54	6.659	.003	.195
		df	f	p					
first-pass regression	I	1, 2032	.652	.420					
	P	1, 2032	26.308	.001*					
	G	2, 2032	.238	.788					
	I x G	2, 2032	2.093	.124					
	P x G	2, 2032	.504	.604					
	I x P	1, 2032	13.515	.001*					
	I x P x G	2, 2032	.198	.821					

Note. In the first-pass regression analysis, both the subject and item factors were entered as random factors. I = island constraints factor, P = plausibility factor, G = group, * = $p < .001$.

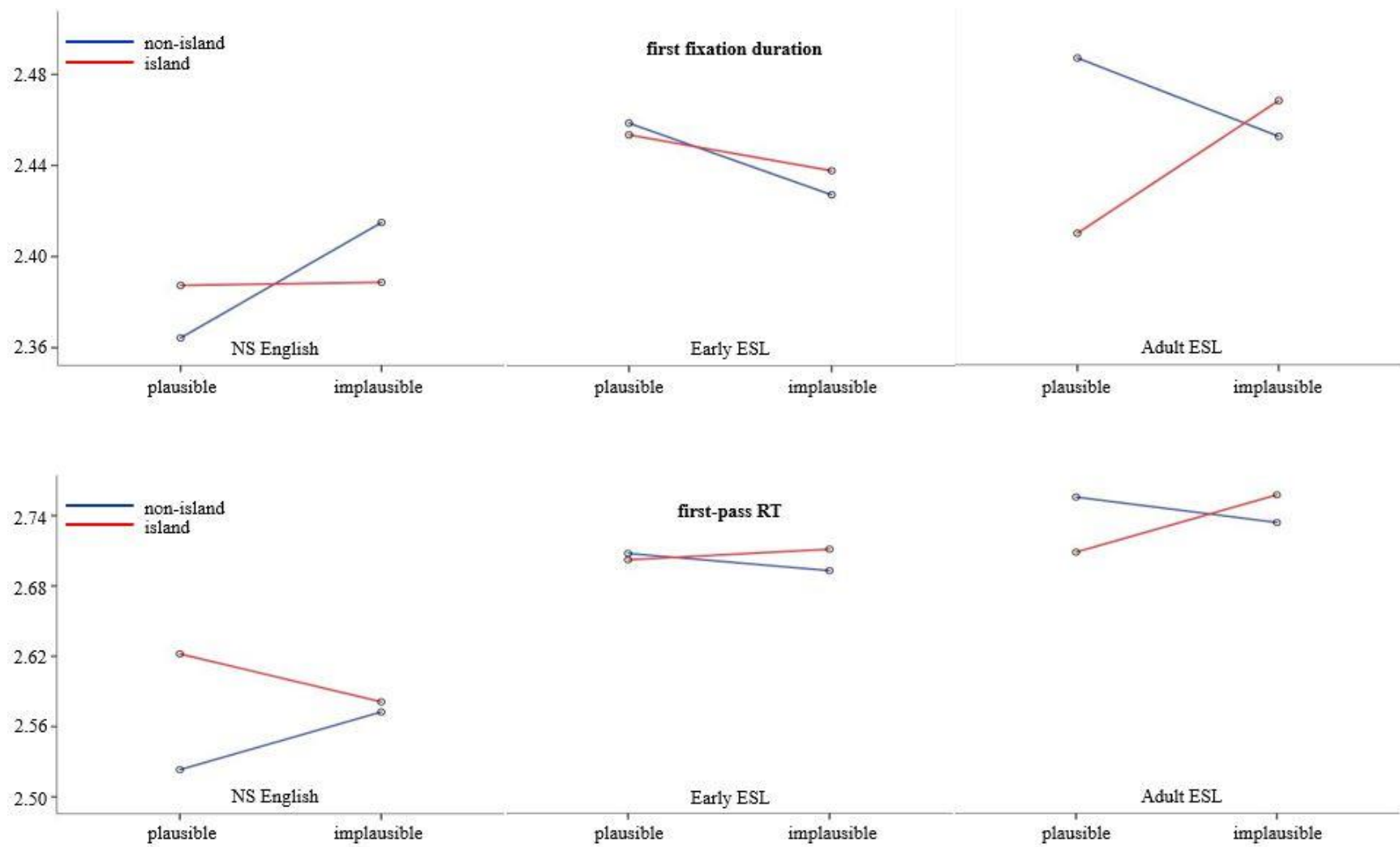


Figure 14. Reading patterns of the three groups during early stages of processing at Region4

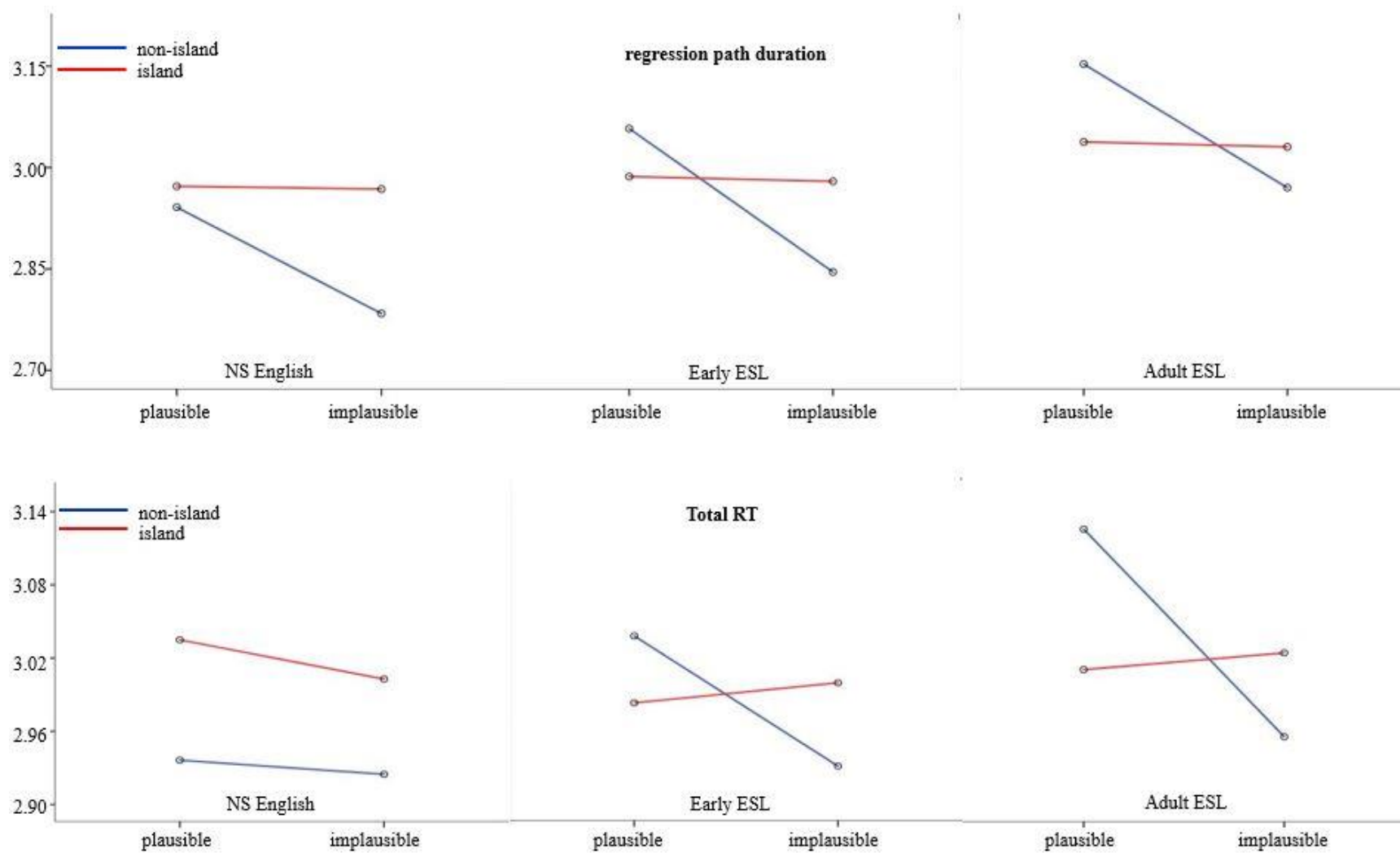


Figure 15. Reading patterns of the three groups during late stages of processing at Region4

First, a significant interaction of island and plausibility was found for regression path duration, first-pass regression, and Total RT, likely indicating the expected (reversed) plausibility effect only in the non-island condition for those measures. The group related interactions were found for all measures except in the first-pass regression. First, a significant group by plausibility interaction was found on first fixation duration ($p_1 = .028$, $p_2 = .018$), and marginally on Total RT ($p_1 = .094$, $p_2 = .075$) and first-pass RT in the by-items analyses ($p_2 = .086$). A significant interaction of group and island was found on Total RT ($p_1 = .026$, $p_2 < .001$) and regression path duration in the by-items analysis ($p_2 = .016$). Lastly, there was a significant 3-way interactions (group x island x plausibility) on first fixation duration, first-pass RT, and Total RT. The results of the follow-up analyses for each group on those four measures are provided below:

First fixation duration & first-pass RT First of all, the early ESL group showed neither main effects nor interaction effect in both measures, suggesting that their reading profiles were similar to one another across the experimental conditions for these two measures: first fixation duration—Island (I) ($p_1 = .927$, $p_2 = .592$); Plausibility (P) ($p_1 = .179$, $p_2 = .187$), ; Interaction (I x P) ($p_1 = .630$, $p_2 = .525$), and first-pass RT—I ($p_1 = .777$, $p_2 = .959$); P ($p_1 = .823$, $p_2 = .939$); I x P ($p_1 = .504$, $p_2 = .518$). For the NS English group, they showed a main island effect on first-pass RT, $F_1(1, 23) = 4.680$, $p = .041$, $\eta_p^2 = .169$; $F_2(1, 27) = 3.999$, $p = .056$, $\eta_p^2 = .129$, in that their RTs in the island condition were significantly longer than those in the non-island condition. The adult ESL group also showed a main island effect on first fixation duration, $F_1(1, 23) = 6.232$, $p = .019$, $\eta_p^2 = .188$; $F_2(1, 27) = 5.112$, $p = .032$, $\eta_p^2 = .159$, in that they spent more time in the non-island (particularly for plausible reading) than in the island condition. In regard to the interaction effect, both the NS English and the adult ESL group showed a significant island by

plausibility effect for both measures, except in the first fixation duration of the NS English group in the by-items analysis: NS English, [$F_1(1, 23) = 4.376, p = .048, \eta_p^2 = .160$; $F_2(1, 27) = 1.772, p = .200, \eta_p^2 = .060$] for first fixation duration, and [$F_1(1, 23) = 6.266, p = .020, \eta_p^2 = .214$; $F_2(1, 27) = 5.870, p = .022, \eta_p^2 = .179$] for first-pass RT; the adult ESL group, [$F_1(1, 23) = 18.369, p < .001, \eta_p^2 = .405$; $F_2(1, 27) = 10.656, p = .003, \eta_p^2 = .283$] for first fixation duration, and [$F_1(1, 23) = 5.754, p = .024, \eta_p^2 = .176$; $F_2(1, 27) = 3.609, p = .068, \eta_p^2 = .118$] for first-pass RT.

However, subsequent paired *t*-tests for each island condition found that the ways the two factors (i.e., island & plausibility) interacted differ between the two groups. First, the NS English group continued to show the same pattern in the island condition, with no significant differences between the two plausibility conditions for both measures, first fixation duration ($p_1 = .962, p_2 = .877$); first-pass RT ($p_1 = .177, p_2 = .221$). In the non-island condition, the NS English group displayed significant RT differences between the two plausibility conditions for both measures; first-fixation duration, [$t_1(23) = 2.954, p = .007, d = .503$; $t_2(27) = -2.187, p = .038, d = .608$], and first-pass RT [$t_1(23) = 2.279, p = .032, d = .351$; $t_2(27) = 1.613, p = .118, d = .405$].

However, as noted above, the direction of the effect was opposite to the pattern they showed at Region3, with significantly longer RTs in the implausible rather than in the plausible condition for both measures. On the other hand, the adult ESL group showed a significant difference in the island condition, rather than in the non-island condition, for first fixation duration, [$t_1(27) = -3.878, p = .001, d = .854$; $t_2(27) = -3.941, p = .002, d = .868$], and marginally for first-pass RT, [$t_1(27) = -1.861, p = .074, d = .366$; $t_2(27) = -1.897, p = .069, d = .493$], in that their RTs in the implausible condition were found to be significantly longer than the RTs in the plausible counterpart. In the non-island condition, the adult learners spent more times in reading plausible sentences for both measures; however, it was only first fixation duration that showed

significance; first fixation duration: [$t_1(27) = 2.297, p = .030, d = .530$; $t_2(27) = 1.736, p = .094, d = .500$.]; and first-pass RT ($p_1 = .318, p_2 = .418$).

Regression path duration & Total RT The follow-up analysis for each group found that the cause of the significant interaction of group and island for regression path duration appeared to be the increased RTs of the two ESL groups in the [non-island, plausible] condition (see Table 12), which increased the overall RTs of the two groups in the non-island condition, contributing to their generally slower RTs in the island condition. Consequently, the two learner groups did not present a significant island effect for regression path duration: the early ESL ($p_1 = .582, p_2 = .878$); and the adult ESL ($p_1 = .597, p_2 = .123$). On the other hand, a significant island effect was found for the NS English group, $F_1(1, 23) = 5.888, p = .023, \eta_p^2 = .204$; $F_2(1, 27) = 9.735, p = .004, \eta_p^2 = .265$, with significantly longer RTs in the island than in the non-island condition. There was a significant plausibility effect for all groups, most likely due to increased RTs in the [non-island, plausible] for all groups; NS English, $F_1(1, 23) = 6.583, p = .017, \eta_p^2 = .223$; $F_2(1, 27) = 2.985, p = .095, \eta_p^2 = .100$; early ESL, $F_1(1, 20) = 13.312, p = .002, \eta_p^2 = .400$; $F_2(1, 27) = 6.738, p = .015, \eta_p^2 = .200$; and the adult ESL, $F_1(1, 27) = 11.819, p = .002, \eta_p^2 = .304$; $F_2(1, 27) = 23.328, p < .001, \eta_p^2 = .464$. There was also a significant interaction of island and plausibility across the groups; NS English, $F_1(1, 23) = 4.527, p = .044, \eta_p^2 = .164$; $F_2(1, 27) = 1.384, p = .250, \eta_p^2 = .049$; early ESL, $F_1(1, 20) = 7.164, p = .015, \eta_p^2 = .264$, $F_2(1, 27) = 9.041, p = .006, \eta_p^2 = .251$, and the adult ESL, $F_1(1, 27) = 4.951, p = .035, \eta_p^2 = .155$; $F_2(1, 27) = 6.077, p = .020, \eta_p^2 = .184$. The subsequent paired t -tests confirmed that the source of this significant interaction was the significant RT differences between the plausible and implausible condition that was found only in the non-island condition for all three groups: NS English, [non-island: $t_1(23) = 3.043, p = .006, d = .769$; $t_2(27) = 2.077, p = .047, d = .497$; island: $t_1(23) = .094, p$

= .926, $d = .025$; $t_2(27) = .251$, $p = .804$, $d = .073$]; early ESL, [non-island: $t_1(20) = 4.406$, $p < .001$, $d = .989$; $t_2(27) = 4.217$, $p < .001$, $d = 1.212$; island: $t_1(20) = .144$, $p = .887$, $d = .040$; $t_2(27) = .179$, $p = .859$, $d = .051$]; and the adult ESL, [non-island: $t_1(27) = 3.575$, $p = .001$, $d = .726$; $t_2(27) = 4.538$, $p < .001$, $d = 1.092$; island: $t_1(27) = .164$, $p = .871$, $d = .038$; $t_2(27) = .408$, $p = .686$, $d = .094$].

Analysis of Total RT also found patterns similar to those in regression path duration, in that it was only the NS English group that showed a main island effect, $F_1(1, 23) = 9.854$, $p = .005$, $\eta_p^2 = .300$; $F_2(1, 27) = 27.514$, $p < .001$, $\eta_p^2 = .505$. On the other hand, a significant plausibility effect was found only in the two learner groups, arguably due to their longer Total RTs in the [non-island, plausible] condition for both groups; early ESL, $F_1(1, 20) = 5.495$, $p = .030$, $\eta_p^2 = .216$; $F_2(1, 27) = 21.454$, $p < .001$, $\eta_p^2 = .443$; and the adult ESL, $F_1(1, 27) = 21.153$, $p < .001$, $\eta_p^2 = .439$; $F_2(1, 27) = 17.507$, $p = .020$, $\eta_p^2 = .393$. The early and adult ESL groups also exhibited a significant interaction of island and plausibility, early ESL, $F_1(1, 20) = 4.828$, $p = .040$, $\eta_p^2 = .194$; $F_2(1, 27) = 5.565$, $p = .026$, $\eta_p^2 = .171$; and the adult ESL, $F_1(1, 27) = 16.453$, $p < .001$, $\eta_p^2 = .379$; $F_2(1, 27) = 34.280$, $p < .001$, $\eta_p^2 = .559$, but not for the NS English group ($p_1 = .548$, $p_2 = .647$). The planned paired t -tests confirmed that the significant interactions found in the ESL groups were due to significant RT differences only in the non-island condition (i.e., $RT_{\text{plausible}} > RT_{\text{implausible}}$); early ESL, [non-island: $t_1(20) = 2.727$, $p < .013$, $d = .684$; $t_2(27) = 4.453$, $p < .001$, $d = .903$; island: $t_1(20) = .590$, $p = .562$, $d = .104$; $t_2(27) = .253$, $p = .802$, $d = .064$]; and the adult ESL, [non-island: $t_1(27) = 5.507$, $p < .001$, $d = 1.190$; $t_2(27) = 7.407$, $p < .001$, $d = 1.626$; island: $t_1(27) = .536$, $p = .597$, $d = .094$; $t_2(27) = 1.307$, $p = .309$, $d = .231$].

4.3.2.3. Interim summary of the results— Ultimate gap

Table 15. Summary of the findings at the ultimate gap

Significant island x plausibility interaction?			Major implications
	(Critical) Region 3	(Spillover) Region 4	
NS English	<u>YES</u> ✓ FFD ✓ first-pass RT ✓ REGR ✓ RPD ✓ Total RT	<u>YES</u> ✓ REGR ✓ RPD	<ul style="list-style-type: none"> The NS English and early ESL groups showed evidence for filler-gap reanalysis from earlier stages of processing when reading in the non-island condition, demonstrating sensitivity to the structural cues that signal the need for a reanalysis.
		<u>NO</u> ✓ FFD ✓ First-pass RT ✓ Total RT	
Early ESL	<u>YES</u> ✓ FFD ✓ first-pass RT ✓ RPD ✓ Total RT	<u>YES</u> ✓ FFD ✓ REGR ✓ RPD ✓ Total RT	<ul style="list-style-type: none"> The adult ESL group showed evidence for filler-gap reanalysis only during late stages of processing, as measured by RPD and Total RT. Their reading patterns during early stages of processing did not present any plausibility effects, indicating delayed gap identifications, compared to the early ESL and NS English group.
	<u>NO</u> ✓ REGR	<u>NO</u> ✓ First-pass RT	
Adult ESL	<u>YES</u> ✓ first-pass RT ✓ RPD ✓ Total RT	<u>YES</u> ✓ FFD ✓ REGR ✓ RPD ✓ Total RT	<ul style="list-style-type: none"> No group displayed plausibility effects in the island condition, indicating no effect of plausibility manipulations. This corroborates the results found at the previous regions that the participants did not postulate a gap in the island environment.
	<u>NO</u> ✓ FFD ✓ REGR	<u>NO</u> ✓ First-pass RT	

4.4. The effect of individual differences in working memory capacity

In order to examine how individual differences in WMC influence the ways the early and adult ESL learners deal with a dislocated filler during online reading of filler-gap dependencies in their L2 English, a series of repeated measures ANCOVA analyses for RT measures, and a logistic random effects regression analysis for first-pass regressions, were carried out separately for each group²⁰.

4.4.1. The effect of WMC at the earliest gap Region1 and spillover Region2

The observation of the parameter estimates ($\hat{\beta}$ coefficient) for the WM span scores showed a general trend that for all three groups, those with higher WMC, compared to those with lower WMC participants in the same group, generally tended to read slightly faster and make less regression (during first pass) both at Region1 and Region2, especially in the non-island condition, but to lesser degrees or occasionally in the opposite direction in the island condition. This appeared to be relatively more so for the adult ESL learners. For example, the regression path duration data of the adult ESL group at Region1 showed that when the WM span score was increased by one unit, the reading times (regression path durations) were reduced by approximately 18 percent²¹, $(1 - (10^{\beta=-.085}))$ in the [non-island, plausible condition], and about 15 percent, $((1 - (10^{-.072})))$ in the [non-island, implausible condition]. On the other hand, the native English speakers' regressing path durations were reduced by about 6.7 percent, and

²⁰ Recall that the WM data from one early ESL learners and one adult ESL learners were removed from the analyses due to their inconsistent performance across the two WM span tests. Thus, the sample sizes of the ESL groups were adjusted to $N = 20$ for the early ESL, and $N = 27$ for the adult ESL group in the analyses.

²¹ As noted earlier, the changes in the reading time outcome variables are presented in percent rather than changes in actual reading times because the raw data were log transformed.

3.3 percent with one unit increase in the WM span, respectively in the same conditions. When considering the fact that the RTs of the ESL learners were much slower than the native English speakers (see Table 6), the degrees of the changes in RTs as a function of the WM score increase would seem to be relatively larger. With this in mind, the results of the repeated measures ANCOVAs and logistic regression analyses at the critical (Region1) and spillover regions (Region2) for each group are provided in Table 16. In the following, the results of each group are reported.

NS English At Region1 and Region2, the NS English group showed neither a reliable main effect of WM nor a significant interaction associated with WM on these two regions across all measures, indicating that different WMCs among the native English speakers did not have much effect on native English speakers' reading behaviors on these regions across different experimental conditions. Compared to the results obtained from the mixed ANOVA analyses reported in the previous sections, the results for the other non-WM-related factors from the ANOCOVA analyses did not show much change. Crucially, the significant interaction of island constraints and plausibility found at Region1 in the ANOVA analyses appeared to remain intact, for all dependent measures.

Table 16. *Summary of the WM effect analyses at Region1 and Region2*

		NS English				Early ESL				Adult ESL			
		Region1		Region2		Region1		Region2		Region1		Region2	
		<i>f</i>	<i>p</i>	<i>f</i>	<i>p</i>	<i>f</i>	<i>p</i>	<i>f</i>	<i>p</i>	<i>f</i>	<i>p</i>	<i>f</i>	<i>p</i>
first-fixation duration	I	6.931	.015	4.065	.056	.600	.449	.305	.588	.868	.360	.269	.608
	P	7.085	.014	5.678	.026	8.491	.009	.121	.732	8.505	.007	2.116	.158
	I x P	10.504	.004	2.106	.161	11.912	.003	14.027	.001	.004	.951	3.942	.044
	WM	.151	.702	.106	.748	3.643	.072	.700	.414	4.333	.048	.455	.506
	I x WM	2.732	.113	1.098	.306	2.751	.114	.762	.394	.083	.775	.170	.684
	P x WM	.126	.726	.021	.887	.007	.935	.017	.899	.190	.667	.000	.992
	I x P x WM	.692	.414	1.686	.208	2.836	.109	1.022	.325	1.933	.177	2.503	.126
first-pass RT	I	5.916	.024	13.308	.001	4.236	.054	.068	.797	10.832	.003	.026	.874
	P	10.854	.003	2.468	.130	6.073	.024	.159	.694	3.499	.073	4.725	.039
	I x P	12.986	.002	.355	.557	5.573	.030	2.520	.130	.499	.486	.893	.354
	WM	1.249	.276	.516	.480	1.381	.255	1.170	.292	2.053	.164	.221	.642
	I x WM	1.902	.182	.746	.397	.575	.458	.663	.426	1.098	.305	.540	.469
	P x WM	.215	.647	1.091	.310	4.484	.048	.770	.392	.299	.590	3.216	.085
	I x P x WM	.041	.841	1.637	.214	2.118	.163	.001	.980	1.858	.185	.315	.580
regression path duration	I	.656	.427	.415	.526	4.128	.057	.025	.877	5.773	.024	.549	.021
	P	7.085	.014	.374	.547	5.492	.310	12.144	.003	12.922	.001	38.794	.001*
	I x P	24.793	.001*	4.400	.048	16.327	.001	9.183	.007	20.841	.001*	18.502	.001*
	WM	.670	.422	.076	.786	1.920	.183	1.444	.245	1.915	.179	.045	.834
	I x WM	.017	.896	.293	.593	.062	.806	.798	.384	.184	.672	3.607	.069
	P x WM	.125	.727	1.095	.307	.735	.403	.353	.560	1.108	.303	2.090	.161
	I x P x WM	.136	.716	.141	.711	1.113	.305	.507	.485	.443	.512	.328	.572

Table 16 (cont'd)

		NS English				Early ESL				Adult ESL			
		Region1		Region2		Region1		Region2		Region1		Region2	
		<i>f</i>	<i>p</i>	<i>f</i>	<i>p</i>	<i>f</i>	<i>p</i>	<i>f</i>	<i>p</i>	<i>f</i>	<i>p</i>	<i>f</i>	<i>P</i>
Total RT	I	15.984	.001	25.966	.001*	1.084	.312	.014	.909	.480	.495	.005	.947
	P	2.706	.114	1.701	.206	5.906	.026	.021	.885	24.767	.001*	.177	.678
	I x P	4.758	.044	.594	.449	11.474	.003	1.022	.325	20.388	.001*	.481	.494
	WM	1.854	.187	.600	.447	.587	.454	1.300	.269	.426	.520	.046	.831
	I x WM	3.225	.085	1.078	.310	.529	.476	1.279	.273	5.159	.032	3.536	.072
	P x WM	.694	.414	.370	.549	.906	.354	.052	.822	1.661	.209	.113	.740
	I x P x WM	1.430	.244	.005	.947	.340	.567	.063	.805	.303	.587	.201	.658
first-pass regression	I	5.579	.018	1.632	.202	9.152	.003	.715	.398	21.346	.001*	.974	.324
	P	6.601	.010	.241	.623	12.060	.001	8.182	.004	8.941	.003	48.943	.001*
	I x P	16.785	.001*	2.606	.107	14.735	.001*	2.559	.110	4.972	.026	25.962	.001*
	WM	.004	.951	.345	.557	.479	.489	.462	.497	.381	.537	.236	.628
	I x WM	.554	.457	.222	.638	.459	.498	.168	.682	.728	.394	5.814	.016
	P x WM	.207	.650	.058	.809	2.295	.130	.058	.810	.178	.178	.061	.805
	I x P x WM	.220	.639	1.532	.216	9.749	.002	1.216	.271	.154	.154	.613	.434

Note. I = island constraints factor, P = plausibility factor, WM = WM covariate, // .001* = $p < .001$.

Early ESL The early ESL learners exhibited WM-related significant interactions on two measures at Region1: a significant 2-way interaction between plausibility and WM for first-pass RT ($p = .048$), and a significant 3-way (WM x island x plausibility) interaction for first-pass regression, $F(1, 18) = 9.749$, $p = .002$, $\eta_p^2 = .130$. As noted in the previous chapter (see section 3.4.2.), the early ESL learners were divided into two sub-groups based on their WM span scores, the higher WM ($n = 10$), and lower WM ($n = 10$), to obtain the descriptive statistics of the two subgroups for those two measures. See Table 17.

Table 17. *First-pass RT and first-pass regressions by higher- and lower-WM early ESL*

	First-pass RT			OR	REGR	
	$\hat{\beta}$	H-WM <i>M (SD)</i>	L-WM <i>M (SD)</i>		H-WM <i>M (SD)</i>	L-WM <i>M (SD)</i>
[non-island, plausible]	.88	432 (158)	468 (245)	.93	0.13 (.16)	.17 (.13)
[non-island, implausible]	.86	540 (266)	584 (248)	.75	0.24 (.15)	.36 (.15)
[island, plausible]	1.08	441 (188)	399 (130)	.72	.29 (.17)	.40 (.18)
[island, implausible]	.84	375 (145)	458 (142)	1.29	.37 (.14)	.27 (.17)

In regard to first-pass RT, the repeated ANCOVA showed a main effect of plausibility ($p = .024$), a marginal effect of island ($p = .054$), and a significant interaction of island and plausibility ($p = .030$), reflecting a clear sign of plausibility effect only in the non-island condition. With respect to the significant interaction of plausibility and WM, $F(1, 18) = 4.484$, $p = .048$, $\eta_p^2 = .199$, the observation of the beta coefficients for the WM span scores hinted that the cause of the significant interaction might be the different reading patterns in the island condition between the two subgroups. In the [island, plausible] condition, the trend was about 8% increase of first-pass RT with one unit increase in the WM span, (i.e., positive direction), but

it was the opposite in the implausible counterpart; the amount of the change was the largest with the negative relationship of 16% in the [island, implausible] condition. This tendency could be observed in the descriptive statistics as well (see Table 15). Whereas the two groups displayed similar reading patterns in the non-island condition (i.e., $RT_{\text{implausible}} > RT_{\text{plausible}}$), the RT pattern of the higher WM group appeared to be reversed in the island condition (i.e., $RT_{\text{plausible}} > RT_{\text{implausible}}$), thus reducing the magnitude of the plausibility effect in the non-island condition (i.e., no or less plausibility effect overall). In contrast, the lower WM group showed a similar pattern across the island conditions (i.e., $RT_{\text{plausible}} < RT_{\text{implausible}}$), thereby likely causing a greater plausibility effect when the island conditions are collapsed to lump the two plausibility conditions together. Consequently, different degrees of the plausibility effect between the two WM subgroups appeared to lead to a significant interaction of plausibility and WM.

The analysis of the first-pass regression data at Region1 also displayed a clear plausibility effect that occurred only in the non-island condition, as revealed by a significant island by plausibility interaction ($p < .001$), and a main plausibility effect ($p = .001$) as a result of the increased regressions in the [non-island, plausible] condition. The 3-way interaction ($p = .002$) found in the analysis is interesting here, as it showed the entirely opposite patterns the two subgroups displayed on their first-pass RT discussed above. The source of the interaction was arguably the peak of the higher WM group in the [island, implausible] condition, in that the probability of making a first-pass regression was increased by about 29 percent ($OR = e^{0.2580}$) with one unit increase in the WM span scores (i.e., positive). The observation of the descriptive statistics showed this trend, in that the mean regression ratio of the higher WM group was about 10 percent higher ($M = .37$, $SD = .14$) than that of the lower WMC group ($M = .27$, $SD = .17$) in the [island, implausible] condition. This pattern was reversed in the plausible counterpart, in that

it was the lower WM group that showed much higher mean regression ratio ($M = .40$, $SD = .18$), compared to the higher WM group ($M = .29$, $SD = .17$). In the non-island condition, both groups had more regressions in the implausible condition. See Table 15 above. As a result, similar regression patterns across the island conditions by the higher WM learners (i.e., $REGR_{\text{plausible}} < REGR_{\text{implausible}}$), accompanied with the different regression pattern between the island conditions (i.e., $REGR_{\text{plausible}} < REGR_{\text{implausible}}$ in the non-island, but $REGR_{\text{plausible}} > REGR_{\text{implausible}}$ in the island condition) by the lower WM learners appeared to be the likely source of the 3-way interaction.

Adult ESL The analysis of the adult ESL learner data showed significant WM-related effects for two measures at Region1, first fixation duration and Total RT, and one measure, first-pass regression at Region2. First, for first fixation duration at Region1, there was a main effect of WM, $F(1, 25) = 4.333$, $p = .048$, $\eta_p^2 = .148$. However, given that there was no WM-related interaction in the analysis, and also considering small values and ranges of odds ratios across the experimental conditions (ranging from .930 to .975), the main WM effect seemed to be a reflection of relatively shorter fixation duration by the higher WM adult learners across the experimental conditions. The results on the other factors remained almost intact compared to the ANOVA analysis performed earlier, with no interaction of island and plausibility ($p = .951$).

At the same region (Region1), the analysis of their Total RT showed a significant island by WM interaction, $F(1, 25) = 5.159$, $p = .032$, $\eta_p^2 = .171$. The parameter estimates for the WM span scores were examined first. It showed that the relationships between the WM span scores and Total RT yielded negative relationships for all experimental conditions, except in the [island, implausible] condition that had a positive relationship: non-island: $\hat{\beta}_{\text{plausible}} = .872$; $\hat{\beta}_{\text{implausible}} = .903$; island: $\hat{\beta}_{\text{plausible}} = .976$; $\hat{\beta}_{\text{implausible}} = 1.061$. This suggests that the RTs of the lower WM adult

learners in the non-island conditions would likely be longer than those of the higher WM adult learners at the least. As before, the adult ESL group was divided into two WM subgroups—higher WM ($n = 14$), lower WM ($n = 13$)—to supplement the interpretation of the interaction. The descriptive statistics showed that whereas both groups exhibited a clear plausibility effect in the non-island condition, [higher WM: $M_{\text{plausible}} = 836\text{ms}$; $M_{\text{implausible}} = 1435\text{ms}$; lower WM: $M_{\text{plausible}} = 1243\text{ms}$; $M_{\text{implausible}} = 1844\text{ms}$], the RTs of the lower WM group in the [non-island, plausible] condition were particularly high, thus likely resulting in their overall reading time in the non-island condition being relatively longer than their overall reading time in the island condition when the two plausibility conditions were lumped [$M_{\text{non-island}} = 2987$; $M_{\text{island}} = 2743$]. On the other hand, the higher WM group appeared to have spent relatively more time in the island condition, [$M_{\text{non-island}}: M = 2271\text{ms}$; $M_{\text{implausible}} = 2767\text{ms}$], presumably to deal with the structurally more complex part of the sentences at this point, while they were generally faster and more efficient in performing filler-gap processing in the non-island condition, compared to the lower WM group. In sum, the significant interaction of island and WM by the adult ESL group thus could be attributed to the lower WM adult learners' heavier processing difficulties in dealing with implausible interpretation in the [non-island, implausible] condition, which appeared to have been extended until the later stages of processing as measured by Total RT.

Lastly, the adult ESL group also showed another significant interaction of island and WM for first-pass regression at Region2, $F(1, 748) = 5.814$, $p = .016$. The ORs for the WM span scores across the experimental conditions showed that the relationship between the WM span scores and the outcome variables (i.e., REGR) turned to the opposite way between the non-island and island condition. Specifically, there was a negative relationship in the non-island condition ($\text{OR}_{\text{plausible}} = .87$, and $\text{OR}_{\text{implausible}} = .79$), signaling decreases in probabilities of making a first-

pass regression (about 13% and 21% respectively) as a function of one unit increase in the WM span scores. In contrast, the odds of making a first-pass regression was increased by about 29 ($OR_{\text{plausible}} = 1.29$) and 37 ($OR_{\text{implausible}} = 1.37$) percent with one unit increase in the WM span scores in the island condition. This trend was also reflected in the descriptive statistics. First, in the island condition, whereas both groups did not show much differences between the two plausibility conditions, overall first-pass regression ratios by the higher WM group ($M_{\text{plausible}} = .26$; $M_{\text{implausible}} = .27$) were higher than those of the lower WM group ($M_{\text{plausible}} = .18$; $M_{\text{implausible}} = .17$), conforming to the positive ORs above. At least partly because of such higher regressions by the higher WM group in the island condition, their overall mean regression ratios in the non-island condition ($M_{\text{plausible}} = .12$, and $M_{\text{implausible}} = .33$) were lower than their own regression ratios in the island condition, despite a peak in regressions in the [non-island, implausible] condition (i.e., $REGR_{\text{non-island}} < REGR_{\text{island}}$). In contrast, the lower WM group appeared to have higher regression ratios in the non-island ($M_{\text{plausible}} = .14$, and $M_{\text{implausible}} = .47$) than in the island condition. As was the case in their Total RT above, it could be interpreted as suggesting that the lower WM adult ESL learners had more processing difficulties in dealing with semantic anomalies in the non-island condition.

For the higher WM adult ESL learners, it is interesting to find that they made relatively more regressions than the lower WM adult learners in the island conditions at this spillover region. In fact, however, such regression patterns that the higher WM adult learners displayed between the two island conditions (i.e., $REGR_{\text{non-island}} < REGR_{\text{island}}$) are more similar to the reading behaviors that the NS English group showed in the island condition at this region, as they were measured by their first-pass regression, regression path duration, and Total RT at this region (see Table 8). One possible explanation would be that more regressions in the island

condition by the higher WM adult learners may be a reflection of their more active filler-gap processing at an earlier point of reading than the lower WM adult learners, in an attempt to construct a dependency between the second filler (i.e., *journalist*) and the verb *wrote* inside the island as early as possible, compared to the lower WM counterparts. That is, when the parser encounters the relative pronoun *who* in the island condition, which signals the opening of another relative clause, the parser must identify another filler to carry (i.e., *the journalist_j* co-indexed with *who_j*) in addition to the filler that the WM already holds (i.e., *the book_i/the city_i*). In the subsequent processing at *wrote*, the parser would need to attempt to link *the journalist_j* (not *the book_i or the city_i*) with its subcategorizing verb *wrote*, forming a filler-gap dependency which is licit (i.e., *the journalist* as the subject of the embedded relative clause). This is a highly complex syntactic computation especially when taking into account the fact that there is still another filler that has not been resolved yet at this point (i.e., *the book_i/the city_i*). This is assumedly why the native English speakers spent more time and made more regressions reading sentences in the island condition even at this spillover region.

At Region2, the higher WM adult learners indeed appeared to spend more time in the island condition than the lower WM adult learners at this region. Although there was only a marginally significant interaction of WM and island constraints, the reading patterns in regression path duration ($p = .069$: $\hat{\beta}_{\text{plausible}} = 1.07$; $\hat{\beta}_{\text{plausible}} = 1.04$) and Total RT ($p = .072$: $\hat{\beta}_{\text{plausible}} = 1.08$; $\hat{\beta}_{\text{plausible}} = 1.04$) were shown to be same as those in the first-pass regression reported above. What is interesting here is that the direction of the relationship between WM span scores and reading times in the non-island condition was negative for those measures, meaning that the higher WM adult learners' reading times in the non-island condition were faster than the lower WM adult learners, similarly to the first-pass regression result discussed above:

regression path duration ($\hat{\beta}_{\text{plausible}} = .92$; $\hat{\beta}_{\text{plausible}} = .92$); Total RT ($\hat{\beta}_{\text{plausible}} = 1.07$; $\hat{\beta}_{\text{plausible}} = 1.04$). Considering these reading patterns in the two island conditions from multiple measures together, it might be reasonable to assume that the higher WM adult learners initiated complex structure building at a relatively earlier point during reading compared to the lower WM adult learners. It should be also noted that more regressions by the higher WM adult learners do not appeared to be the consequence of illicit filler-gap formations (i.e., the first filler *the book_i or the city_i* as the object of *wrote*), given that they did not display any mismatched plausibility effect. This was the case for the lower WM adult learners as well.

4.4.2. The effect of WMC at the ultimate gap at Region3 and spillover Region4

As discussed earlier, Region3 contains the canonical position of the filler, and is the place where a plausibility effect is expected only in the non-island condition, but in a reversed direction to the plausibility effect that was found to be present at Region1. The reason behind this expectation was that an initial filler-gap analysis that results in a plausible interpretation become more challenging for the parser to withdraw it for reanalysis, potentially resulting in longer RTs and more regressions (i.e. $RT_{\text{plausible}} > RT_{\text{implausible}}$). With this in mind, the results of the repeated measures ANCOVAs and logistic regression analyses at Region3 and the following Region 4 for each group are provided in Table 18.

Table 18. *Summary of the WM effect analyses at Region3 and Region4*

		NS English				Early ESL				Adult ESL			
		Region3		Region4		Region3		Region4		Region3		Region4	
		<i>f</i>	<i>p</i>	<i>f</i>	<i>p</i>	<i>f</i>	<i>p</i>	<i>f</i>	<i>p</i>	<i>f</i>	<i>p</i>	<i>f</i>	<i>p</i>
first fixation duration	I	3.972	.059	.003	.959	1.616	.220	.028	.869	9.481	.005	7.048	.014
	P	9.521	.005	5.513	.028	5.612	.029	1.456	.243	6.143	.020	.959	.337
	I x P	28.663	.001*	3.609	.071	9.164	.007	.317	.580	.093	.763	14.042	.001
	WM	.827	.373	.033	.859	1.155	.297	.274	.607	.042	.839	1.142	.295
	I x WM	.056	.815	.265	.612	1.138	.300	1.990	.175	1.865	.184	1.183	.287
	P x WM	1.180	.289	.010	.921	.339	.568	.260	.616	.875	.359	.014	.907
	I x P x WM	2.937	.101	.385	.542	.862	.366	.287	.599	6.392	.018	3.422	.076
first-pass RT	I	21.070	.001*	4.046	.057	44.026	.001*	.001	.992	38.330	.000	.090	.767
	P	1.085	.309	.012	.914	.362	.555	.013	.910	.323	.575	.693	.413
	I x P	10.670	.004	6.416	.019	3.483	.078	.330	.573	.252	.620	4.336	.048
	WM	.899	.353	.052	.822	1.382	.255	.251	.623	1.374	.252	.673	.420
	I x WM	.010	.921	.768	.390	3.254	.088	.333	.571	.181	.674	.032	.860
	P x WM	1.200	.285	.686	.417	.733	.403	1.288	.271	3.532	.072	1.247	.275
	I x P x WM	1.592	.220	.394	.536	.393	.539	.148	.705	1.324	.261	.074	.788
regression path duration	I	50.905	.001*	6.384	.019	12.848	.002	.099	.757	57.671	.000	.704	.410
	P	1.171	.204	2.107	.084	7.415	.014	12.847	.002	1.181	.288	13.894	.001
	I x P	21.814	.001*	3.925	.060	16.985	.001	7.909	.012	4.321	.048	7.114	.013
	WM	.410	.529	.210	.681	2.011	.173	.287	.599	1.430	.243	1.384	.250
	I x WM	.971	.335	.884	.357	2.242	.152	.829	.374	.106	.747	1.946	.175
	P x WM	1.085	.309	2.034	.097	.050	.826	.944	.344	.267	.610	.561	.461
	I x P x WM	.662	.425	1.814	.144	.618	.442	.998	.331	.872	.197	.775	.387

Table 18 (cont'd)

		NS English				Early ESL				Adult ESL			
		Region3		Region4		Region3		Region4		Region3		Region4	
		<i>f</i>	<i>p</i>	<i>f</i>	<i>p</i>	<i>f</i>	<i>p</i>	<i>f</i>	<i>p</i>	<i>f</i>	<i>p</i>	<i>f</i>	<i>P</i>
Total RT	I	139.274	.001*	11.477	.003	11.561	.003	.045	.834	1.865	.184	1.051	.315
	P	8.988	.007	1.171	.291	11.435	.003	4.767	.042	14.626	.001	24.418	.001*
	I x P	5.669	.026	.515	.480	14.652	.001	4.618	.046	4.492	.044	16.698	.001*
	WM	.659	.426	.817	.376	3.382	.082	.575	.458	.103	.751	6.308	.019
	I x WM	1.285	.269	2.150	.157	.090	.767	.683	.420	.207	.653	1.693	.205
	P x WM	5.810	.025	.064	.803	.585	.454	.010	.921	.012	.913	1.489	.234
	I x P x WM	1.755	.199	.626	.437	1.431	.247	.117	.736	.069	.795	1.939	.176
first-pass regression	I	23.089	.001*	4.883	.027	.064	.800	.238	.626	12.652	.001*	1.047	.306
	P	4.326	.042	5.298	.022	12.014	.001	8.767	.003	.612	.434	18.118	.001*
	I x P	13.019	.001*	7.445	.007	3.502	.032	5.737	.017	.035	.852	3.752	.053
	WM	.125	.724	1.078	.300	2.105	.147	.781	.377	.633	.426	4.130	.042
	I x WM	1.071	.301	.011	.915	.001	.995	1.623	.203	.127	.722	.544	.461
	P x WM	.641	.510	1.968	.161	2.339	.127	.744	.389	.007	.935	3.371	.078
	I x P x WM	.649	.421	1.698	.193	.368	.544	.421	.516	8.171	.004	.244	.621

Note. I = island constraints factor, P = plausibility factor, WM = WM covariate, // .001* = $p < .001$.

NS English The results of the NS English group showed a WM-related interaction only on one measure at the critical region (Region3), and there was no WM-related effect in the spillover region (Region4). At Region3, there was a significant interaction between plausibility and WM for Total RT, $F(1, 22) = 5.820, p = .025$. To better interpret this interaction, the parameter estimates ($\hat{\beta}$ coefficient) for the WM span scores were examined first. There was a trend that higher WM native speakers tended to read faster than lower WM counterparts in both [non-island, plausible: $\hat{\beta} = .83$] and [non-island, implausible: $\hat{\beta} = .92$] conditions, but the estimated degree of a decrease was larger in the plausible than in the implausible condition (about 17% vs. 8%). On the other hand, there was a positive relationship in both [island, plausible: $\hat{\beta} = 1.01$] and [island, implausible: $\hat{\beta} = 1.01$] conditions. The descriptive statistics confirmed this trend in the non-island condition, in that the higher WM participants' RTs ($M_{\text{plausible}} = 830\text{ms}$, $M_{\text{implausible}} = 763$) were about 300ms faster than those of the lower WM participants ($M_{\text{plausible}} = 1129$, $M_{\text{implausible}} = 793$) in the [non-island, plausible] condition, with only about 30ms difference in the implausible counterpart. On the other hand, although the parameter estimates for WM span in the island condition, albeit very marginal (1%), was positive (i.e., increases in reading time with increases in WM span), the mean RTs of the two subgroups showed that the higher WM group was slightly faster in the plausible condition, [higher WM: $M_{\text{plausible}} = 1661\text{ms}$, lower WM: $M_{\text{plausible}} = 1812$], displaying some discrepancies with the information from the parameter estimates. In the [island, implausible] condition, the mean Total RT of the two groups showed a marginal difference, [higher WM: $M_{\text{implausible}} = 1748$, lower WM: $M_{\text{implausible}} = 1713$]. Thus, it appeared that the primary source of the interaction was longer plausible reading times of the lower WM group in the both island conditions, particularly in the [non-island, plausible] condition. This might be taken to suggest that while both the higher and

lower WM native speakers experienced more processing difficulties in revising their initially computed plausible dependencies, and the recovery from the misanalysis took longer for those with the lower WMC until later stages of processing.

Early ESL The results of the early ESL group at Region3 and Region4 did not show any significant main effect of WM and WM-related interactions, implying that different WMCs among the early ESL learners did not have much influence on the way they processed the target sentences at these regions. When comparing their results for the other non-WM-related factors to the results from the ANOVA analyses reported in the previous sections, it did not display much changes for both regions, showing a significant interaction of island and plausibility interactions at Region3 for all dependent measures, except first-pass RT that showed an interaction that was marginal ($p = .078$).

Adult ESL Whereas the NS English and early ESL group did not present much effect of WM and associated interactions at those two regions, the analyses of the adult ESL learners' data elicited some more WM-related effects. First, the analysis of the first fixation data showed a significant 3-way interaction (island x plausibility x WM) at Region3, $F(1, 25) = 6.392$, $p = .018$, $\eta_p^2 = .204$. The parameter estimates for the WM span scores revealed that the direction of the relationship between first fixation duration and the WM span scores were all mixed. Specifically, in the [non-island, plausible] condition, the parameter estimates for the WM span scores yielded a positive relationship ($\hat{\beta} = 1.09$) whereas it showed a negative relationship in the [non-island, implausible] condition ($\hat{\beta} = -.94$). The directions of these two were found to be reversed in the island condition although the degrees of the positive relationships were rather small ($\hat{\beta}_{\text{plausible}} = .97$, $\hat{\beta}_{\text{implausible}} = 1.01$). The descriptive statistics of the two subgroups were in line with this trend. In the [non-island, plausible] condition, the mean first fixation duration of

the higher WM group ($M = 347\text{ms}$) was slower than that of the lower WM group ($M = 280$), but they were slightly faster in the implausible counterparts (higher WM: $M = 279$, lower WM: $M = 297$). In contrast, in the [island, plausible condition], the mean fixation duration of the higher WM group ($M = 323$) was faster than that of the lower WM group ($M = 369$). The mean fixation duration of the two subgroups in the [island, implausible] condition was very close to one another, $M = 326$, and $M = 324$, respectively for the higher and lower WM group. To summarize, although the adult ESL group as a whole did not show any significant interaction of plausibility and island interaction ($F < 1$, $p = .763$), it was shown that the higher WM adult learners' reading patterns in the non-island condition (347ms for plausible against 279ms for implausible reading) conformed to the patterns found in the NS English and early ESL group, suggesting that the adult ESL learners with higher WMC detected the need for a reanalysis more immediately as soon as they encountered Region3, compared to their counterparts with lower WMC.

Another 3-way interaction found at the same region (Region3) was on first-pass regression, $F(1, 748) = 8.171$, $p = .004$. The ORs for the WM span scores across the experimental conditions showed that the relationship between the WM span scores and the outcome variables were all negative, indicating that a probability of making a first-pass regression decrease by the amount of OR values (i.e., $1 - \text{OR}$) with one unit increase in the WM span scores: non-island: [$\text{OR}_{\text{plausible}} = .939$, $\text{OR}_{\text{implausible}} = .695$], and island: [$\text{OR}_{\text{plausible}} = .670$, $\text{OR}_{\text{implausible}} = .880$]. Thus, the degrees of decrease in probability of making a regression appeared to be larger in the [non-island, implausible; about .30.5% less] condition, and in the [island, plausible; about 33% less] condition. The two subgroups indeed displayed the largest differences in their mean first-pass regression ratios in those two conditions; [non-island, implausible]: $M = .29$, and $M = .46$ for the higher and lower WM group, and [island, plausible]:

$M = .36$, and $M = .54$ for the higher and lower WM group in that order. Such relatively larger differences in those two experimental conditions appeared to be the source of the differences in the reading patterns between the two subgroups. For the higher WM group, their mean regressions in the [non-island, plausible] condition was slightly higher than its implausible counterpart ($M_{\text{plausible}} = .32$, $M_{\text{implausible}} = .29$), whereas it was the opposite direction for the lower WM group ($M_{\text{plausible}} = .39$, $M_{\text{implausible}} = .46$). These patterns between the two groups were reversed in the island condition; the higher WM group regressed slightly more in the implausible condition ($M_{\text{plausible}} = .36$, $M_{\text{implausible}} = .40$), while the lower WM group made about 7% more regressions in the plausible condition ($M_{\text{plausible}} = .54$, $M_{\text{implausible}} = .47$). As a result, it seemed that at least for the adult ESL learners with higher WMC appeared to show reading behaviors that were similar to the native English speakers as well as the early ESL learners. Lastly, the adult group showed a main effect of WM on two measures, Total RT, $F(1, 25) = 6.308$, $p = .019$ and first-pass regression, $F(1, 748) = 4.130$, $p = .042$, at the spillover region (Region). However, given that there were no other WM-related interactions associated with these main effects, and that the results for the other non-WM-related factors (island and plausibility) remain intact, the effects appeared not to provide any particularly useful information.

4.4.3. Summary of the results— the effect of WM

Table 19. *Summary of the findings— The WM effect*

	Any WM-related Effect?		Major Implications
	Initial gap	Ultimate gap	
NS English	<u>No</u>	<u>YES (Region4)</u> ✓ Total RT	<ul style="list-style-type: none"> The Total RT of the NS English group showed a significant plausibility by WM interaction at Region4, which suggested that NSs with lower WM had more spillover effects in the non-island condition than those with higher WM. The results of the early ESL learners displayed somewhat contradictory patterns. In the analysis of first-pass RT at Region1, it was the lower WM learners that showed evidence for illicit gap postulations in the island environment. However, the results on REGR revealed that this non-native-like pattern was shown by the higher WM learners, thus making it difficult to interpret the direction of the WM effect for this group.
	<u>YES (Region1)</u> ✓ First-pass RT ✓ REGR	<u>No</u>	
Early ESL			
Adult ESL	<u>YES (Region1)</u> FFD Total RT	<u>YES (Region3)</u> FFD REGR	<ul style="list-style-type: none"> The adult ESL learners with higher WMC presented filler-gap reanalysis effects from early stages of processing at Region3, demonstrating more native-like reading patterns.
	<u>YES (Region2)</u> REGR	<u>YES (Region4)</u> Total RT	

CHAPTER 5: DISCUSSION

One of the main concerns in the current L2 processing literature has been whether the types of parsing heuristics and linguistic resources adult L2 learners put to use during online processing are qualitatively similar or different from those used by native speakers of the target language. Whereas the current L2 processing literature provides evidence for both qualitative similarities and differences between L1 and adult L2 processing, Clahsen and Felser (2006a, 2006b, 2006c) suggest through their shallow structure hypothesis (SSH) that the nature of L2 processing by adult L2 learners is fundamentally and qualitatively different from L1 processing. The SSH claims that the type of syntactic representations the parser computes in adult L2 processing are shallower and hierarchically less detailed for two possible reasons: First, the SSH characterizes the L2 grammatical representations of adult L2 learners as being incomplete and divergent from the target language norms. In consequence, the parser fed by such deficient L2 grammar representations is restricted in constructing a sufficiently detailed representation for the input. Second, the SSH views that even if relevant L2 grammar is somehow available (e.g., in the offline), adult L2 learners are less likely to be able to utilize it in real time presumably due to their limited and inefficient L2 processing capacities to rapidly integrate their knowledge into the parse, even at a highly proficiency level. For these reasons, the L2 parser depends largely on non-syntactic representations instead, such as lexical-semantic verb-argument information and pragmatic information, which makes what adult L2 processing fundamentally different from native processing.

The present study attempted to address these issues to provide more insight into the nature of adult L2 syntactic processing, by investigating how advanced early and adult ESL learners deal with a dislocated filler to create an association with its ultimate gap for

comprehension, whether they are able to make use of island constraints in a timely manner to build a structurally detailed grammatical representation for the input, and whether learners' different working memory capacities have effects on their online processing. This chapter discusses the research findings and their implications in the light of the research questions.

5.1. The effect of age of acquisition

In exploring the role of age of acquisition in L2 processing, the first research question sought to examine how advanced early and adult ESL learners (A) process island constraints at the earliest possible gap position (Region1 and Region2), and (B) perform a filler-gap (re)analysis (Region3 and Region4), as they are compared to native English speaker controls. This section discusses the findings obtained from the analysis at the first critical region (Region1) and the following spillover region (Region2). The examples of the four experimental conditions illustrated in (27) ~ (30) are repeated in (31) and (32) below for readers' convenience.

(27) [non-island, plausible 'the book' & implausible 'the city']

The book_i / The city_i that the journalist [Region1] wrote *ti* [Region2] fairly regularly
[Region3] about *ti* was [Region4] named for an explorer.

(28) [island, plausible 'the book' & implausible 'the city']

The book_i / The city_i that the journalist who [Region1] wrote [Region2] fairly regularly
[Region3] mentioned *ti* was [Region4] named for an explorer.

Use of active filler strategy & online application of island constraints Recall that Region1 is the linearly closest gap site for the filler in the non-island condition, but not in the island

condition. In the non-island condition, the parser may postulate a gap at this region in accordance with the active filler strategy, which subsequently should create either plausible or implausible interpretations of the sentences by the experiment design. An implausible interpretation would likely render more processing burden to the parser during the syntactic and semantic integration processes, as measured by increased reading times and more regressions in the implausible than in the plausible condition, thus resulting in a plausibility effect. However, such plausibility effect may not occur in the island condition because there is no structurally posited gap in the grammatical representation; but this should be so only if the parser makes use of full-fledged knowledge of island constraints.

Bearing this in mind, the results at Region1 suggest that all three groups exhibited, by and large, fairly similar reading behaviors across the experimental conditions at these regions. The results on first-pass RT, first-pass regression, regression path duration, and Total RT revealed that all three groups showed a reliable interaction of plausibility and island constraints, largely due to the plausibility effects that were captured only in the non-island condition. Crucially, this interaction was not modulated by the group factor for those measures, which could be taken as suggesting that both the NS English and the two ESL groups successfully blocked illicit filler-gap formations inside the island constructions, arguably by virtue of integrating the appropriate grammatical constraints into the parse from early stages of processing.

In comparing the two learner groups, the early ESL learners appeared to have a certain degree of speed advantage compared to the adult ESL learners. However, apparently both groups showed a plausibility effect only in the non-island across those four measures. The only exception came from the result on first fixation duration, which showed a significant interaction

of plausibility and island that was modulated by the group factor in the by-subject analysis, and marginally in the by-item analysis ($p_1 = .029$, $p_2 = .097$). This was due to the fact that the adult ESL group read implausible sentences relatively more slowly than plausible sentences in both island conditions, although the extent to which the reading times slowed down in reading implausible sentences was relatively smaller in the island condition. The follow-up analysis showed that it was the adult ESL group that did not have a significant plausibility and island interaction, whereas the other two groups showed an expected reliable interaction. One plausible explanation for such different reading patterns between the adult ESL and the other two groups would be that the adult ESL learners' application of island constraints might have been slightly delayed during the very early stage of processing, thus momentarily experiencing a mild plausibility effect in the island condition. However, when considering the results from the subsequently following early measures (first-pass RT and first-pass regression) at this region, it was shown that the adult ESL learners nevertheless made use of the relevant grammatical constraints fairly early, although this might not have been as immediate and efficient as the early ESL learners and the native English speakers.

The results at the following spillover region (Region2) showed somewhat similar, but different reading behaviors among the groups, especially with respect to the degrees of a spillover effect found in the non-island condition. For the native English learners, they did show a plausibility effect only on their first fixation durations in the non-island condition, which was modulated by the island constraints factor, as suggested in the results from the preliminary analysis. This result could be interpreted as a spillover effect carried over from Region1. The results showed, however, neither a reliable plausibility effect, nor an interaction of the two factors in the other subsequent measures, except in the by-subject analysis on regression path

duration ($p = .046$). These results suggest that the native English speakers rapidly overcame the processing difficulties derived from implausible interpretations in the non-island condition at Region1, thus exhibiting no plausibility effect in both island conditions.

On the other hand, the results from the early and adult ESL groups suggest that both groups had rather clearer spillover effects from an early to later stage of processing. For the early ESL group, a significant interaction of plausibility and island constraints was found in their first fixation duration, regression path duration, and marginally on first-pass regression ($p = .077$). Again, the apparent source of those significant interactions was on the significant plausibility effects that took place only in the non-island condition, with longer reading times and more regressions in reading implausible sentences. In contrast, their reading patterns between the two plausibility conditions were comparable to each another in the island condition, as revealed by the subsequent paired t -tests.

The adult ESL learners did not differ much from the early ESL learners in this regard. The adult learners also exhibited clear spillover effects in the non-island condition, reflecting extended processing difficulties in dealing with semantic anomalies, as measured by first fixation duration, regression path duration, and first-pass regression. However, their reading patterns in the island condition showed no such plausibility effect, contributing to a reliable interaction of plausibility and island constraints for those measures.

Taken together, the results of the analysis of the participants' eye-movement data at Region1 and Region2 showed that all three groups actively sought to fill the gap in the non-island condition by postulating a grammatically licit gap at the earliest possible position, which is, by and large, in line with the findings from a number of previous studies (e.g., Cunnings et al., 2010; Juffs, 2005; Kim et al., 2015; Omaki & Schulz, 2011, Traxler & Pickering, 1996, among

others). Keep in mind that although this was not the main focus of the present study (i.e., plausibility effect in the non-island condition at the earliest gap), this finding provides a crucial basis for evaluating the reading patterns in the island condition, specifically in regard to whether such plausibility effects observed in the non-island condition would disappear by virtue of the filler's application of the island constraints in real time. That is, if the parser makes use of the active filler strategy (i.e., processing) without much consideration of computing syntactic details such as island constraints, relying instead lexical-semantic and verb-argument information as the SSH would predict for L2 processing, then what would follow is that the same plausibility effect must take place in the island condition as well. However, no groups in the present study showed such plausibility effect in the island condition at these earliest possible gap sites. Thus, this could be taken as suggesting that both the early and adult ESL learners successfully deployed the knowledge of island constraints during their initial processing and blocked the parser's illegal gap formations inside the island constructions. This finding is generally in line with the result of the Chinese [-wh] and German [+wh] ESL learners in Cummings et al.'s (2010) eye-tracking reading, German ESL learners in Felser et al.'s (2012) eye-tracking reading, and Spanish [+wh] ESL learners in Omaki and Schulz's (2001) and Kim et al.'s (2015) self-paced reading studies, although it should be noted that some of those studies found significant interaction of plausibility and island constraints only on later measures, or at the spillover regions.

As discussed, the only exception came from the adult ESL learner group on their first fixation duration at Region1, which showed the similar reading patterns in both island conditions, with longer reading times in the implausible than in the plausible condition. This result may be consistent with the result of Korean ESL learners in Kim et al.'s self-paced reading experiment to some extent, which found evidence for L2 learners' use of island constraints in

their stop-making-sense judgment task, but not during their self-paced reading. However, more finely grained eye-movement data analyzed in the present study showed that such tendencies were not carried over to the following processes at the same region (e.g., first-pass, first-pass regression), as well as at the following region—i.e., no spillover effect in the island condition, whereas there was a great deal of spillover effects in the non-island condition. There may be a couple of possible accounts that could explain the different results between the current study and Kim et al. in terms of adult ESL learners' gap creations in the island environment. First, as also noted by the authors, Kim et al.'s self-paced reading was accompanied with the stop-making-sense judgment task, which might have potentially added more task burden to the adult ESL learners as they had to do the two different tasks simultaneously (i.e., segment-level reading and stop-making-sense judgment). This consequently might have affected learners' reading behaviors. On the other hand, in the current study, the participants read sentences in a more natural way as there was no other task required during reading. Another possibility to consider is that, adult learners' different amount of L2 immersion experience between the two studies might have resulted in different results between the two studies. Although this is an estimation based on the information from their article, comparing the participants' length of residence, it appears that overall the adult ESL learners in this study had more exposure to an English-speaking environment ($M = 4.76$, .33-20 years), compared to the Korean ESL learners in Kim et al.'s study ($M = 3.6$, 1-8 years). When taking into account some empirical evidence that shows a positive relationship between the amount of exposure and L2 processing, generally more immersion experience by the adult ESL learners in the current study might have allowed them to deploy the relevant grammatical information more efficiently (e.g., Dussias & Sagarra, 2007; Frenck-Mestre, 2002, 2005; Pliatsikas & Marinis, 2013)

The effect of filler-gap (re)analysis Region3 includes the canonical position of the filler for all experimental condition, but its function is slightly different between the two island conditions. For the non-island condition, this is the second and ultimate gap site where the parser must cancel its earlier misanalysis at Region1 and perform an immediate reanalysis as soon as it identifies the missing of an object of the preposition. This reanalysis can be more challenging for the parser especially when it needs to give up its initially constructed plausible interpretation (due to readers' deeper commitment). A reanalysis in the implausible condition may be relatively easier, when taking into account the fact that this region may be the very place for the parser to resolve the mystery of the implausible interpretation it experienced at Region1. Such mismatched processing difficulties (i.e., reanalysis effect) between the two plausibility conditions would present a plausibility effect, but in a reverse way to the plausibility effects observed at Region1, with longer reading times and more regressions in the plausible than in the implausible condition. On the other hand, Region3 serves as an initial gap site in the island condition given that there should be no gap postulation at the previous regions. Therefore, the parser should be free from a reanalysis effect in the island condition, meaning that both plausible and implausible sentence reading should yield comparable reading patterns.

With this in mind, the results at Region3 showed the expected reading patterns for all groups, at least when it comes to the late measures, namely regression path duration and Total RT. As the results from the nonparametric tests and their mean descriptive statistics revealed, all three groups showed an expected reanalysis effect with longer reading times in reading plausible than implausible sentences in the non-island condition. Crucially, no group showed a significant plausibility effect in the island condition.

Some group differences were found in the analysis on some of the early measures, on first fixation duration and first-pass regression in particular. For first fixation duration, it was the adult ESL group that showed a difference. Whereas both the NS English and the early ESL group showed a reliable reanalysis effect that was modulated by the factor island constraints, there was no significant interaction of the two factors for the adult ESL group—partly due to the fact that their reading patterns were similar in both island conditions with slightly longer reading times in the plausible condition, and also partly due to the lack of significant plausibility effect in both island conditions, as tested in separated paired *t*-tests—compared to the other groups. This may imply that the adult ESL learners might not have detected the need for a reanalysis as rapidly and efficiently as the early ESL learners and the native English speakers at this point (i.e., less sensitive to a gap identification). Perhaps this might have been a similar case in their first-pass RT as well. Although the preliminary analysis yielded a significant interaction of plausibility and island constraints which was not modulated by the group factor, the difference between reading plausible and implausible reading in the non-island condition that the adult ESL group presented in their descriptive statistics was relatively smaller, compared to the other two groups. Assuming that adult ESL learners' online processing is generally less efficient than the early ESL learners, or the native English controls at the least, the task of reanalysis would have been more burdened for the adult learners, meaning that they could have displayed greater processing difficulties in revising plausible initial misanalysis, just as was the case in their regression path duration (1761ms and 1030ms, for plausible and implausible reading) and Total RT (1393ms and 944ms, for plausible and implausible reading). Considering that, it seems reasonable to interpret that the adult learners' reanalysis process might not have started yet during their first-pass reading.

Another piece of evidence that is in line with this interpretation comes from adult ESL learners' first-pass regression patterns. In their first-pass regression, the adult ESL group did not exhibit much difference across the experimental conditions. Furthermore, in contrast to the other two groups that exhibited a clear reanalysis effect on their first-pass regressions (approx. 10% more regression in the non-island plausible condition compared to the implausible counterpart), the adult ESL learners even made slightly more regressions in reading implausible sentences (approx. 2%), suggesting no clear reanalysis effect.

Finally, the results at the following spillover region (Region4) showed spillover effects for all groups. For the NS English group, they showed a significant reanalysis effect on first-pass regression and regression path duration with more regressions and longer reading times in the plausible than in the implausible condition. This was so only in the non-island condition, as revealed by a significant plausibility effect interacting with island constraints and the subsequent pairwise comparisons for both measures. The two ESL groups also showed a clear spillover effect on those measures with the patterns comparable to the NS English controls. In addition, it seems that their processing difficulties for reanalysis lasted longer, as their Total RTs also yielded a significant main plausibility effect that was interacting with island constraints, whereas the NS English group showed no such effect in this late measure.

Taken together, the results of analysis on the participants' eye-movement data at the ultimate gap Region3 and the following Region4 suggest that the native English speaker controls and the early ESL learners patterned similarly to one another, in that both groups presented longer reading times and more regressions in reading plausible than in reading implausible sentences in the non-island condition across the measures at Region3, suggesting that they identified a gap and initiated a reanalysis from a fairly early stage of processing at this region,

and this was so only in the non-island condition, and no plausibility effect appeared to be present in the island condition. The reading profiles of the adult ESL learners also suggest that the filler-gap reanalysis took place from the critical region (Region3), but this was the case for the two late measures (regression path duration and Total RT) only. This could be taken to suggest that although the adult ESL learners were slower in identifying a gap and initiating a reanalysis, they were eventually able to do so during later stages of processing at the least. Finally, the adult ESL group, like the other two groups, showed no plausibility (or reanalysis) effect in the island condition, displaying comparable reading patterns between the two plausibility conditions. This last result in the island condition corroborates the findings that the adult ESL learners avoided illicit filler-gap formations at earlier regions in the island condition.

Based on the results discussed above, the findings of the present study provide evidence for qualitative similarities between L1 and L2 processing, in that both the early and adult ESL learners demonstrated sensitivity to the relative clause island constraint, as was the case for the native English controls. As discussed above, island configuration entails abstract and hierarchically detailed locality constraints that restrict an extraction of the filler from certain structures such as relative clause (islands) under investigation, which according to the SSH may not be available to those adult ESL learners for use during online processing. Evidence for non-application of island constraints (i.e., shallower processing) would have been a plausibility effect in both non-island and island conditions, with longer reading time in reading implausible sentences at Region1, and with longer reading time in reading plausible sentences at Region3. Contrary to what the SSH would predict, however, the adult ESL learners displayed no such plausibility effect in the island environment across the board, except in their first fixation duration at Region1. This demonstrates that the adult ESL learners made use of the relevant

knowledge of syntactic constraints to build suitable structural representations of the island constructions from the early stages of processing, although they might not be as fast and efficient as the early ESL learners and the native English speakers when deploying such knowledge.

In comparing the performance of the adult ESL learners to that of the early ESL learners, the early learners appeared to have some processing advantages over the adult learners; their reading was relatively faster, and their application of the constraints appeared to have occurred slightly earlier than that of the adult ESL learners that showed no plausibility effect modulated by island constraints in their first fixation duration at the initial gap site. In addition, the early ESL learners patterned more similarly to the native English speakers when it came to the reanalysis processes at the ultimate gap. As noted earlier, both the NS English and early ESL group showed a clear reanalysis effect at this region from the very early stages of processing, suggesting that they rapidly identified the structural gap that needs to be filled, and initiated the reanalysis early. On the other hand, the adult ESL group showed this effect in the late measures only (regression path duration & Total RT), meaning that their identification of a structurally posited gap was delayed until later stages of processing. The result of the adult ESL learners on the filler-gap reanalysis is in line with the results found in previous studies (e.g., Felser et al., 2012; Williams et al., 2011; Williams, 2006, but cf. Kim et al., 2015). Of those studies, Felser et al. (2012) claimed that learners' weaker sensitivity to structural information during early stages of processing and availability of relevant grammatical representations (e.g., structural cue, such as a missing object of the preposition at Region3) only during later stages of processing suggest that these learners may employ shallow processing because it would work faster for them. However, delayed processing relative to native speakers should not undermine the fact the adult ESL learners had the relevant knowledge of filler-gap representations and utilized that

knowledge for the filler-gap processing (Juffs & Rodriguez, 2015, see also Dekydtspotter et al., 2006). In other words, it should be construed as more of quantitative rather than qualitative difference between L1 and L2 processing (i.e., efficiency rather than representational deficit).

It should be also noted that the early and adult ESL groups in the present study were not matched in terms of their L2 proficiency and length of residence, in that the proficiency test score of the early ESL group was significantly higher than that of the adult ESL group; although their self-rated proficiency on L2 reading and L2 grammar were not statistically different from one another. The length of residence by the early ESL group ($M = 16.44$ years) was also significantly longer than that of the adult ESL group ($M = 4.76$ years), thus potentially rendering some advantages to the early ESL group. As a result, relatively less efficient and less immediate processing performance by the adult ESL group might not necessarily be due to their late ages of immersion, as the aforementioned factors (proficiency and length of residence) might have functioned as confounds.

To summarize, the results of the adult ESL learners in the present study do not support the claims of the SSH that adult L2 processing is fundamentally and qualitatively different from L1 processing, and the grammatical representations used by adult L2 learners during online processing are shallower and structurally less detailed. Despite the cross-linguistic difference between their L1s and the target language, the adult ESL learners (wh-in-situ: L1 Korean & L1 Chinese) in this study showed that they have acquired relevant syntactic representations of wh-constructions (wh-movement) in the L2 (e.g., White & Juffs, 1998). Although their processing was not as immediate and efficient as the early learners and the native speakers (either because of late ages of immersion, relatively shorter length of residence, lower proficiency, or any combinations of these three factors), the results demonstrated clearly that the

adult ESL learners made use of the relevant and detailed structural representations early during their online reading. The reading patterns that the adult ESL learners showed in the present study have several methodological and theoretical implications for the discussion of the age-related effects in L2 processing and learning. First of all, the results of this study are not compatible with the claims of Johnson and Newport (1989, 1991) and Dekeyser (2000) that suggested gradual declines in L2 performance and L2 learning ability over age of arrival up until puberty (around the age of 17), with poorer L2 performance and larger performance variability for adult (or late) learners whose L2 immersion occurred in the adulthood. Although the present study did not perform a separate analysis to examine whether there were linear decreases in L2 performance as a function of age (of arrival) for each ESL group, the adult ESL learners demonstrated reading behaviors comparable to those of the early ESL learners, which were not different from those of the native English speaker controls in many respects. One possibility that could account for such different findings between this study and Johnson and Newport and Dekeyser's study might be different research methods. Using the eye-movement monitoring techniques, this study tested learners' knowledge and usage of L2 grammar in a more naturalistic reading setting, and analyzed various types of eye-tracking measures to track down their processes of reading more in detail from early to later stages of processing. On the other hand, the two aforementioned studies used the audio grammaticality judgment tasks. As discussed earlier, however, with the type of the data the GJT provides, it is difficult to determine on what basis learners come to their grammaticality judgments. Thus, there is a risk that learners' judgments might have been based on a number of different factors within and across the participants. In addition, the auditory task might have been more challenging than the reading task for adult L2 learners in their study (cf. Johnson, 1992). Furthermore, given the fact that the mean ages of the early and late learner

groups in Dekeyser's study were quite high (43.2 and 60.00 years old, respectively), with the oldest age of the participant at the time testing was 81 years old, the auditory task might have been difficult especially for some of the elderly participants.

Another implication of the findings of the present study may be related to the adult ESL learners' high levels of education and/or extended period of schooling in the L2 environment. As discussed earlier, Birdsong (2014) found from his analysis of Dekeyser's (2000) data that these two factors (i.e., years of schooling and levels of education) were strongly correlated with learners' grammatical proficiency. Similarly, Hakuta et al. (2003) also suggested the role of formal education in L2 learning. The authors analyzed the U.S. Census data of nearly 2.3 million immigrants with L1 Spanish or L1 Chinese backgrounds, and found that the amount of formal education was a one of the crucial factors in predicting how well those immigrants learn the L2. In this respect, all participants in the present study were at least college graduates, with 24 graduate students (18 doctoral and 6 MA students), and one MA graduate and 3 participants with a Ph.D. degree at the time of their participation, all of whom have had higher and/or graduate education in the United States. It is not entirely clear how exactly years/levels of education affect the development of L2 grammar and L2 processing. Note, however, that the structurally complex relative clause constructions tested in the current study are more frequently provided in written texts. When taking this into account, one plausible account might be that perhaps these adult ESL participants might have had more extensive reading experience through their studies and careers, which led them to have sufficient processing experiences with more written input, including long-distance filler-gap dependency and relative clause constructions. As a result, the adult ESL learners in this study likely had more opportunities to develop their knowledge of the

target language grammar in question, as well as the parsing abilities to make use of the acquired knowledge in real time effectively.

A theoretical implication, but related to the discussion above, is that the amount, type (written/spoken), and quality of L2 (classroom/naturalistic) input may play a crucial role in the development of L2 grammar and processing, especially for adult L2 learners. Thus, although a general consensus is that adult L2 learners are less efficient and less consistent than early L2 learners, it does not necessarily mean that they (adult L2 learners) are restricted in their acquisition of L2 grammar and processing heuristics. With more exposure to the naturalistic and relevant L2 input and more processing practices, it may be possible for adult L2 learners to develop their knowledge of L2 grammars and parsing mechanisms.

The last implication, which is both theoretical and methodological, is that L2 learners' exposure to the target language input prior to their L2 immersion may need to be treated with more caution when investigating age-related effects in L2 acquisition and processing. As reported earlier, all adult ESL learners in the present study had received formal English instruction from their home country at early ages prior to their arrival (mean age = 11.61, $SD = 1.59$). Although there has been a tendency in recent L2 critical period and processing research to consider age of arrival as a more important factor in determining status of L2 learners (i.e., early versus late/adult learners), the adult learners' L2 experience at early ages might also influence the way they learn and process the target language structures in the long term, at least to a certain degree, especially when the target language in question is more commonly taught and used second/foreign languages such as English, the input of which is not only accessible from their classroom settings, but also available from outside of their classroom as well especially in these days (e.g., internet and TV shows). Considering this possibility, then it is an empirical question

whether adult L2 learners whose exposure to the target language before puberty is close to zero can develop their L2 grammar and processing skills, through rich and quality L2 input, to a degree comparable to the early L2 learners or native speakers of the target language.

5.2. The role of working memory in L2 processing of island constraints

The second research question sought to investigate the potential role of individual differences in working memory capacity (WMC) on L2 processing of island constraints. Specifically, the main interests were in how different WMC of individual learners affects (2.1.) the way they respond to the plausibility manipulation at Region1, and (2.2.) the way they apply the relevant grammatical representation of island constraints at Region1 and Region3, and (2.3.) the way they perform a filler-gap reanalysis at Region3.

The analysis of the WM data together with the eye-tracking measures of the three groups brought some interesting results, especially for the adult ESL learners. For the native English speakers, the results yielded neither a significant WM effect nor a WM-related interaction for any measure across the regions, except one. As a result, even after the WM effect has been controlled in the model, the other non-WM related effects appeared to have remained for most of the time across the measures, compared to the results from the ANOVA analyses reported earlier, suggesting a non-significant role of WM on online processing of island constraints for this group. The only exception that yielded a significant WM-related effect was on their Total RT at Region3, where a reanalysis effect was expected (e.g., *about was*). The result yielded a significant interaction of plausibility and WM ($p = .025$). The follow-up analysis found that those with lower WMC native English speakers had more processing difficulties reading plausible sentences in the non-island condition, reflecting a greater reanalysis effect, compared to

those with higher WMC. When considering that the measure was Total RT, it could be taken to interpret that the lower WMC native speakers might have needed extended period of time to recover from initial misanalysis until the later stages of processing at this region. However, given the fact that this was the only result that showed a WM-related effect among the measures, this result must be taken with caution.

The results of the early ESL group also provided non-significant WM effects for most measures across the regions. The results of the other non-WM related factors also did not show notable changes, compared to their results from the ANOVA analyses. There were only two cases that showed a significant WM-related effects for this group, which were first-pass RT at Region1 that yielded a significant 2-way interaction between plausibility and WM ($p = .048$), and a significant 3-way interaction between WM and the other two factors ($p = .002$) at the same region. For first-pass RT, it was identified that the cause of the significant interaction was the lower WM early learners' slower reading with implausible sentences in both island conditions, whereas the higher WM learners showed this trend only in the non-island condition. Given the fact that Region1 is the critical point that tests whether the parser makes use of the island constraints and thus avoids a gap postulation inside the relative island structure, the pattern that the lower WM early learners presented in the island condition might be construed as a result of ungrammatical gap creation at the verb (i.e., *who wrote*___*) in the island condition during initial processes, as was the case for the adult ESL learners in their first fixation duration at the same region. However, a problem of coming to such a plausible interpretation came when encountering a result of the 3-way interaction on their first-pass regression at the same region. That is, the analysis of this interaction revealed that the regression patterns of the two WM subgroups in the island condition were in the opposite direction they showed on first-pass RT

discussed above. Specifically, it was the higher WM learners that had more regressions in reading implausible sentences in both island condition. When taking into account these two conflicting findings in the lack of additional information from other measures (e.g., WM-related effects), it seems difficult to determine at this point how exactly early ESL learners' different WMCs played a role on their processing of island constraints.

The results of the adult ESL group yielded some interesting findings. At Region1 and Region2, the lower WM adult learners showed a tendency to read more slowly and make more regressions in the non-island than in the island condition, despite the fact that the sentences in the island condition were structurally more complex to parse. This was largely attributable to the greater magnitude of the plausibility effect that reflects their struggles in dealing with implausible interpretations in the non-island condition, compared to the higher WM adult learners. In contrast, the higher WM adult learners showed a trend that they spent more time and made more regressions than the lower WM adult learners in the island condition at Region2, as measured by first-pass regression, regression path duration, and Total RT, and this appeared to be conforming to the reading patterns of the NS English group. Interestingly, this trend was found to be reversed in the following region (Region3) between the higher and lower WM adult learners. That is, at Region3, it was the lower WM adult learners that spent more time and made more regressions than the higher WM adult learners in the island condition. Taken together, these results could be interpreted to suggest that the higher WM adult learners might have initiated processing of structurally complex embedded relative clauses earlier than the lower WM adult learners, in order to resolve the filler-gap dependency between the second filler (e.g., *the journalist*) and the embedded relative clause verb (e.g., *wrote*). On the other hand, the lower WM adult ESL learners' increased reading time and more frequent regressions in the island condition

at Region3 suggest that their construction of the embedded relative clause was initiated with more delays.

Another intriguing finding with respect to the role of WM came from Region3, which showed a significant 3-way interaction (WM x plausibility x island constraints) in two measures, first fixation duration and first-pass regression. The analysis of those interactions found that while the adult ESL learners as a whole group showed evidence for filler-gap reanalysis only from later stages of processing (regression path duration and Total RT), the higher WM adult learners were shown to display reading patterns that were similar to those of the early ESL learners and the native English speakers on those two early measures. First, for first fixation duration, the results showed that the higher WM adult learners spent more time in reading plausible than implausible sentences in the non-island condition (i.e., reanalysis effect), whereas the lower WM adult learners showed the opposite reading pattern, spending more time in reading implausible sentences. In the island condition, both WM subgroups showed comparable reading patterns in reading plausible and implausible sentences with no sign of a plausibility effect. The result on the first-pass regression showed a similar pattern, in that the higher WM adult learners made more regressions while reading plausible sentences in the non-island condition (i.e., reanalysis effect), whereas the lower WM adult learners showed the opposite pattern with more regressions in reading implausible sentences. These results may be taken to suggest that the higher WM adult learners were more sensitive in identifying a structurally posited gap than the lower WM adult learners at the ultimate gap position, thus initiating filler-gap reanalysis from earlier stages of processing (see Dussias and Piñar, 2010 for a similar result).

To summarize, the results showed some evidence that individual differences in working memory capacity do influence the way adult ESL learners process filler-gap dependency

constructions. First, lower WM adult learners did not have a disadvantage of making use of the syntactic information during online processing, compared to the lower WM adult learners, in that the adult ESL learners did not attempt to postulate a gap in an island environment, as shown by significant interaction of plausibility and island that is not modulated WM. Second, the effect of WMC were found in the non-island condition, in that the lower WM adult learners tended to carry more processing difficulties in dealing with implausible interpretations. This might have drained their relatively limited cognitive resources, consequently making them less sensitive to the gap identification at the ultimate gap during early stages of processing. On the other hand, the higher WM adult learners showed relatively faster recoveries from the implausible interpretations, and demonstrated reading behaviors that were conforming to those of the early ESL learners and the native English speaker controls at the ultimate gap, by presenting evidence that they were sensitive to the structural cues for the filler-gap reanalysis.

The SSH posits that the fundamental differences between adult L2 and L1 processing cannot be explained by different cognitive resource capacities such as working memory, although they confusingly stated in their footnote "... unlike the grammar, parsing is subject to time constraints and capacity limitations... Computationally complex sentences.... tend to be difficult to process even though they are licensed by the grammar" (Clahsen & Felser, 2006b, p.123). However, based on what has been discussed above, the results of the present study do not fully support the claims of the SSH. First of all, the adult ESL learners did not show any evidence through their reading behaviors that they failed to utilize the syntactic information in the first place. Second, the fact that the results of this study yielded some significant WM effects on adult learners' online processing behaviors suggest a possibility that the claims of the SSH may not be applied to all adult L2 learner population. These need further research however.

CHAPTER 6: CONCLUSION

This dissertation mainly aimed to explore the nature of second language (L2) online processing to provide a better insight into ‘how’ and ‘what’ types of parsing mechanisms and information resources second language learners bring into real time processing to manage the target language input, and ultimately achieve comprehension. Given that no language can be acquired or learned without using it, understanding this HOW and WHAT may provide us with valuable information that will help advance our understanding of how learners develop their interlanguage grammar system through the input they encounter.

One of the primary issues in the current L2 processing literature is whether adult L2 learners can make use of relevant grammatical information to construct fully detailed and appropriate structural representations to accommodate incoming target language input, like native speakers do. In this regard, one position stands that the nature of adult L2 processing is qualitatively different from L1 processing, mainly due to adult learners’ L2 grammar that feeds the parser being “incomplete, divergent, or of a form that makes unsuitable for parsing” or because of their limited ability to compute detailed grammatical representations during online processing even if they have one (Clahsen & Felser, p.118). The current study attempted to test the validity of these claims by investigating the way advanced early and adult ESL learners deal with structurally complex relative island constructions during filler-gap processing. Additionally, learners’ working memory capacity was also measured to examine whether individual learners’ different resource capacity has a role in their online application of grammatical representations.

Overall, the results demonstrated that although there was a slight delay, compared to the early ESL learners and the native English speakers, the adult ESL learners were sensitive to the structural cues (e.g., relative pronoun *who*), and were able to deploy the relevant and

hierarchically detailed knowledge of island constraints in a timely manner from the early stages of processing, thus avoiding postulating an illicit gap inside the island environment. The adult learners' filler-gap reanalysis occurred with delays in that they showed evidence of filler-gap reanalysis only during later stages of processing. However, the result of the working memory analysis revealed higher working memory adult learners had reading patterns similar to those of the early ESL learners and the native speaker controls, showing evidence of filler-gap reanalysis at earlier stages of processing compared to the lower working memory adult learners.

Based on the results discussed above, online processing of filler-gap dependencies and the application of the structurally complex island constraints performed by the early and adult ESL learners in this study were not qualitatively different from the native English speakers. The adult learners were relatively less efficient and less sensitive in identifying the gap at some point and during certain stages of processing, but they showed evidence that they had the knowledge of the target structures under investigation, they made use of the knowledge during their online reading, and that they managed to comprehend the target sentences accurately.

One of the contributions of the current study is that the use of eye-tracking added more ecological validity to the reading task, and provided the participants with a more naturalistic reading environment. Given the complexity of the target sentences used in this study, implementing other types of methods such as the noncumulative self-paced reading would have added more task burden to the learners, which may have functioned as a confound. In addition, it also made it possible to observe readers' multi-stages of processing of the text during their reading, and the analysis of various types of time course measures and the movement patterns provided invaluable information in interpreting the results.

The present study also sheds more lights on the role of individual differences in working memory capacity. In order to gain more accurate working memory-spans of the participants, this study implemented two different types of automated working memory tests that have been widely used in a number of different fields of study, and endeavored to administer the test as accurately as possible. The results of the working memory capacity on the adult L2 brought some interesting findings as discussed above. However, further research is obviously needed to verify the findings.

6.1. Limitations and future research

One limitation of this study is that L2 groups included participants from two L1s, which was done mainly for a practical reason, as it was challenging to recruit enough number of early proficient adult L2 learners from one L1 background. Although the two languages share the same linguistic feature that was the main focus of the study, they differ from one another in many other respects (e.g., word order), some of which could have resulted in different processing patterns between the two L2s, thereby functioning as a confound in the group results.

Another limitation of the present study is that English proficiency and length of residence (or L2 exposure) were not matched between the early and adult ESL group, which consequently made it difficult to examine the role of different ages of immersion, as both L2 proficiency and degrees of L2 exposure have been found to play role in development of L2 grammar and L2 processing. It would be beneficial for future research to control for these factors in order to better examine the age-related effects in L2 processing.

As addressed earlier in the introduction, language processing involves a series of multiple complex linguistic analyses. The present study focuses only on one aspect of those analyses,

syntactic processing. However, given the fact that all these analyses occur more or less concurrently, it is conceivable that one type of linguistic analysis has an influence on another type of linguistic analysis. In this regard, some recent L2 processing studies report that L2 learners' lexical processing could affect their subsequent syntactic processing in one way or another, depending on lexical frequency of the items used in the test (e.g., Hopp, 2016), or cognate status (e.g., Miller, 2014). Thus, it will be more beneficial for future research to take these findings into consideration and further explore how these two areas of processing interact with each other.

APPENDICES

Appendix A. Language background questionnaire

1. Background Questionnaire for L2 learners

BACKGROUND QUESTIONNAIRE

(Participant ID: WM order: Task Type:)

※Please answer the following questions. If you have any questions that you would prefer not to answer, you can just leave them blank. All information you provide will be used only for the research purpose and all data will be kept confidential for your privacy and no information you provide will be directly related with any of your personal information.

A. Are you right-handed or left-handed? ☐ LEFT ☐ RIGHT

B. Are you wearing contact lenses or glasses for the experiment?
☐ YES (Circle one: Contact lenses Glasses) ☐ NO

1. Gender: ☐ Male ☐ Female

2. Age: _____ years old.

3. Education and/or Current Academic Status

☐ Freshman ☐ Sophomore ☐ Junior ☐ Senior ☐ BA (graduated)
☐ MA student ☐ MA ☐ Ph.D. student ☐ Ph.D. ☐ Others _____

4. What is your field of study? _____

5. What is your first/native language? _____

6. In what country and/or language environment did you have your primary (elementary) and secondary education?

I had my primary (elementary) education in _____, and classes were taught in _____ language.

I had my secondary education in _____, and classes were taught in _____ language.

7. HOW OLD WERE YOU...

when you first began acquiring/learning English? I was _____ years old.

7.1.) What was the educational setting for your English learning? Please mark all that apply.

- ☐ English classes at school (Your grade _____)
- ☐ Private tutoring
- ☐ English institute
- ☐ Others (Please explain what the setting was _____)

when you first came/moved to the United States?

- ☐ I was _____ years old.
- ☐ I was born in the U.S. (if applicable)

8. How long have you been living/studying in the U.S. so far? For ___ years ___ months

9. Do you have any other experiences of living in other English-speaking environments/countries PRIOR TO your current residence in the U.S.? Please exclude your travel experiences unless they are more than a year.

- ☐ No
- ☐ Yes (Please provide more details about those living abroad experiences.

For example, I was 2 years old when I moved to Australia and lived there for 3.5 years.

I was _____ years old when I moved to _____(WHERE) and lived there for _____years.

I was _____ years old when I moved to _____(WHERE) and lived there for _____years.

I was _____ years old when I moved to _____(WHERE) and lived there for _____years.

10. Please list all the languages you know, including your mother language and English, from the most to the least proficient order.

(1) _____ [Most proficient]

(2) _____

(3) _____

(4) _____ [Least proficient]

11. Among the languages you listed above.....

which language do you feel most comfortable with for verbal communication? _____

which language do you feel most comfortable with for reading? _____

which language do you feel most comfortable with for writing? _____

12. On a scale from ZERO (Not proficient at all) and TEN (near native-like), how would you rate your English proficiency in each of the following language skills? Please select one and V-check in the appropriate box in the table below.

	0 Not at all	1	2	3	4	5	6	7	8	9	10 Native -like
Listening											
Speaking											
Reading											
Writing											
Grammar											
Overall											

13. Please skip this question, if you have come to the U.S. after the age of 16.

On a scale from ZERO (Not proficient at all) and TEN (native-like), how would you rate your Chinese (For Chinese speakers) or Korean (For Korean speakers) proficiency in each of the following language skills? Please select one and V-check in the appropriate box in the table below.

	0 Not at all	1	2	3	4	5	6	7	8	9	10 Native -like
Listening											
Speaking											
Reading											
Writing											
Grammar											
Overall											

14. Have you taken any form of the standardized English proficiency tests, such as TOEFL, TOEIC, MTELP, or IELTS?

☐ No ☐ Yes

If your answer is YES, please provide what the test was and what the score was for each test. You do not need to answer, if you would like. The scores you provide will be used for the research purpose only.

TOEFL (※ Please circle the test format taken: iBT , CBT , PBT) Your score _____

TOEIC MTELP IELTS OTHERS (Test name_____) Your score _____

15. Is there any other information you can provide about your language background, or any comments? If so, please include it here:

2. Background Questionnaire for native speakers

BACKGROUND QUESTIONNAIRE

(Participant ID: WM order: Task Type:)

※Please answer the following questions. If you have any questions that you would prefer not to answer, you can just leave them blank. All information you provide will be used only for the research purpose and all data will be kept confidential for your privacy and no information you provide will be directly related with any of your personal information.

A. Are you right-handed or left-handed? LEFT RIGHT

B. Are you wearing glasses for the Eye-tracking experiment? YES NO

C. Are you wearing contact lenses for the Eye-tracking experiment? YES NO

1. Gender: ☐ Male ☐ Female

2. Age: _____ years old.

3. Education and Current Academic Status

- ☐ Freshman ☐ Sophomore ☐ Junior ☐ Senior ☐ BA (graduated)
☐ MA student ☐ MA ☐ Ph.D. student ☐ Ph.D.

4. What is your field of study? _____

5. Is there any other information you can provide about your language background, or any comments? If so, please include it here:

Appendix B. List of test items in the LexTALE English proficiency measure

[Words in English, $n = 40$]

ablaze	allied	bewitch	breeding	carbohydrate	celestial
censorship	cleanliness	cylinder	dispatch	eloquence	festivity
flaw	fluid	fray	hasty	hurricane	ingenious
lengthy	listless	lofty	majestic	moonlit	muddy
nourishment	plaintively	rascal	recipient	savoury	scholar
scornful	screech	shin	slain	stoutly	turmoil
turtle	unkempt	upkeep	wrought		

[Nonwords in English, $n = 20$]

abergy	alberation	crumper	destription	exprate	fellick
interfate	kermshaw	kilp	magrity	mensible	plaudate
proom	pudour	pulsh	purrage	rebondicate	skave
spaunch	quirty				

Appendix C. Materials for the eye-tracking experiment

- [non-island, plausible]
1a. The song that the guitarist wrote so passionately for was loved by the audience.
- [non-island, implausible]
1b. The band that the guitarist wrote so passionately for was loved by the audience.
- [island, plausible]
1c. The song that the guitarist who wrote so passionately recommended was loved by the audience.
- [island, implausible]
1d. The band that the guitarist who wrote so passionately recommended was loved by the audience.
- 2a. The diary that the historian read very thoroughly about was found near the castle.
2b. The sword that the historian read very thoroughly about was found near the castle.
2c. The diary that the historian who read very thoroughly studied was found near the castle.
2d. The sword that the historian who read very thoroughly studied was found near the castle.
- 3a. The fish that the chef cooked very uniquely with was introduced in the magazine.
3b. The lady that the chef cooked very uniquely with was introduced in the magazine.
3c. The fish that the chef who cooked very uniquely liked was introduced in the magazine.
3d. The lady that the chef who cooked very uniquely liked was introduced in the magazine.
- 4a. The captain that the spy killed so fiercely for was exposed to the enemy.
4b. The mission that the spy killed so fiercely for was exposed to the enemy.
4c. The captain that the spy who killed so fiercely assisted was exposed to the enemy.
4d. The mission that the spy who killed so fiercely assisted was exposed to the enemy.
- 5a. The actor that the designer dressed very elegantly for was praised by the critics.
5b. The opera that the designer dressed very elegantly for was praised by the critics.
5c. The actor that the designer who dressed very elegantly saw was praised by the critics.
5d. The opera that the designer who dressed very elegantly saw was praised by the critics.
- 6a. The nurse that the doctor texted very urgently about was isolated for further tests.
6b. The virus that the doctor texted very urgently about was isolated for further tests.
6c. The nurse that the doctor who texted very urgently examined was isolated for further tests.
6d. The virus that the doctor who texted very urgently examined was isolated for further tests.
- 7a. The article that the intern wrote very critically about was reviewed by the editor.
7b. The picture that the intern wrote very critically about was reviewed by the editor.
7c. The article that the intern who wrote very critically sent was reviewed by the editor.
7d. The picture that the intern who wrote very critically sent was reviewed by the editor.
- 8a. The suspect that the detective questioned very intensively about was sent to the CIA.
8b. The cocaine that the detective questioned very intensively about was sent to the CIA.
8c. The suspect that the detective who questioned very intensively found was sent to the CIA.
8d. The cocaine that the detective who questioned very intensively found was sent to the CIA.

- 9a. The mailman that the lady asked very angrily about was stuck in the warehouse.
 9b. The package that the lady asked very angrily about was stuck in the warehouse.
 9c. The mailman that the lady who asked very angrily expected was stuck in the warehouse.
 9d. The package that the lady who asked very angrily expected was stuck in the warehouse.
- 10a. The woman that the suspect texted very frequently about was searched by the police.
 10b. The house that the suspect texted very frequently about was searched by the police.
 10c. The woman that the suspect who texted very frequently mentioned was searched by the police.
 10d. The house that the suspect who texted very frequently mentioned was searched by the police.
- 11a. The dinner that the cook prepared very adeptly for was served with red wine.
 11b. The client that the cook prepared very adeptly for was served with red wine.
 11c. The dinner that the cook who prepared very adeptly hosted was served with red wine.
 11d. The client that the cook who prepared very adeptly hosted was served with red wine.
- 12a. The concert that the singer performed very actively for was sponsored by the city.
 12b. The pianist that the singer performed very actively for was sponsored by the city.
 12c. The concert that the singer who performed very actively helped was sponsored by the city.
 12d. The pianist that the singer who performed very actively helped was sponsored by the city.
- 13a. The book that the journalist wrote fairly regularly about was named for the explorer.
 13b. The city that the journalist wrote fairly regularly about was named for the explorer.
 13c. The book that the journalist who wrote fairly regularly mentioned was named for the explorer.
 13d. The city that the journalist who wrote fairly regularly mentioned was named for the explorer.
- 14a. The lectures that the professor prepared very hard for were evaluated by the program.
 14b. The students that the professor prepared very hard for were evaluated by the program.
 14c. The lectures that the professor who prepared very hard taught were evaluated by the program.
 14d. The students that the professor who prepared very hard taught were evaluated by the program.
- 15a. The resort that the housekeeper cleaned very diligently for was charged with tax evasion.
 15b. The lawyer that the housekeeper cleaned very diligently for was charged with tax evasion.
 15c. The resort that the housekeeper who cleaned very diligently sued was charged with tax evasion.
 15d. The lawyer that the housekeeper who cleaned very diligently sued was charged with tax evasion.
- 16a. The scenario that the novelist wrote very frequently about was selected for the filming.
 16b. The mountain that the novelist wrote very frequently about was selected for the filming.
 16c. The scenario that the novelist who wrote very frequently liked was selected for the filming.
 16d. The mountain that the novelist who wrote very frequently liked was selected for the filming.
- 17a. The school that the architect built very dedicatedly for was headlined in the news.
 17b. The artist that the architect built very dedicatedly for was headlined in the news.
 17c. The school that the architect who built very dedicatedly supported was headlined in the news.
 17d. The artist that the architect who built very dedicatedly supported was headlined in the news.
- 18a. The golfer that the trainer advised very thoroughly about was taken to the clinic.
 18b. The monkey that the trainer advised very thoroughly about was taken to the clinic.
 18c. The golfer that the trainer who advised very thoroughly trained was taken to the clinic.
 18d. The monkey that the trainer who advised very thoroughly trained was taken to the clinic.

- 19a. The crocodile that the rangers hunted very eagerly for was filmed for the documentary.
 19b. The zoologist that the rangers hunted very eagerly for was filmed for the documentary.
 19c. The crocodile that the rangers who hunted very eagerly liked was filmed for the documentary.
 19d. The zoologist that the rangers who hunted very eagerly liked was filmed for the documentary.
- 20a. The witness that the lawyer called so hurriedly about was reviewed by the judges.
 20b. The verdict that the lawyer called so hurriedly about was reviewed by the judges.
 20c. The witness that the lawyer who called so hurriedly questioned was reviewed by the judges.
 20d. The verdict that the lawyer who called so hurriedly questioned was reviewed by the judges.
- 21a. The reporter that the senator phoned very recently about was investigated by the police.
 21b. The accident that the senator phoned very recently about was investigated by the police.
 21c. The reporter that the senator who phoned very recently blamed was investigated by the police.
 21d. The accident that the senator who phoned very recently blamed was investigated by the police.
- 22a. The bomb that the soldier threw quite forcefully toward was covered with thick moss.
 22b. The wall that the soldier threw quite forcefully toward was covered with thick moss.
 22c. The bomb that the soldier who threw quite forcefully destroyed was covered with thick moss.
 22d. The wall that the soldier who threw quite forcefully destroyed was covered with thick moss.
- 23a. The proposal that the senator prepared so ambitiously for was tackled by the panel.
 23b. The governor that the senator prepared so ambitiously for was tackled by the panel.
 23c. The proposal that the senator who prepared so ambitiously supported was tackled by the panel.
 23d. The governor that the senator who prepared so ambitiously supported was tackled by the panel.
- 24a. The hotel that the architect designed so intensely for was targeted by the terrorist.
 24b. The queen that the architect designed so intensely for was targeted by the terrorist.
 24c. The hotel that the architect who designed so intensely visited was targeted by the terrorist.
 24d. The queen that the architect who designed so intensely visited was targeted by the terrorist.
- 25a. The team that the athlete trained so intensively for was supported by the fans.
 25b. The game that the athlete trained so intensively for was supported by the fans.
 25c. The team that the athlete who trained so intensively led was supported by the fans.
 25d. The game that the athlete who trained so intensively led was supported by the fans.
- 26a. The document that the lawyer read very thoroughly about was investigated by the FBI.
 26b. The accident that the lawyer read very thoroughly about was investigated by the FBI.
 26c. The document that the lawyer who read very thoroughly reported was investigated by the FBI.
 26d. The accident that the lawyer who read very thoroughly reported was investigated by the FBI.
- 27a. The engineer that the CEO paid quite generously for was disliked by the investors.
 27b. The proposal that the CEO paid quite generously for was disliked by the investors.
 27c. The engineer that the CEO who paid quite generously selected was disliked by the investors.
 27d. The proposal that the CEO who paid quite generously selected was disliked by the investors.
- 28a. The musical that the musician composed so devotedly for was awarded the grand prize.
 28b. The pianist that the musician composed so devotedly for was awarded the grand prize.
 28c. The musical that the musician who composed so devotedly loved was awarded the grand prize.
 28d. The pianist that the musician who composed so devotedly loved was awarded the grand prize.

REFERENCES

REFERENCES

- Adger, David (2003). *Core syntax: a minimalist approach*. Oxford: Oxford University Press.
- Aldwayan, S., Fiorentino, R., & Gabriele, A. (2010). Evidence of syntactic constraints in the processing of wh-movement: A study of Najdi Arabic learners of English. In VanPatten, B., & Jegerski, J. (Eds.), *Research in second language processing and parsing* (pp. 65-86). Philadelphia, PA: John Benjamins Publishing Company.
- Ariji, Kenji, Akira Omaki & Nano Tatsuta. (2003). Working memory restricts the use of semantic information in ambiguity resolution. In Peter Slezak (ed.) *Proceedings of the 4th International Conference on Cognitive Science* (pp. 19-25). Sydney, Australia: University of New South Wales.
- Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, 36, 189–208.
- Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *The Psychological Review*, 105, 158-173.
- Barrouillet, P., & Lepine, R. (2005). Working memory and children's use of retrieval to solve addition problems. *Journal of Experimental Child Psychology*, 91, 183-204.
- Belikova, A. & White, L. (2009). Evidence for the Fundamental Difference Hypothesis or not? Island constraints revisited. *Studies in Second Language Acquisition*, 31, 199-223.
- Bialystok, E. (1997). The structure of age: in search of barriers to second language acquisition. *Second Language Research*, 13, 116-137.
- Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language*, 68, 706-755.
- Birdsong, D. (1999). Introduction: Whys and Why nots of the critical period hypothesis for second language acquisition. In Gass, S., & Schachter, J. (Eds.), *Second language acquisition and the critical period hypothesis* (pp. 1-22). Mahwah, NJ: Lawrence Erlbaum Associates.
- Birdsong, D. (2005). Nativelikeness and non-nativelikeness in L2A research. *International Review of Applied Linguistics (IRAL)*, 43, 319-328.
- Birdsong, D. (2014). The critical period hypothesis for second language acquisition: Tailoring the coat of many colors. In Pawlak, M., & Aronin, L. (Eds.), *Essential topics in applied linguistics and multilingualism. Studies in honor of David Singleton* (pp. 43-50). Heidelberg: Springer.
- Bley-Vroman, R., (1990). The logical problem of second language learning. *Linguistic Analysis*, 20, 3–49.

- Bley-Vroman, R., (2009). The evolving context of the fundamental difference hypothesis, *Studies in Second Language Acquisition*, 31, 175-198.
- Bley-Vroman, R., Felix. S., & G. Ioup. (1988). The accessibility of Universal Grammar in adult language learning. *Second Language Research*, 4, 1-32.
- Carreiras, M., & Clifton, C. (1999). Another word on parsing relative clauses: Eyetracking evidence from Spanish and English, *Memory and Cognition*, 27(5), 826-833.
- Chomsky, N. (1973). Conditions on transformation. In S. Anderson & P. Kiparsky (Eds.), *Festschrift for Morris Halle* (pp. 232-286). New York, NY: Holt, Rinehart & Winston.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.
- Clahsen, H. (2007). Psycholinguistic perspectives on grammatical representations. In Featherston, S., & S. Wolfgang (Eds.), *Roots: Linguistics in its Search of Its Evidence Base*. (pp.97-132). Berlin: Mouton de Gruyter Publishing Company.
- Clahsen, H., & Felser, C. (2006a). Grammatical processing in language learners. *Applied Psycholinguistics*, 27, 3- 42.
- Clahsen, H., & Felser, C. (2006b). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, 27, 107- 126.
- Clahsen, H., & Felser, C. (2006c). How native-like is non-native language processing? *Trends in Cognitive Sciences*, 10, 564- 570.
- Clahsen, H., & Muysken, P. (1996). How adult second language learning differs from child first language development. *Behavioral and Brain Sciences*, 19, 721-723.
- Clifton, C., Staub, A., Rayner K. (2007). Eye movements in reading words and sentences. In van Gompel, R.P.G., Fischer, M.H., Murray, W.S., Hill R.L. (Eds.), *Eye movements: A window on mind and brain* (pp. 341–371). New York: Elsevier.
- Conway, R. A., M. J. Kane, M. Bunting, D. Hambrick, O. Wilhelm & R. Engle (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12, 769–786.
- Cunnings, I., Batterham, C., Felser, C., Clahsen, H. (2010). Constraints on L2 learners' processing of wh-dependencies: Evidence from eye-movements. In VanPatten, B., & Jegerski, J. (Eds.), *Research in second language processing and parsing* (pp. 87-112). Philadelphia, PA: John Benjamins Publishing Company.
- Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic difference in parsing: restrictions on the late-closure strategy in Spanish. *Cognition*, 30, 73-105.

- Cuetos, F., & Mitchell, D. C., & Corely, M. (1996). Parsing in different languages. In M. Carreiras, J. Garcia-Albea, & N. Sebastien-Galles (Eds.). *Language Processing in Spanish* (pp. 145-187). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450-466.
- Declerck, M., Lemhöfer, K., & Grainger, J. (in press). Bilingual language interference initiates error detection: Evidence from language intrusions. *Bilingualism: Language and Cognition*. <http://dx.doi.org/10.1017/S1366728916000845>.
- Dekeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22, 49-533.
- Dekeyser, R. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics*, 31, 413-438.
- Dekydtspotter, L., & Miller, (2009). Probing for intermediate traces in the processing of long-distance wh-dependencies in English as a second language. In Bowles, M., Ionin, T., Montrul, S., & Tremblay, A. (Eds.) *Proceedings of the 10th Generative Approaches to Second Language Acquisition (GASLA, 2009)* (pp. 113-124). Somerville, MA: Cascadilla Proceedings Project.
- Dekydtspotter, L., & Miller, (2013). Inhibitive and facilitative priming induced by traces in the processing of wh-dependencies in a second language. *Second Language Research*, 29, 345-372.
- Dekydtspotter, L., & Renaud, C. (2014). On second language processing and grammatical development: The parser in second language acquisition. *Linguistic Approaches to Bilingualism*, 4, 131-165.
- Dekydtspotter, L., Schwartz, & Sprouse, R. (2006). The comparative fallacy in L2 processing research. *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference*, 33-40.
- Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish-English bilinguals. *Bilingualism: Language and Cognition*, 10, 101-116.
- Dussias, P.E., & Piñar, P. (2010). Effects of reading span and plausibility in the reanalysis of wh-gaps by Chinese-English second language speakers. *Second Language Research*, 26, 443-472.
- Ellis, R. (1991). Grammaticality judgments and second language acquisition. *Studies in Second Language Acquisition*, 13, 161-186.
- Epstein, S., Flynn, S., & Martohardjono, G. (1996). Second language acquisition: Theoretical and experimental issues in contemporary research. *Behavioral and Brain Sciences*, 19, 677-714.

- Featherston, S. (2001). *Empty categories in sentence processing*. Amsterdam, the Netherlands: John Benjamins Publishing Company.
- Felix, S., & Weigl, W. (1991). Universal grammar in the classroom: The effect of formal instruction on second language acquisition. *Second Language Research*, 7, 162-180.
- Felser, C., Roberts, L., & Marinis, T. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, 24, 453-489.
- Felser, C., Sato, M., & Bertenshaw, M. (2009). The on-line application of binding Principle A in English as a second language. *Bilingualism: Language and Cognition*, 12, 485-502.
- Felser, C., Cunnings, I., Batterham, C., & Harald Clahsen (2012). The timing of island effects in nonnative sentence processing. *Studies in Second Language Acquisition*, 34, 67-98.
- Fernandez, E. (1999). Processing strategies in second language acquisition: Some preliminary results. In E. Klein, & G. Martohardjono (Eds.), *The development of second language grammars: A generative approaches* (pp.217-239). Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Fender, M. (2003). English word recognition and word integration skills of native Arabic- and Japanese-speaking learners of English as a second language. *Applied Psycholinguistics*, 24, 289-315.
- Fender, M. (2008). L1 effects on the emergence of ESL sentence processing skills of Chinese and Korean ESL learners: A preliminary study. *Languages in Contrast*, 8, 47-73.
- Ferreira, F., Engelhardt, P., & Jones, M. (2009). Good enough language processing: A satisficing approach. In N. Taatgen, H. Rijn, J. Nerbonne, & L. Schomaker (Eds.), *Proceedings of the 31st Annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.
- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London, England: Sage.
- Fodor, J. D. (1998). Parsing to learn. *Journal of Psycholinguistic Research*, 22, 339-374.
- Frazier, L. (1998). Getting there (slowly). *Journal of Psycholinguistic Research*, 16, 123-146.
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14, 178-210.
- Frazier, L. (1987). Processing syntactic structures: Evidence from Dutch. *Natural Language and Linguistic Theory*, 5, 519-559.
- Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6, 291 - 325.

- Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, 4, 93-126.
- Frenck-Mestre, C. (1997). Examining second language reading: An on-line look. In A. Sorace, C. Heycock, & R. Shillcock (Eds.), *Proceedings of the GALA1997 Conference on Language Acquisition* (pp. 474-478). Edinburgh, UK: Human Communications Research Center.
- Frenck-Mestre, C. (2002). A on-line look at sentence processing in a second language. In R. Herrida and J. Altarriba (Eds.), *Bilingual Sentence Processing* (pp.217-236). North Holland.
- Frenck-Mestre, C. (2005). Eye-movement recording as a tool for studying syntactic processing in a second language: a review of methodologies and experimental findings. *Second Language Research*, 21(2), 175-198.
- Frenck-Mestre, C., & Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *Quarterly Journal of Experimental Psychology*, 50A, 119–148.
- Gass, S. M. (1994). The reliability of second-language grammaticality judgments. In Tarone, E., Gass, S. M., & Cohen, A. D. (Eds.), *Research methodology in second-language acquisition*. Hillsdale, NJ: L. Erlbaum Associates.
- Gass, S. M., & Lee, J. (2011). Working memory capacity, inhibitory control, and proficiency in a second language. In Schmid, M., & Lowie, W. (Eds.), *Modeling Bilingualism: From Structure to Chaos* (pp. 59-84). Amsterdam, The Netherlands: John Benjamins Publishing Company.
- Gass, S. M., & Selinker, L. (2008). *Second Language Acquisition: An introductory course (3rd edition)*. New York, NY: Routledge.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1-76.
- Gibson, E., & Hickok, G. (1993). Sentence processing with empty categories. *Language and Cognitive Processes*. 8, 147-171.
- Gibson, E., Pearlmutter, N., Canseco-Gonzalez, E., & Hickok, G. (1996). Recency preference in the human sentence processing mechanism. *Cognition*, 56, 23-59.
- Gibson, E. and Warren, T. (2004): Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax*, 7, 55–78.
- Ha, J. (2005). Age-related effects on syntactic ambiguity resolution in first and second languages: Evidence from Korean-English Bilinguals. In Laurent Dekydtspotter et al

- (Eds), *Proceedings of the 7th Generative Approaches to Second Language Acquisition* (pp.111-123). Somerville, MA: Cascadilla Proceedings Project.
- Hakuta, K., Bialystok, E., & Wiley, E. (2003). *Psychological Science*, 14, 31-38.
- Harrington, M., & Sawyer, M. (1992). Working memory capacity and L2 reading skill. *Studies in Second Language Acquisition*, 14, 25-38.
- Havic, E., Roberts, L., van Hout, R., Schreuder, R., & Haverkort, M. (2009). Processing subject-object ambiguities in the L2: A self-paced reading study with German L2 learners of Dutch. *Language Learning*, 59, 73-112.
- Hawkins, R. (2001). *Second language syntax: A generative introduction*. Oxford: Blackwell Publishers.
- Hawkins, R., & Chan, C. (1997). The partial availability of Universal Grammar in second language acquisition: The 'failed functional features hypothesis.' *Second Language Research*, 13, 187-226.
- Hawkins, R., & Hattori, H. (2006). Interpretation of English multiple wh-questions by Japanese speakers: a missing uninterpretable feature account. *Second Language Research*, 22, 269-301.
- Herschensohn, J. (2000). *The second time around minimalism and L2 acquisition*. Philadelphia, PA: John Benjamins Publishing Company. (AGE EFFECT CHECK)
- Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, 22, 369-397.
- Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, 120, 901-931.
- Hopp, H. (2014). Working memory effects in the L2 processing of ambiguous relative clauses. *Language Acquisition*, 21, 250-278.
- Hopp, H. (2016). The timing of lexical and syntactic processes in second language sentence comprehension. *Applied Psycholinguistics*, 37, 1253-1280.
- Hummel, K. (2009). Aptitude, phonological memory, and second language proficiency in nonnative adult learners. *Applied Psycholinguistics*, 30, 225-249.
- Jackson, C. N. (2008). Proficiency level and the interaction of lexical and morphosyntactic information during L2 sentence processing. *Language Learning*, 58, 875-909.
- Jackson, C. N., & Dussias, P. E. (2007). Cross-linguistic differences and their impact on L2 sentence processing. *Bilingualism: Language and Cognition*, 12, 65-82.

- Jegerski, J., VanPatten, B., & Keating, G. (2011). Cross-linguistic variation and the acquisition of pronominal reference in L2 Spanish. *Second Language Research*, 27, 481-507.
- Jiang, N. Selective integration of linguistic knowledge in adult second language learning. *Language Learning*, 57, 1-33.
- Jiang, N., Novokshanova, E., Masuda, K., & Wang, X. (2011). Morphological congruency and the acquisition of L2 morphemes. *Language Learning*, 61, 940-967.
- Johnson, J., & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 69-99.
- Johnson, J., & Newport, E. (1991). Critical period effects on universal properties of language: The status of subadjacency in the acquisition of second language. *Cognition*, 39, 215-258.
- Johnson, J. (1992). Critical period effects in second language acquisition: The effect of written versus auditory materials on the assessment of grammatical competence. *Language Learning*, 42, 217-248.
- Juffs, A. (1998). Some effects of first language argument structure and morphosyntax on second language sentence processing. *Second Language Research*, 14, 406-242.
- Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society*, 102, 199-225.
- Juffs, A. (2005). The influence of first language on the processing of wh-movement in English as a second language. *Second Language Research*, 21, 121-151.
- Juffs, A., & Harrington, M. (1995). Parsing effects in second language sentence processing: subject and object asymmetries in wh-extraction. *Studies in Second Language Acquisition*, 17, 483-516.
- Juffs, A., & Harrington, M. (1996). Garden path sentences and error data in second language sentence processing. *Language Learning*, 46, 283-326.
- Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, 44, 137-166.
- Juffs, A., & Rodriguez, G. (2015). *Second language sentence processing*. New York, NY: Routledge.
- Just, M. A., Carpenter, P., & Woolley, J.D. (1982). Paradigms and processes and in reading comprehension. *Journal of Experimental Psychology: General*, 3, 228-238.
- Kann, E., Ballantyne, J., & Wijnen, F. (2015). Effects of reading speed on second language sentence processing. *Applied Psycholinguistics*, 36, 799-830.

- Keating, G. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning*, 59, 503-535.
- Kim, E., Baek, S., Tremblay, A. (2015). The role of island constraints in second language sentence processing. *Language Acquisition*, 22, 384-416.
- Kim, J., Christianson, K. (2013). Sentence complexity and working memory effects in ambiguity resolution. *Journal of Psycholinguistic Research*, 42, 393-411.
- Lardiere, D. (2008). Feature assembly in second language acquisition. In: Liceras J., Zobl, H. and Goodluck H. (Eds.), *The role of formal features in second language acquisition* (pp. 106-140). New York, NY: Lawrence Erlbaum Associates.
- Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York: Routledge.
- Lee, M-W. (2004). Another look at the role of empty categories in sentence processing (and grammar). *Journal of Psycholinguistic Research*, 33, 51-73.
- Leeser, M. (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: Topic familiarity and working memory. *Language Learning*, 57, 229-270.
- Lemhöfer K and Broersma M (2012) Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, 44, 325-343.
- Lenneberg, E., (1967). *Biological foundations of language*. New York, NY: Wiley.
- Lewis, R. (1998). Reanalysis and limited repairing parsing: leaping off the garden path. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 247-228). Dordrecht: Kluwer Academic Publishers.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 1j01, 676-703.
- Marinis, T. (2003). Psycholinguistic techniques in second language acquisition research. *Second Language Research*, 19, 144-161.
- Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language processing. *Studies in Second Language Acquisition*, 27, 53-78.
- Marquardt, D. (1980). A critique of some ridge regression methods: Comment. *Journal of the American Statistical Association*, 75, 67-91.
- Nakano, Y., Felser, C., & Clahsen, H. (2002). Antecedent priming at trace positions in Japanese long-distance scrambling. *Journal of Psycholinguistic Research*, 31, 531-571.

- Meara, P. M. (1996). *English Vocabulary Tests: 10 k*. Swansea. UK: Center for Applied Language Studies.
- Miller, A. K. (2014). Accessing and maintaining referents in L2 processing of wh-dependencies. *Linguistic Approaches to Bilingualism*, 4, 167-191.
- Miller, A. K. (2015). Intermediate traces and intermediate learners: Evidence for the use of intermediate structure during sentence processing in second language French. *Studies in Second Language Acquisition*, 37, 487-516.
- Mirdamadi, F., & De Jong, N. (2015). The effect of syntactic complexity on fluency: Comparing actives and passives in L1 and L2 speech. *Second Language Research*, 31, 105-116.
- Ojima, S. et al. (2005) An ERP study of second language learning after childhood: effects of proficiency. *Journal of Cognitive Neuroscience*, 17, 1212-1228
- Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constraints in second-language sentence processing. *Studies in Second Language Acquisition*, 33, 563-588.
- O'Rourke, P. (2013). The interaction of different working memory mechanisms and sentence processing: A study of the P600. In Knauff, M., Sebanz, N., Pauen, M., & Wachsuth, I. (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 1097-1102). Austin, TX: Cognitive Science Society.
- Osaka, M., & Osaka, N. (1992). Language independent working memory as measured by Japanese and English reading span tests. *Bulletin of the Psychonomic Society*, 30, 287-289.
- Oswald, F., McAbee, S., Redick, T., & Hambrick, D., (2015). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods*, 1, 1343-1355.
- Papadopolou, D. (2006). *Cross-linguistic variation in sentence processing*. Dordrecht: Springer.
- Penfield, W., & Roberts, L. (1959). *Speech and brain mechanisms*. New York, NY: Atheneum
- Pickering, M., & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes*, 6, 229-259.
- Pickering, M., & Traxler, M. (1998). Plausibility and recovery from garden path: An eye-tracking study. *Journal of Experimental Psychology*, 24, 940-961.
- Pliatsikas, C., & Marinis, T. (2013). Processing empty categories in a second language: When naturalistic exposure fills the (intermediate) gap. *Bilingualism: Language and Cognition*, 16, 167-182.
- Pollard, C., & Sag, I. (1994). *Head-driven phrase structure grammar*. Chicago: University of Chicago Press and Stanford: CSLI Publications.

- Pritchett, B. (1988). *Garden path phenomena and the grammatical basis of language processing*. *Language*, 64, 539-576.
- Pritchett, B. (1992). *Grammatical competence and parsing performance*. Chicago: University of Chicago Press.
- Rahman, S. S. (2010). Acquisition of wh-movement in L2 learning: A cross-linguistic analysis. *The Dhaka University Journal of Linguistics*, 2, 185-199.
- Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.
- Redick, T., Broadway, J., Meier, M., Kuriakose, P., Unsworth, N., Kane, M., & Engle, R. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, 28, 164-171.
- Roberts, L., Marinis, T., Felser, C. and Clahsen, H. (2007). Antecedent priming at gap positions in children's sentence processing. *Journal of Psycholinguistic Research*, 36, 175-188.
- Roberts, L., & Siyanova-Chanturia (2013). Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition*, 35, 213-235.
- Robinson, P. (2002). Effects of individual differences in intelligence, aptitude and working memory on incidental SLA. In Robinson, P. (Ed.), *Individual differences and instructed language learning*. Philadelphia: John Benjamins.
- Rothman, J. (2008). Why all counter-evidence to the critical period hypothesis in second language acquisition is not equal or problematic. *Language and Linguistic Compass*, 2/6, 1063-1088.
- Ross, J. R. (1967). *Constraints on variables in syntax*. Unpublished Ph.D. thesis. MIT.
- Sabourin, L. and Haverkort, M. (2003). Neural substrates of representation and processing of a second language. In van Hout, R., Hulk, A., Kuiken, F., & Towell, R. (Eds.), *The Lexicon-Syntax Interface in Second Language Acquisition* (pp. 175-195). Amsterdam: John Benjamins.
- Sagarra, N., & Ellis, N. C. (2013). From seeing adverbs to seeing morphology. Language experience and adult acquisition of L2 tense. *Studies in Second Language Acquisition*, 35, 261-290.
- Sagarra, N. & J. Herschensohn (2010). The role of proficiency and working memory in gender and number agreement marking in processing in L1 and L2 Spanish. *Lingua*, 120, 2022-2039.
- Schacter, J. (1989). Testing a proposed universal. In S. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 73-88). Cambridge, UK: Cambridge University Press.

- Schachter, J. (1990). On the issue of completeness in second language acquisition. *Second Language Research*, 6, 93-124.
- Schachter, J., & Yip, V. (1990). Grammaticality Judgments: Why does anyone object to subject extraction? *Studies in Second Language Acquisition*, 12, 379-392.
- Schwartz, B., & Sprouse, R. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, 12, 40-72.
- Segalowitz, N., & Segalowitz, S. J. (1993). Skilled performance practice and differentiation of speedup from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, 13, 369-385.
- Singleton, D. (2005). The critical period hypothesis: A coat of many colours. *International Review of Applied Linguistics*, 43, 269-285.
- Slabakova, R. (2006). Is there a critical period for semantics? *Second Language Research*, 22, 302-338.
- Staub, A. & Clifton, C. (2006). Syntactic prediction in language comprehension: Evidence from either... or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 425-436.
- Stowe, L. (1986). Evidence for on-line gap location. *Language and Cognitive Processes*, 1, 227-245.
- Sturt, P., Pickering, M. J., & Crocker, M. W. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language*, 40, 136-150.
- Swinney, D. (1979). Lexical access during sentence comprehension: (Re) consideration of context effect. *Journal of Verbal Learning and Verbal Behavior*, 18, 645-659.
- Swinney, D., Ford, M., Frauenfelder, U., & Bresnan, J. (1988). On the temporal course of gap-filling and antecedent assignment during sentence comprehension. In Grosz, B., Kaplan, R., Macken, M., & Sag, I. (Eds.), *Language structure and processing*. Stanford, CA: CSLI.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632-1634.
- Tremblay, A. (2005). Theoretical and Methodological perspectives of the use of grammaticality judgment task in linguistic theory. *Second Language Studies*, 24, 129-167.
- Traxler, M., & Pickering, M. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, 35, 454-475.

- Trenkic, D., Mirkovic, J., & Altmann, G. (2014). Real-time grammar processing by native and non-native speakers: Constructions unique to the second language. *Bilingualism: Language and Cognition*, 17, 237-257.
- Ullman, M. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, 92, 231-270.
- Van Gompel, R. P. G., & Pickering, M. (2007). In Gaskell, M., & Altmann, G. (Eds.), *The Oxford handbook of psycholinguistics*. (pp. 289-307). Oxford: Oxford University Press.
- VanPatten, B., & Jegerski, J. (2010). Second language processing and Parsing: The issues. In VanPatten, B., & Jegerski, J. (Eds.), *Research in second language processing and parsing* (pp. 3-26). Philadelphia, PA: John Benjamins Publishing Company.
- Vasishth, S., Drenhaus, H. (2011). Locality in German. *Dialogue and Discourse*, 2, 59-82.
- Wagers, M., & Phillips, C. (2014). Going the distance: Memory and control processes in active dependency construction. *The Quarterly Journal of Experimental Psychology*, 67, 1274-1304.
- Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience*, 8, 231-256.
- White, L. (1992). Subjacency violations and empty categories in L2 acquisition. In H. Goodluck & M. Rochemont (Eds.), *Island constraints* (pp. 445-464). Dordrecht: Kluwer.
- White, L. (2003). *Second language acquisition and Universal Grammar*. Cambridge, England: Cambridge University Press.
- White, L., & Juffs, A. (1998). Constraints on wh-movement in two different contexts of nonnative language acquisition: Competence and processing. In Flynn, S., Martohardjono, G. and O'Neill, W. (Eds.), *The generative study of second language acquisition* (pp. 111-130). Hillsdale, NJ: Erlbaum.
- Williams, J. N. (2006). Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition*, 9, 71-88.
- Williams, J., Moebius, P., & Kim, C. (2001). Native and non-native processing of English wh-questions: Parsing strategies and plausibility Constraints. *Applied psycholinguistics*, 22, 509-540.
- Zawiszewski, A., Gutierrez, E., Fernandez, B., & Laka, I. (2011). Language distance and non-native syntactic processing: Evidence from event-related potentials. *Bilingualism: Language and Cognition*, 14, 401-411.

Zufferey, S., Mak, W., & Degand, L. (2015). Advanced learners' comprehension of discourse connectives: The role of L1 transfer across on-line and off-line tasks. *Second Language Research*, 31, 389-411.