SANDHI-VARIATION AND THE COMPREHENSION OF SPOKEN ENGLISH FOR JAPANESE LEARNERS

A Dissertation
Submitted
to the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

By
Brett Collins
May, 2018

Examining Committee Members

David Beglar, Advisory Chair, Teaching and Learning
Tomoko Nemoto, Teaching and Learning
Rob Waring, External Member, Notre Dame Seishin University
Kazuya Saito, External Member, Birkbeck, University of London

ABSTRACT

In this study I addressed three problems related to how *sandhi-variation*, the adjustments made by speakers to the speech stream, filters comprehension for second language listener processing. The first was the need to better understand proficiency problems encountered by L2 listeners as they decode the speech stream with the phonological features of sandhi-variation, elision and assimilation, by investigating the item difficulty hierarchy of the phenomena. The second was the scarcity of research on aural processing abilities of second language learners in relation to their understanding sandhi-variation in aural texts. The third concerns the lack of research investigating links between learners' backgrounds and their ability to handle listening texts, especially variations in the speech stream in target aural texts. The purpose of this study was threefold. My first purpose was to investigate the item difficulty hierarchy of sandhi-variation types that learners have in relation to L2 listening proficiency. My second purpose was to evaluate links between aural input containing elision and assimilation and second language aural processing, to provide insight into how learners deal with sandhi-variation as they process such input. My third purpose was to investigate through the use of interviews the aural input that participants have encountered prior to the interventions of this study, to help explain which types of aural input can facilitate intake.

Twenty-five first- and second-year Japanese university students participated in the current study. The participants completed a series of instruments, which included (a) a Test of English as a Foreign Language Paper-Based Test (TOEFL PBT), (b) a Listening Vocabulary Levels Test (LVLT), (c) a Modern Language Aptitude Test–Elementary (MLAT-E), (d) a Pre-Listening in English questionnaire, (e) an Elicited

Imitation Test (EIT), and (f) a Background and Length of Residency interview. The EIT was used as a sandhi-variation listening test with two component parts (i.e., elision and assimilation) and two sub-component parts (e.g., two different utterance rates), using elicited imitation. Finally, the participants were interviewed about their language backgrounds to gauge their understanding and feelings about English.

An empirical item hierarchy for elision and assimilation was investigated, along with the determinants of the hierarchy. Overall, the tendency was for items with elision and assimilation to be more difficult. Results also indicated that the two input rate variables combined with elision and assimilation affected the non-native participants' listening comprehension. Moreover, the strength of the relationship between two measures of the participants' language ability, proficiency and aptitude, and their comprehension of items with and without the phonological features of elision and assimilation, were investigated. The results confirmed a positive relationship between language aptitude as measured by the MLAT-E and the comprehension of the phonological features of elision and assimilation. Finally, the results indicated that there were no significant, positive correlations between *English language proficiency* scores and both the Pre-Listening Questionnaire, which measured the participants' feelings about second language listening, and the Background and Length of Residency Interview. More research needs to be conducted to determine how learners' backgrounds are related to listening comprehension in order to better prescribe aural input in second language listening classrooms.

ACKNOWLEDGMENTS

I would first like to acknowledge myself, or, at least, the person I was when I started the process. That kid was naive. I have learned to buck up better since then thanks to the work needed to finish this degree. Also, I have learned so much from the mountain of mistakes I made while writing this dissertation, and that in the mistakes lie lessons. Dr. David Beglar was there throughout my trials and errors, and he modeled academics for me. When I look back over the feedback that Dr. Beglar gave over the many years, I am in awe, and I am inspired to help others. Thank you, Dr. Beglar, for your tremendous support and advice. Your edits were always right.

Great thanks goes to the members of my committee. Dr. Rob Waring helped me through a panic at the proposal stage. He was instrumental in pointing me in new directions, and he gave good suggestions at the defense. Dr. Tomoko Nemoto gave many good comments at the defense, which helped me gain a different perspective on pronunciation and transcription. Dr. Kazuya Saito gave excellent advice to help me expand my perspective on the study, especially on language aptitude.

Other help came from readers, especially Dr. Phillip Clark. His many close readings of the manuscript and suggestions on style and structure made the study more presentable. Dr. Rick Derrah also read parts of the manuscript and made suggestions. Also, early on, Dr. Mark Sawyer helped me formulate ideas for this study. Thanks to Yuki Tokashiki and Masahiro Toma for translation work.

Thank you very much to the members of the Temple University Japan 2008 cohort, especially Phillip Clark, Rick Derrah, John Rylander, and Scott Shinall, for support from the early days on. Thanks to the White Bear crew, including Kevin Ballou. Thank you to Miles Davis, especially for *On The Corner*, *Big Fun*, and *Bitches Brew*. These albums got me through some rough days. Thanks to my close friends who are still around after all this time, and apologies to those friends who are not. Thanks and love to Jittima for her daily kindness and love.

Finally, the imperfections within the manuscript are absolutely my fault.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The Background of the Issue

A persistent challenge for learners of a new language is listening to fluent speech (Field, 2008a; Flowerdew & Miller, 2005; Kennedy & Blanchet, 2014; Wolvin, 2005). Simply, speech is as a connection of sounds created by a combination of movements made as the vocal folds, the tongue, the teeth, and the larynx form the vocal stream into meaningful noises (Jusczyk, 2000; Laver, 1994; Pinker, 2007). Words in the speech stream are woven together by phonetic rules, and one general rule for this weave is referred to as *sandhi-variation* (Ahn, 1987; Brooks, 1966; Brown & Kondo-Brown, 2006; Henrichsen, 1984). Historically, sandhi-variation is the phenomenal effect of the speech process on contextual segments as they combine to form utterances (Ahn, 1987; Henrichsen, 1984). Reasons given for the phenomena of sandhi-variation are varied, and include the influence of sounds on neighboring sounds (Celce-Murcia, Brinton, Goodwin, & Griner, 2010/2016) and the relaxation by speakers due to, for example, the abundance of contextual clues given alongside the spoken message (e.g., physical gestures). Changes, such as elision and assimilation, occur within and between contextual segments (Field, 2008b), which are made up of what are traditionally labeled *phonemes* (Jones, 1956). These changes are all aspects of sandhi-variation. Elision occurs, for example, when a medial vowel is deleted after a strong syllable; for instance, *favorite* [feɪvərɪt] is often elided as [feɪvrɪt]. Assimilation occurs, for example, when one sound behaves differently because of another sound. For instance, *in perpetuity* [ɪn pɜrpɪtjuɪti] can assimilate as [ɪm pɜrpɪtjuɪti]. In this case, the final nasal [n] in the word *in* adjusts to the following consonant.

Listeners must assign boundaries, or segments within the speech stream, based on their perception of the context of *the linguistic message*. The linguistic message here refers to the perceived meaning of the speaker's utterance, and the dynamics (e.g., syllables) in the physical message (Abercrombie, 1968; Vandergrift & Goh, 2012). These boundaries relate correctly or incorrectly to speakers' actual *signals*, the spoken connections of sounds speakers use to convey meaning (Tatham & Morton, 2006). A listener's perception of the context of the signal might differ from the actual context of the signal produced by the speaker for numerous reasons, for instance, misunderstanding a word within the speech stream.

Second language learners must negotiate the speech stream in order to develop accurate perceptions of phonological context, the intended meaning of the speaker. For example, if a speaker wants to say *excuse me?* (*x*) and then says *excuse me?* (*x*) to a listener, the speaker wants the listener to receive the message *excuse me?* (*x*) (i.e., not *y* or another contextual idea not intended, for instance, *accuse me?*). Moreover, second language learners often must negotiate the speech stream from the unknown and unique utterances that they might encounter in authentic language, such as in conversations with native speakers, so that they can gradually acquire fluency in comprehending aural input (Bradlow, Nygaard, & Pisoni, 1999; Cutler, 2015; Rost & Ross, 1991). Second language listeners must derive meaning from spoken input, yet they must first have enough anchored contexts, sections of the target where they can access semantic meaning in the second language, to be able to develop abstract notions of the boundaries before assigning meaning to the whole utterance (Field, 2003; Kennedy & Blanchet, 2014).

For second language learners, listening comprehension depends on their experience and knowledge of the second language (Cutler, 2000, 2015), much in the same way that for native speakers, listening comprehension depends on similar variables in their first language (Jarvella, 1971; Tomasello, 2003). Fundamentally, the more cultural literacy, experience, and knowledge that learners have of listening, vocabulary, locations, emotions, attitudes, prosody, and a host of other variables, the better chances they have to accurately perceive and comprehend aural input (Ellis, 2002). Listening comprehension breakdowns occur when the listeners' ability to parse the myriad intricacies of the string of sounds in target utterances is challenged. Comprehension of what is referred to as *the utterance plan* (Tatham & Morton, 2006), which can be roughly defined as the acoustic goal of the speaker, is rooted in how well listeners are able to process the audio stream they hear based on what they already know (Bundgaard-Nielsen, Best, & Tyler, 2011; Liberman & Mattingly, 1985; Rost, 2016; Tatham & Morton, 2006). Even minor listening errors, such as a missed morpheme, can lead listeners down an incorrect path or garden path when inferring meaning (Juffs & Harrington, 1996; Lightbown & Spada, 1990; Rost & Ross, 1991). Sandhi-variation has been shown to be one feature of spoken language that triggers difficulty for second language listeners' comprehension.

Furthermore, second language learners' backgrounds are unique. They carry with them unique historical experiences of target language intake that, in part, are influenced by their L1. These intake experiences can influence the activation of schema when learners process new input (Hammer, Jia, & Uchikoshi, 2011; Mai, Ngoc, & Thao, 2014). Broadly, for aural input, L1 language differences, including L1 phonology, are reflected in the way learners initially develop listening comprehension skills in their L1 (Cutler, 2000; Cutler & Otake, 1994; Fay & Buchweitz, 2014; Fox &

McGory, 2007). For example, language learners whose first language is Japanese, a *mora-timed* language that has rhythmically timed units, but who are learning English, a *stress-timed* language with units of speech timed by stress, have to develop an aural knowledge and ability to process the stress-timed structure of English in order to accurately comprehend the language (Mixdorff & Munro, 2013). Therefore, understanding individual listeners' abilities to comprehend target aural input with sandhi-variation and their English language learning background can add to a broader understanding of listening comprehension as a skill. By exploring links between individuals' backgrounds and their ability to comprehend sandhi-variation, the types of historical input that learners have received might be linked to the varying degrees to which they comprehend sandhi-variation.

However, studies about general second language learning background (e.g., experiences such as homestay) and listening comprehension are scarce. For example, Fujio (2010), in her preliminary study on the use of background and context clues, found no correlation between background and listening comprehension. In addition, she found that listeners used context clues more than background information. Other recent studies into study abroad and listening comprehension have traced significant proficiency gains over time (Allen & Herron, 2003; Beattie, Valls-Ferrer, & Pérez-Vidal, 2014; Kinginger, 2008), but these researchers did not investigate a direct link to the phenomena of sandhi-variation. More research needs to be conducted to investigate how second language listeners use their background in an L2 environment when processing specific types of aural input (e.g., elision).

With aural input, second language learners develop the ability to *detect* information. This process, whereby listeners find information within the speech stream without being prompted (Rost, 2016), occurs when lexical items and grammatical forms compete for cognitive capacity. For example, if a learner hears an utterance in which the speaker uses a low-frequency verb tense such as past perfect progressive, the listener might require more cognitive capacity to recall the structure, in turn limiting their recall of the lexical items in the utterance. Even processing constraints imposed at or below a second language learner's proficiency level, such as those that are controlled for speed, context, and lexis, can disable their ability to detect new linguistic information if, for instance, the lexical combination in an utterance is new to the listener (Field, 2008a; Rost, 2016; Skehan & Foster, 2001). With more research about L2 listening comprehension, teachers of L2 listening might be able to help learners develop their abilities to process a second language speech stream, and, by default, increase the learners' cognitive efficiency. This might be accomplished by drawing learners' attention to particular aspects of the speech stream. For example, shadowing or transcribing elided targets of language that are comprehensible to the learners might help them adjust to the appearance of elision and assimilation. However, prior to developing instructional interventions, understanding learners' second language proficiency and the comprehension of the phonological features of sandhi-variation would help to identify parts of the speech stream that cause comprehension difficulties. Instructional materials for practicing sandhi-variation, such as practicing linking consonants and vowels, have done this (Celce-Murcia, Brinton, Goodwin, & Griner, 2010/2016), but empirical evidence to support such materials is sparse. Without more research, the skills required by second language listeners to process the speech stream will remain uncertain.

Statement of the Problem

In this study I address three problems related to second language learners' comprehension of L2 speech. The first is the need to better understand proficiency problems encountered by L2 listeners as they decode the speech stream with the phonological features of sandhi-variation, elision and assimilation, by investigating the item difficulty hierarchy of the phenomena. Learners' ability to parse a speech stream for meaning is rooted in the comprehension of an acoustic signal. Nonetheless, the sandhi-variation that is embedded in acoustic signals in spoken English has received scant investigation in second language research. Results of database searches of Linguistics and Language Behavior Abstracts, EBSCO, ERIK, and Gale Academic OneFile for the search terms *applied linguistics* and *sandhi-variation* consistently contained two extensive studies by Ahn (1987) and Henrichsen (1984) and no others of relevance to this study. Expanding the search to include *connected speech, second language, foreign language, elision, assimilation,* and combinations of these terms, turned up approximately 40 articles published since the year 2000 that varied in relevance to the topic of this study. For example, many of the articles focused on very specific aspects of sandhi-variation, such as how experience effects non-native vowel contrasts (Simon & D'Hulster, 2012), or predicting the ease of acquisition for similar English and Catalan sounds (Fabra, 2005). I have used both the Ahn (1987) and Henrichsen (1984) studies as a starting point for this study because the researchers showed that sandhi-variation creates significant difficulty for non-native listeners in identifying words and comprehending sentential input. In both studies the researchers also showed that no significant difficulty exists for native speakers of English when asked to identify the same words identified by non-native listeners in a speech stream with sandhi-variation. However, while the studies of Ahn and Henrichsen were important additions to the literature, both were conducted prior to the commonality of computers. Much has changed in terms of understanding phonetics in relation to the acoustic signal (Cawley & Green, 1991; Díaz-Campos, 2004; Wojtaszek & Arabski, 2011). Few researchers have explored the effects of aural processing using sandhi-variation targets for the purpose of providing insight into the sandhi-variation phenomena and L2 spoken language processing. None have investigated the item difficulty hierarchy of targets embedded with the phonological effect of sandhi-variation.

The second problem addressed in this study is the scarcity of research on aural processing abilities of second language learners in relation to their understanding sandhi-variation in aural texts. How well and how easily listeners understand L2 speech, in other words the *comprehensibility* of target utterances (Saito, Tofimovich, & Isaacs, 2015), hinges, for non-native learners of English, on the challenges of understanding fluent speech (Nunan, 2002; Rubin, 1994). Comprehensibility involves the negotiation of lexical load, speech rate, and syntactic complexity, among other factors. Prosodic, lexical, syntactic, morphemic, and cultural variations (to name a few) create referential ambiguities; if learners have to focus on a target audio text that contains, for example, lexical ambiguity, any assessment of the listening process might be skewed because the origin of a breakdown in comprehension cannot be related directly to processing connected speech (Field, 2003). The breakdown might be associated with an inability to comprehend content (Krashen, 1981; Rost, 2016) or with distraction caused by the content (Field, 2003). There is a lack of understanding about aspects of aural texts that are challenging for second language learners and how different challenges, such as the amount of elided speech, affect aural processing.

The third problem concerns the lack of research investigating links between learners' backgrounds and their ability to handle listening texts, especially variations in the speech stream in target aural texts. Currently no researchers have linked the type of L2 input that learners receive with the act of processing the acoustic signal. While there is a growing body of research on the relationship between learners' L2 language learning backgrounds and their performance on standardized English assessment tests, no researchers have explored the relationship between language learners' backgrounds and their ability to process the acoustic signal, which includes sandhi-variation. Learners' experiences with the L2, for example during a homestay, might alter their ability to comprehend certain acoustic signals more than others. One's historical input may scaffold or cause confusion for one's present aural input.

Purposes and Significance of the Study

My first purpose is to investigate the item difficulty hierarchy of sandhi-variation types that learners have in relation to L2 listening proficiency. In this study, I narrow problems encountered while decoding the speech stream by L2 listeners through testing the participants' comprehension with two sandhi-variation types, elision and assimilation. By using vocabulary and syntax in the listening instrument pitched below the participants' levels of receptive vocabulary and syntax, I can focus on the participants' ability to process the sound stream because the lexis and syntax have been controlled. Also, by looking at the types of problems learners have when decoding the speech stream on the elicited imitation test used in this study, I can explore an empirical item hierarchy for the phonological features of sandhi-variation, elision, and assimilation. Furthermore, the item difficulty hierarchy might correlate with the type of English input learners have been exposed to (e.g., in a study abroad program) as revealed through interviews, as well as with the Test of English as a Foreign Language Paper-based Test (TOEFL PBT), the Listening Vocabulary Levels Test (LVLT), and the Modern Language Aptitude Test–Elementary (MLAT–E). An eventual identification of any item difficulty hierarchy of sandhi-variation types that learners follow as they develop L2 listening proficiency would improve our current understanding of L2 learners' listening comprehension, and also improve the design of target audio texts.

My second purpose is to evaluate links between aural input containing elision and assimilation and second language aural processing, to provide insight into how learners deal with sandhi-variation as they process such input. Prior studies of sandhi-variation and aural processing have provided a basic understanding of how processing is applied by second language listeners (Ahn, 1987; Henrichsen, 1984), and, more recently, how the brain handles the processing of processing utterances (Ding, Melloni, Zhang, Tian, & Poeppel, 2016). Gauging the processing ability of individual second language listeners provides additional research findings, provides instructors with insight into the needs of listeners, and might inform classroom content, such as aural-text type used in listening lessons.

My third purpose is to investigate through the use of interviews the aural input that participants have encountered prior to the interventions of this study, to help explain which types of aural input can facilitate intake. Interview data can reveal variations in participants' historical experiences with English. Furthermore, by cross-referencing the participants' abilities to process language containing sandhi-variation with their experiences of aural input, I show a correlation between input type and ability to parse speech. The aural experiences that participants have encountered relates to participants' abilities to process sandhi-variation.

The Audience for the Study

This study will benefit researchers by more fully highlighting the aspects and influence of sandhi-variation on second language learners' listening comprehension. Researchers have identified problems associated with listening comprehension in a second language, but these problems have remained largely theoretical: There is insufficient empirical evidence to conclusively support many of the claims of L2 listening researchers. Moreover, L2 aural processing models need to include the complexity of spoken language such as sandhi-variation; L2 researchers might take elements of complexity into consideration to a greater degree when developing research designs related to L2 speech processing or when assigning, for example, phonetic transcription onto an utterance without acknowledging the possible dynamic aspects of the utterance (e.g., intonation). Also, researchers might be further motivated to test claims made about various perceptual difficulties related to second language listening comprehension.

This study will benefit instructors tasked with teaching listening and speaking classes. Building a better understanding of the processes that second language learners use to comprehend aural texts is integral to improving learners' communicative skills. Instructors' abilities to choose appropriate listening texts for classroom use are important in helping learners improve their listening skills, as is knowing that comprehending authentic materials poses a more difficult challenge for second language learners than materials graded for learners' proficiency levels. The importance of both the choice of texts and knowledge of input is true for a variety of reasons, including lexical load. By becoming more aware of the difficulties second language learners encounter with aural texts, and identifying elements that might be challenging for them when used for instructional purposes, instructors can work to lower the cognitive load students experience during listening tasks.

Materials designers would benefit from understanding an item difficulty hierarchy for processing sandhi-variation. While it is common for second language listening researchers to recommend authentic listening audio for learners, there does not seem to be much regard for the difficult aspects of the speech stream embedded in authentic listening texts. This study highlights the challenges of processing listening texts, and thereby clarifies a need to focus more on processing sandhi-variation in listening materials.

Delimitations

The first delimitation of this study concerns the single L1, Japanese, of the participants. The participants' L1 greatly limits the generalizability of the hierarchical effect explored in this study because participants with an L1 other than Japanese might be able to process English phonology more or less easily because of L1 transfer (Cutler, 2015). In other words, L2 segmentation for listeners with L1s that encourage the same rhythm-based segmentation procedure as English might have advantages on the Elicited Imitation Test. L2 segmentation for listeners with L1s that differ from the L2 segmentation procedure, for instance, speakers of mora-based languages such as Japanese, might be disadvantaged. Thus, because the participants in the current study were all speakers of a single L1, the results cannot be generalized to speakers of other L1s.

The second delimitation concerns the age of the participants. The narrow range of participant ages, between 18 and 21 years of age, limits the generalizability of the results outside the age group represented by the participants. A wider span of ages would decrease homogeneity in terms of certain aspects of their generational culture (e.g., similar elementary school curriculum), so it is unclear how well the results can be generalized to English language learners who grew up in a different generation.

The third delimitation concerns the English proficiency levels of the participants as measured by the standardized (e.g., TOEFL IPT) and non-standardized (e.g., Elicited Imitation Test) proficiency tests used in this study. The results best represent similar learners who are motivated to graduate into the Liberal Arts program in the school, and with intermediate to upper intermediate level English proficiency. This delimitation suggests that the results can only be generalized to learners with lower English proficiency.

Organization of the Study

This study is divided into seven chapters. Chapter 2 contains the Literature Review and is divided into three sections: Sandhi-Variation, Language Processing in English, and Background and Length of Residency. At the end of Chapter 2, I describe the gaps in the literature, purposes of the study, and the research questions as informed by the literature review. In Chapter 3, Methodology, I describe the participants, instrumentation, research design, procedures, and analyses used to address the research questions. In Chapter 4, Instrument Validation, I analyze the Listening Vocabulary Levels Test, the Modern Language Aptitude Test, and an elicited imitation test to provide validity evidence for their stated purposes. In Chapter 5, Results, I present the results for each research question. In Chapter 6, Discussion, I interpret the results and discuss the theoretical and pedagogical implications. In

Chapter 7, Conclusion, I present a summary of the findings, limitations of the research, suggestions for future research, and my conclusions.

CHAPTER 2

REVIEW OF THE LITERATURE

In this chapter I review the relevant literature on the subject of sandhi-variation and comprehension of spoken English for Japanese learners. The review is divided into the following four sections: Sandhi-Variation, Spoken Language Processing in English, Language Learning Aptitude, and Background and Length of Residency. The Sandhi-Variation section is situated first, as it serves as background for the topic; it covers (a) an introduction of sandhi-variation in English, (b) sandhi-variation in Classical Phonetics, (c) L2 sandhi-variation research in English, and (d), sandhi-variation in current linguistic theory. Second, the Spoken Language Processing in English section contains (a) language processing theory, (b) L2 spoken language processing theory, (c) reasons for comprehension breakdowns of L2 learners, and (d) an explanation of syllables and mora. Third, the Second Language Aptitude section discusses the connection between aptitude and L2 acquisition. Fourth, the Background and Length of Residency section covers (a) a definition of *background*, (b) studies that utilized the participants' background information to link to L2 acquisition and phonology, and (c) length of residency and L2 acquisition. I conclude the chapter by discussing the gaps in the literature, the purposes of the study, and by listing the research questions that guide this study.

Sandhi-Variation

In this section, sandhi-variation is introduced as phonological phenomena. First, I introduce sandhi-variation in Classical Phonetics theory, discussing the phonological features of elision and assimilation. Examples of L2 research of sandhi-variation in English are provided, and this research is contrasted with sandhi-variation in current linguistic theory, including a brief consideration of *coarticulation theory* and a model of speech perception. Finally, I give an explanation to reflect my theoretical standpoint for the phonological types.

Sandhi-variation is an umbrella term for a variety of spoken phenomena. Kaisse (1985) defined sandhi-variation, from the Sanskrit word □□□□ (*saṃdhi*), meaning euphonic combination, as a variety of phonological processes that takes place at morpheme or word boundaries. Examples include the combination of different sounds across word boundaries and the changing of sounds due to neighboring sounds or to the grammatical function of words adjacent to a dominant sound (Ahn, 1987; Rost, 2016). Celce-Murcia et al. (2010/2016) defined sandhi-variation as the tendency of words to "run together" in spoken discourse (p. 163), and recommended teaching the phenomena, including teaching the different types of sandhi-variation—such as *elision*, where sounds within or between words disappear, or *assimilation,* where sounds within or between words combine—to L2 learners from low-proficiency onward. Tatham and Morton (2006) defined the concept as an abstract notion, different and idiosyncratic for the speaker and for the listener, and dependent on variables, such as the varying intensity of the vocal folds in use. Though the definitions differ, researchers have suggested that learners need to comprehend a variety of speech possibilities, such as differences in stressed syllables or adjustments to the pronunciation of speech due to utterance rate, because spoken language always contains variance from one exemplar to the next (Bowen, 1975; Jusczyk, 2000).

Sandhi-variation, the general phenomena of spoken variations in the connections of words, is frequent in English speech. However, it is typically ignored in spelling. Contractions are the exception. For example, *don't* is produced from *do* *not* by eliding the /o/ in *not*, a phenomenon that is sometimes called *blending*. Therefore, a burden rests on listeners to parse the spoken language without a textual guide and to code sounds onto words and sentences that do not always, if ever, align with the written language (Hieke, 1990; Jusczyk, 2000).

Researchers have attempted to partially explain the phenomenon of sandhi-variation in two ways. The first explanation is as an artifact of *the principle of least effort* (Wells, 1982), which is the idea that speakers choose utterances that involve the least articulatory effort without losing meaning. The second explanation, more recently offered, *the principle of ease of articulation* (Ladefoged & Johnson, 2014), is the idea that speakers tend to conserve energy when using articulatory organs. There is no existing empirical evidence for either of these two principles; nonetheless, sounds in spoken utterances are connected and vary in their degrees of connectedness. In the next section I discuss how sandhi-variation has been defined historically in Classical Phonetics.

Sandhi-Variation in Classical Phonetics

Classical Phonetics is the idea that the spoken utterances of language are made up of strings of discrete sounds (Tatham & Morton, 2006). The idea that they contain discrete sounds means, in turn, that these spoken utterances can be uniformly and completely represented by symbols in, for example, a phonetic transcription. In Classical Phonetics, the symbolic standard representation is meant to accurately model perceived spoken utterances and to prescribe pronunciation. I have annotated all specific forms that follow using the International Phonetic Alphabet (IPA) General American transcription, which is one accepted method for representing speech (Field, 2008a). I have done so in order to satisfy the need for a representation of possible pronounced sounds of target examples within the soundwave of an utterance. By doing so I am not, however, making an argument in support of Classical Phonetics or strings of discrete sounds. The examples should be understood as both one possible way of transcribing the specific targets and, above all, theoretical.

In Classical Phonetics, there are two main types of sandhi-variation, *internal* and *external* sandhi-variation. Internal sandhi-variation features the alteration of sounds at morpheme boundaries. Internal sandhi-variation happens, therefore, within words. Examples of internal sandhi-variation in English are the flapping /t/ in *hotter* [hɑɾər], the glottalization of /t/ in *mitten* [mɪʔən], the velarization of /l/ in *film* [fɪɫm], and the coarticulation (a term used differently in current theory, discussed later) of /t/ to the preceding /n/ in *interact* (Ahn, 1987). External sandhi-variation, however, refers to changes found at word boundaries (Kaisse, 1985). Examples of external sandhi-variation in English include the elided consonant cluster reduction in *land reform* [lænrɪfɔrm], in which the consonant /d/ in *land* (i.e., clustered between /n/ and /r/) is elided to make pronunciation easier, and the coalescent assimilation in *watch you* [wɑtʃu], in which the final alveolar consonant sequence *tch* in *watch* coheres with the neighboring word-initial /y/ in *you* to create a [tʃu] sound not available alphabetically (adapted from Ahn, 1987; Field, 2008a).

Furthermore, for external sandhi-variation, there are *proclitic* forms, where a word is treated as part of the following word, and *enclitic* forms, where a word is treated as part of the preceding word. Proclitic words are said to lean forward; for example, *it was* becomes *'twas*, as the first word *it* moves forward (i.e., leans) to become a part of the second, *was*. Enclitic words are said to lean backward; for example, the phrase *want to* becomes *wanna*, as the second word *to* moves backward (i.e., leans) to become a part of the first word, *want*. Some enclitic words in English form a written contraction (e.g., *don't*) and some, for example, *going to* (always formally written in full form), depend upon the availability of a written contracted form in English (Ahn, 1987; Berkling, Zissman, Vonwiller, & Cleirigh, 1998).

Finally, there are basic functions of elision and assimilation. I have adapted the following definitions and examples of elision, and assimilation—including *progressive assimilation*, *regressive assimilation*, and *coalescent assimilation*—from Celce-Murcia, Brinton, Goodwin, and Griner (2010/2016). The forms are represented symbolically.

Elision. Also referred to as *deletion*, *ellipsis*, or *omission*, elision happens when sounds are removed or not clearly articulated. Possibilities for elision can occur as follows:

- When /nt/ is placed between two vowels or before a syllabic /l/, the /t/ sound is deleted: /t/ printer [prɪnər], pronto [prɑnoʊ], center [sɛnər].
- When /t/ or /d/ appears second in a sequence of consonants or in the middle of two other consonants: /t/ restless [rɛsləs], compactly [kəmpækli]; /d/ blindness [blaɪnnəs].
- When /t/ or /d/ is clustered with another consonant at a word end, and succeeded by an initial consonant:

  East + side: [issaɪd]

  blind + man: [blaɪnmæn].
- Elision of a medial vowel if the medial vowel is unstressed and follows a strongly stressed syllable: chocolate [ʧɔklət], every [ɛvri], camera [kæmrə].
- *Aphesis* occurs when an unstressed initial vowel is elided: because: 'cause [kɔz], about: 'bout [baʊt], around: 'round [raʊnd].

- Elision of ending /v/ in *of*, when the succeeding word is a consonant: lots of time: [lɑtsə taɪm], waste of money: [weɪstə mʌni].
- Elision of initial /h/ and /ð/ in pronominal forms: ask her [æskɜr], help them [hɛlpɛm]. (Celce-Murcia et al., 2010/2016, pp. 172-173)

Elision in spoken language can create confusion for L2 learners. Elided forms of words have been shown to create competitors with existing vocabulary for L2 learners (Spinelli & Gros-Balthazard, 2007). Word recognition can become more difficult for learners when elided forms are used because one elided word can resemble another word or phrase. For example, the final /g/ in the word *living* is often elided as livin' [lɪvɪn] and confused, when heard by low-proficiency Japanese learners in my classroom environment, as two separate words, *live in*. This confusion confounds not only the grammar of the heard utterance for the learner but also the meaning, as the two lexical items differ semantically.

Assimilation. Assimilation refers to the physical adjustment of one sound, the assimilating sound, that takes on the characteristics of a neighboring or conditioning sound. For the purposes of this study, I review three types of assimilation: *progressive assimilation*, *regressive assimilation*, and *coalescent assimilation*. Progressive assimilation happens when the conditioning sound moves forward to affect the sound that follows it, the assimilated sound. Celce-Murcia et al. (2010/2016) suggested that most progressive assimilation is rooted in morphology, and "historically motivated" (p. 394), where the phonological changes (e.g., the voiced /s/ below) are a result of older forms of the language:

Conditioning sound	Assimilated sound
-*s* ending	

flags [flægz] voiced /s/

sacks [sæks] voiceless /s/

-*d* ending

grooved [gruvd] voiced /d/

wished [wɪʃt] voiceless /d/.

Voiced here refers to the vibration of the vocal folds to a noticeable degree by the speaker, with the vibration varying in intensity. *Voiceless* means that the vocal folds are not vibrating. The differences between voiced and voiceless inflections are said to help listeners identify differences. For example, a voiced /g/ leads to the –s being pronounced as /z/ in *bags*, which might aid listeners in differentiating the word from *backs*.

Regressive assimilation precedes the conditioning sound, yet is affected by it:

have + to [həv tʊ]: [hæftə]

has + to [həz tʊ]: [hæstə]

used + to [juzd tʊ]: [juzstə]

Assimilated sound	Conditioning sound
[hæf	tə]
[hæs	tə]
[jus	tə].

In the examples above, the plosive /t/ in *to* moves the labiodental lenis (i.e., voiced) /v/ to fortis (i.e., voiceless) /f/. This occurrence of voicelessness is often termed *assimilation of voice*. An alternatively accepted form of prescribed regressive assimilation allows for the voicing of /s/ in the example *used to* [juzs tə]: [juz tʊ] (Celce-Murcia et al., 2010/2016).

Regressive assimilation can also account for changes to the manner of articulation; for example, *give me* [gɪv mi] can be assimilated as [gɪmi]. This type of manner-based regressive assimilation occurs in informal speech (Celce-Murcia et al., 2010/2016), whereas other examples of assimilation, such as the assimilation of voice, can occur in formal speech. Celce-Murcia et al. (2010/2016) did not define differences between formal and informal speech or give reasons for the occurrences.

Coalescent assimilation occurs when the blending of two sounds creates a third sound. In English, this occurs commonly in cases where alveolar consonants, for example, /t/, are followed by initial palatal /y/, such as in *you*, *your*, or *yourself*:

Rule	*Examples*
/s/ + /y/: / ʃ /	tissue
	He's proposing this year.
/z/ + /y/: / ʒ /	treasure
	Does your mom care?
/t/ + /y/: / tʃ /	stature
	Is that your phone?
/ts/ + /y/: / tʃ /	She lets your dog out.
	He hates your friend.
/d/ + /y/: / dʒ /	procedure
	Would you help with this?
/dz/ + /y/: / dʒ /	She needs your paper.
	He always heeds your advice.

(Celce-Murcia et al., 2010/2016, pp. 168-171)

Defining coalescent assimilation is contentious for researchers. Some researchers have claimed that it is a form of progressive assimilation (Roach, 2009), and others that it is a form of regressive assimilation (Sobkowiak, 2001). Most researchers have described it as *reciprocal* (i.e., moving back or forth depending upon the assimilating case) (Collins & Mees, 2013).

Moreover, all three forms of assimilation—progressive, regressive, and coalescent—vary for individual speakers (Ellis & Hardcastle, 2002), and might be manifest in many languages but in different ways (Cutler, 2015). Variations occur in both the use of assimilation in spoken utterances and degree of application. Cutler (2015) provided examples from English, Dutch, and French: "In English, place of articulation is matched: [tb] becomes [pb], for instance in *ratbag*, which can then sound like *rapbag*. In Dutch or French, it is likely to be the voicing that is matched: [tb] can turn to [db], as in Dutch *houtblok*, block of wood, or French *lutte brutale*, brutal fight" (p. 200). The examples provide evidence that assimilation is a natural aspect of articulating speech sounds. Furthermore, research on the perception of spoken assimilation has shown that listeners also adjust to assimilation. For example, Darcy, Peperkamp, and Dupoux (2007) claimed that speakers of English detect place-assimilated words more systematically than voice-assimilated words, but that speakers of French detect the inverse pattern. The varieties of assimilation expose the complexity of spoken utterances and the challenge of comprehending spoken utterances, which is a process that must be relearned in an L2.

It is important to note that even definitions of types of sandhi-variation have differed for researchers. Rost (2016), for example, included elision as a form of assimilation, and he described both elision and assimilation as *coarticulation*, as rapidly spoken sounds, and, simultaneously, as products of minimal energy used by speakers. These descriptions differed from Celce-Murcia et al. (2010/2016), who did not combine the two definitions of phonological features, elision and assimilation, into one (viz., assimilation) and did not mention coarticulation as an aspect of the phonological features, but described the phonological features of sandhi-variation as one method for speakers to regulate rhythm. However, the inclusion of elision and assimilation definitions and examples, despite the inconsistencies between researchers, provides an overview of the sandhi-variation phenomena. Specific examples of elision and assimilation in the instruments used in this study are highlighted in Chapter 3.

Sandhi-Variation in L2 Research

In this section, I provide an overview of two major studies focused on sandhi-variation. Next, I look at relevant current research in sandhi-variation and spoken word recognition. I discuss basic models of perception, aural perception, and an example of a waveform for a spoken utterance.

Henrichsen (1984), in a study of sandhi-variation, focused on perceptual saliency in second language acquisition, which is the hypothesis that less language knowledge equals greater signal dependence, while more language knowledge equals greater signal independence. Comprehension depends on how much of the language is recognized in the acoustic stream. Henrichsen theorized that native speakers, when tested for listening, would show no significant difference between items with sandhi-variation and items without. He theorized that non-native speakers, however, would show a significant difference. In line with Henrichsen's hypothesis, native speakers showed no significant difference in comprehension between the presence and absence of sandhi-variation, whereas both the high- and low-proficiency non-native participants showed significantly lower levels of comprehension when exposed to sandhi-variation. Henrichsen's simple experiment revealed the difficulty of processing the spoken language signal. Furthermore, the results suggested the importance of phonological proficiency.

Ahn (1987) conducted quantitative research on the influence of sandhi-variation and affective factors on listening comprehension of spoken English in an EFL context. He used a combined background and attitude/motivation questionnaire as one instrument and a sandhi-variation test as a second instrument. Ahn found that participants' scores were significantly lower for test items that included sandhi-variation; however, he found a weak or complete absence of correlation between attitudes/motivation regarding listening comprehension and the presence or absence of sandhi-variation. Ahn concluded that, although some affective factors might influence comprehension of speech, their influence is minor when compared to sandhi-variation.

It can be assumed that the two studies by Henrichsen (1984) and Ahn (1987) use the idea that discrete phonological segmentation occurs in a speech stream. Neither researcher mentioned coarticulation theory.

More recently, researchers have continued to look at the general phenomena of sandhi-variation. Wong et al. (2015) examined the role of perception on contraction, elision, and assimilation, which they labeled *reduced forms* of speech. They used this term in place of the term *sandhi-variation*, though *reduced forms* should not be confused with *reduction*, which is a term used by some researchers (e.g., Field, 2008a) for words that are dropped or shortened because they lack importance in the overall meaning of the utterance. They identified and tested five phonological skills related to the perception of reduced forms: *spoken word discrimination*, the facility to discriminate and categorize speech sounds in English; *part-word* *recognition*, the ability of the listener to make use of a noisy waveform with relatively few acoustic-phonetic signals; *phonemic awareness*, which involves knowledge of the phonemic structure of English, such as rhyme; *vocabulary knowledge*, which involves top-down influences, such as predicting upcoming words; and, *phonological memory*, which is the ability to hold longer streams of speech in memory and is associated with L2 vocabulary acquisition. Wong et al. (2015) hypothesized that L2 comprehension difficulties were the result of differences between *canonical* (i.e., words pronounced in isolation with all syllables intact) and *reduced* (part-word) word forms. They looked at (a) the relationship between the participants' perception of reduced forms and listening comprehension, and (b) which aspects of phonological representation best predicted the accurate perception of reduced forms. They found that receptive vocabulary and the ability to recognize reduced forms in native English significantly predicted perception for the 60 participants, who were naïve speakers of Chinese. The results were in line with previous studies that identified the recognition of vocabulary and reduced word forms as interdependent when participants perceived meaning in reduced words (Trueswell, Tanenhaus, & Garnsey, 1994). Moreover, the results were in line with a similar study that showed a link between vocabulary and phonology. Vowel perception, a type of segmental awareness (i.e., a listener's ability to segment the waveform of a spoken utterance), was also related to the participants' vocabulary size (Bundgaard-Nielsen, Best, Kroos, & Tyler, 2012).

Kennedy and Blanchet (2013) showed a relationship between language awareness, demonstrated through reflective journals, and the development of L2 (French) sandhi-variation for participants in a listening course. Their research was focused on formal instruction—whether it was useful for improving the perception of familiar and standard connected speech processes—and whether the quality of learner awareness was related to connected speech perception. The participants completed activities such as taking dictation based on daily lessons that were focused on a particular form of connected speech (e.g., elision). The dictations were also used as a measure of the developing perception of target forms. The learners completed weekly journal entries about their learning experiences and language use. The results showed that participant perception improved most when they did not focus on the specific speech connection, but instead when they focused on a method of using the speech connection to garner meaning from utterances. In other words, when focused on context, they were more oriented toward an awareness of connected speech processes. Kennedy and Blanchet concluded that L2 learners can benefit from explicit instruction of sandhi-variation processes.

Research at the neurolinguistic level has been conducted to look at how listeners process sandhi-variation. Ding, Melloni, Zhang, Tian, and Poeppel (2015) isolated spoken language activity in the auditory cortex, the part of the temporal lobe of the brain responsible for processing aural input, by correlating speech waves with electro-encephalographic (EEG) data while participants listened to spoken language. This procedure is called *cortical tracking*. Ding et al. found that, "In speech, hierarchical linguistic structures do not have boundaries that are clearly defined by acoustic cues and must therefore be internally and incrementally constructed during comprehension. We found that, during listening to connected speech, cortical activity of different timescales concurrently tracked the time course of abstract linguistic structures at different hierarchical levels, such as words, phrases and sentences" (p. 158). At the very least, their results implied the need for L2 learners to develop ways of negotiating acoustic cues that are not clearly defined or discrete.

Furthermore, current research in neurolinguistics is not based on the notion that the speech stream is made up of discrete sounds. The idea of discrete sounds has been discarded based on the understanding that spoken utterances are more variant and dynamic. For example, Kösem, Basirat, Azizi, and van Wassenhove (2016), in their article on neural activity and the prediction of word parsing in ambiguous speech streams, define a *speech stream* by stating that spoken linguistic structures do not have boundaries that are clearly defined by acoustic cues and must therefore be internally and incrementally constructed during comprehension. Although neurolinguistic models of word parsing have sparked controversy by pointing to an internal, innate mechanism for processing as opposed to a developed understanding through usage, the models presuppose a dynamic model of a speech stream.

Other research has been focused on specific phenomena under the umbrella of sandhi-variation, such as regressive assimilation. Cutler (2015), for example, referred to the phenomena of elision and assimilation together as "phoneme sequence distortion" (p. 277), but when she dealt with processing, she focused more on types of assimilation and reasons for it than on elision. She concluded that listeners differ based on their L1, so the task of listening differs from language to language.

In the next section, I consider how sandhi-variation fits into current linguistic theory and how this fit might be situated in relation to L2 language learners.

Sandhi-Variation in Current Linguistic Theory

In this section I discuss the current theory of phonetics that places different emphases on spoken utterances, which I include for the purpose of highlighting the complexity of speech, especially for L2 learners, and to contrast current theory with Classical Phonetics. More recent theory and experimental work has moved on to *coarticulation theory*, and, consequently, the redefinition of assimilation. Speech by definition is connected, so the phonological features of sandhi-variation still exist within any speech stream and within current theory. However, the theory behind the purpose and generation of the phenomena has changed. Coarticulation theory places greater importance on the relationship between the speaker and listener (Jusczyk, 2000; Tatham & Morton, 2006), and disposes of the idea that adjustments in a speech stream, for instance, elision, are a consequence of strings of discrete phonological sounds (Kühnert & Nolan, 1999; Tatham & Morton, 2006).

Coarticulation refers broadly to the effect of an influence or overlap of an articulatory segment on adjacent segments, whereby the articulatory segment is not realized identically in every environment (Díaz-Campos, 2004; Kühnert & Nolan, 1999; Tatham & Morton, 2006). For example, the word *happy* [hæpi] might be coarticulated by initially moving the tongue into a position to pronounce [æ], which could then be heard during the [h]. At [æ], the lips are closing for the [p], the tongue moves into position to pronounce [i], and still the lips remain in a position from the [p]. This example of [hæpi] could change in relation to the word or sound coming before or after it in a speech stream. Coarticulation theory presupposes the theory of classical phonetics by attempting to explain the "high level phonological planning" (Tatham & Morton, 2006, p. 20), instead of only labeling potential aspects of speech output. High level phonological planning refers to phonological production that includes neuro-motor, articulatory, and acoustic aspects (e.g., the physical movements of the vocal folds to produce an utterance) that goes into speech production. The theory suggests that these aspects must be represented in realistic models of speech output, in contrast to the rudimentary string of discrete sounds found in Classical Phonetics. Furthermore, coarticulation theory concerns how acoustic signals are assigned to as-yet-unspoken words and sentences, and allows for a range of contextual variation with limits determined only by the specific language or accent (Keating, 1990). The theory frees up the availability of dynamic aspects of speech, for example, allowing for a speaker's mood or the variation of a momentary adjustment of sounds combining, that is not accounted for in Classical Phonetics (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). In other words, any word falling within a speech stream can take a variety of forms—within the spectrum of physical possibility—depending on combinations of accents, prosody, speed, words preceding and following the target word, use in the context, and other variables.

Assimilation in current usage refers to the influence or influences that phonological segments have on one another, and differs from coarticulation in two specific ways. According to Tatham and Morton (2006), assimilation is "voluntary or optional" (p. 23). It occurs prior to an utterance; therefore, the speaker can manipulate the target utterance. For example, the word *winter* [wɪntər] contains the alveolar production *nt* and can produce a minimalized /t/, so the finished utterance would sound like [wɪnər]. When the word *wintry* [wɪntri] is uttered, the /t/ remains pronounced. The degree of sound change of the /t/ for the speaker from [wɪntər] to [wɪnər] is optional (Yu, 2013). In contrast, coarticulation is not voluntary, and happens during an utterance, though speakers can also manipulate the effect.

There are many ways to say any one word. Coarticulation theory aims to explain the complexity of spoken utterances for both the speaker and the listener, so utterances are recognized as the idiosyncratic output of speakers and idiosyncratic input for listeners. The International Phonetic Alphabet (IPA) is used in coarticulation theory only to represent potential utterances, for example, if an utterance is repeated in a different way or a listener hears an utterance differently than the speaker intends it. IPA does not represent the articulations that produce sound variation. For example, the IPA transcription of the word *cruise* is [kruz], but often when the word is pronounced, the lips are rounded for the duration of the utterance. Lip rounding is needed for the pronunciation of /u/ and the /r/, but, according to coarticulation theory, the cognitive command to round the lips is sent to the articulators at the initial /k/. Furthermore, the rounding of the lips remains in effect after the end of the /u/ and into the final /z/. IPA transcription does not reveal this effect, which explains one reason why IPA is recognized as an abstract way of representing spoken utterances (Tatham & Morton, 2006).

Coarticulation theory also views listeners as playing an active role. Listeners segment the signal as they have perceived it, and the segmentation they do loosely correlates with the actual spoken utterance. In other words, the conceptual idea as generated by the listener of a speaker's utterance is *bound*, or portioned off into segments, by the listener in order to comprehend it. For L2 language learners, developing the ability to bind utterances and create meaningful segments for comprehension requires practice. For L2 learners, the theory suggests that working with a greater variety of spoken language targets can serve to expand their phonological options and automatize the process of perception more authentically.

Another theoretical aspect of coarticulation theory that can challenge L2 learners' aural comprehension is that each speaker produces utterances idiosyncratically, meaning that individuals develop their own solution to the integration of a speech stream within the physical spectrum of speech (Kühnert & Nolan, 1999). Instead of one rule for language-specific sandhi-variation, variations of rules are potentially allowed for by the mechanisms used for the production of each utterance. For example, variations of intonation, stress, and rhythm, or dialectical variations such as the nasalization of vowels, can provide different renderings of utterances with sandhi-variation. This characteristic of spoken utterances explains the need for listeners to adjust to different dialects of an L1, and why L2 listeners perform better at processing a target text where the speaker's accent is familiar (Baker & Bradlow, 2009; Brouwer & Bradlow, 2014). Perhaps, in addition to the volume of aural input, a variety of speaker input, in the form of different accents and dialects, is important for L2 learners.

For researchers to adopt coarticulation theory, the accepted conditions of the perception of a target acoustic signal must also change. According to Tatham and Morton (2006), the perception of language is as follows:

> Perception is essentially an act of interpretation since it is clear that there is no linear correlation or "direct path" between the acoustic signal and the assigned symbolic representation. What perceivers do not do is discover or register the linguistic objects of what they hear—this would imply that these objects and their labels are in the soundwave. What they do is interpret what is heard in a complex process of assignment, from what they already know, of symbolic representations. (p. 20)

Phonetic transcriptions (e.g., IPA) are symbolic representations linked to aural possibilities and the consequences of the dynamic interactions of speech production. Underlying the coarticulation theory of speech production and perception are dynamic, observable phenomena that the simplicity of Classical Phonetics cannot articulate or explain. For instance, sounds made by speakers are subjective. A speaker might report a sound, for example, [s], as one just uttered, but this idea ignores the necessity of how they might have produced the sound with variant, idiosyncratic articulatory movements. IPA simplifies complex phenomena, which perhaps creates an underestimation of the difficulty that listeners, especially L2 listeners, face.

Figure 1 shows a basic model of speech perception (Tatham & Morton, 2006); the speaker creates an idealized utterance, renders it, and then the acoustic signal degrades, that is, it is affected by noise (e.g., a mispronounced affix or an external noise, such as a nearby train), or any detraction from the rendered acoustic signal as it transfers to the listener. Listeners actively assign a symbolic representation based on their prior knowledge of the semantic content and the context clues available. The listeners have to reverse the process of sandhi-variation to recover the *utterance plan*, the idea the speaker has and the idea that the listeners think the speaker has embedded in the utterance, by piecing together segments in their mind. In this model, speech perception is active because listeners must actively construct the conceptual content of what is heard (Best & Tyler, 2007; Klatt, 1979). L2 listeners must be able to create an utterance plan for spoken utterances based on their conception of the utterance plan of speakers and of the target language that they can cognitively activate.

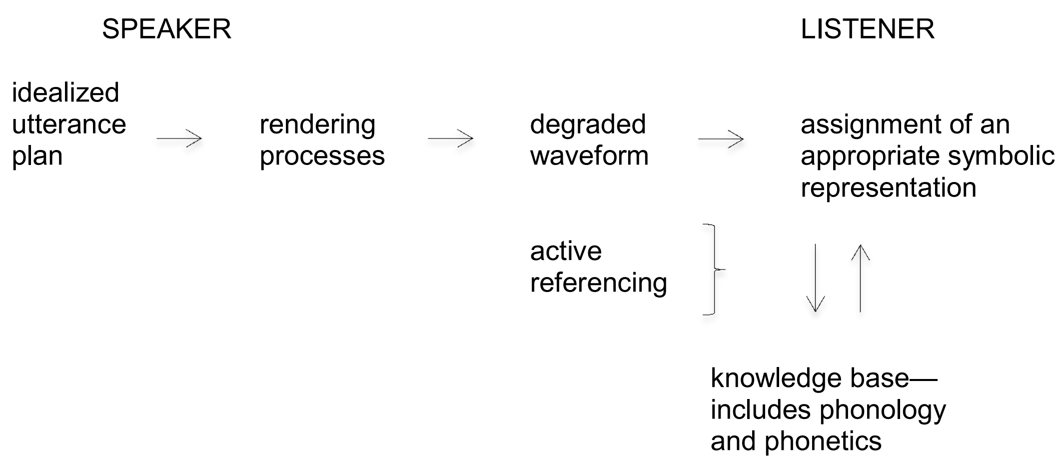

Figure 1. Basic model of speech perception: the perceptual task of listeners. Adapted from *Speech Production and Perception* (p. 199), by M. Tatham and K. Morton, 2006, New York: Palgrave MacMillan.

The acoustic signal does not contain noticeable phonological segments. This phenomenon can be shown within a waveform. Figure 2 is an abstract representation of a waveform for the spoken phrase, *where are the segments*. A physical spoken waveform is continuous from start (i.e., *where*) to finish (i.e., *segments*); it is a single stream of sound, and listeners need to create segments from the steady stream of sound while they listen.

Figure 2. Waveform for the phrase, *where are the segments.*

I conclude this section by restating it as relevant to this study of L2 language learners. For this study, I hold the theoretical point of view that spoken utterances are not strings of uniformly discrete sounds, but, instead, are products of deeper combining factors (e.g., speaker idiosyncrasy), and this idea might differ from the views taken by some researchers (e.g., in the choice of items on one of the instruments and in the discussion of the data). Some L2 researchers still use Classic Phonetics (e.g., Edwards & Zampini, 2008; Kuhl, Stevens, Hayashi, Deguchi, Kiritani, & Iverson, 2006; Kuhl, Tsao, & Liu, 2003; Major, 2001). In my opinion, the theory that discrete sounds produce the dynamic process of speech does a disservice to the processes required by L2 language learners as they develop listening comprehension in English. Thus, I am motivated in this study to investigate proficiency problems encountered by L2 learners as they decode speech streams. Moreover, because sandhi-variation is a blanket term used for several possible phonological changes to any speech stream, I use the specific terms *elision* and *assimilation* when referencing items in this study.

Spoken Language Processing in English

In this section, Spoken Language Processing in English is introduced. First, I introduce the theory of *anticipatory system*, a concept of how learners predict, as a model for processing, and give an example of the *feed-forward anticipatory model* that relates to spoken language processing. Examples of studies in L2 processing are then provided, including a brief consideration of L2 language aptitude. Next, an explanation of language acquisition through the theory of sound mapping is given followed by a discussion of usage-based listening acquisition. I then discuss language differences that lead to breakdowns in perception, and use syllables as an example of such a language difference.

Anticipatory System and Feed-Forward Anticipatory Model

In this section, the theory of *anticipatory system*(s) as background for the *feed-forward anticipatory model* is explained followed by a discussion of the feed-forward anticipatory model as a theory for listeners of spoken English. This information is included as background for the section that follows it, L2 Spoken Language Processing. Furthermore, this section highlights listeners' primary task: to understand spoken language.

Linking L2 language processing with current theoretical models of language processing might provide insight into how these same learners negotiate sandhi-variation in speech signals. Current L1 and L2 language processing theories, such as the speaker/listener collaboration of the feed-forward anticipatory model, stem from models of language (Jackendoff, 2002), and cognitive and biological models (Toates, 2005), though no model is suited to all purposes (Tatham & Morton, 2006). Jackendoff, for example, suggested a multiple layered and multiple direction system that works for phonological, syntactic, and conceptual structure of language that requires processing by the listener (see p. 199 for more on parallel grammar implemented as processing architecture). *Higher-order processing* refers to the act of extrapolating beyond immediate sensory input, based on predictions. This type of processing occurs when humans engage in listening, and suggests that different means, such as the use of prediction, can be exploited to reach a given goal. Toates (2005) argued that this phenomenon is part of an evolutionary process. He also suggested that such processing is not innately developed, and develops over time as learners' consciousness becomes refined to the nuances of stimuli. These models are also task dependent, so processing utilizes different areas of the brain depending on the task (Hickok & Poeppel, 2007). Thus, some current theories are based on ideas that processing is collaborative, flexible, and develops over time.

In this study, I accept a model that stems from *anticipatory systems* (Guenther & Vladusich, 2012; Pezzulo, 2008), which is a widely accepted theory within both cognitive and biological models of L1 language processing. There is evidence for such a system in brain imaging studies (Porro et al., 2002). Rosen (1985) provided a well-known definition of (an) *anticipatory system*, as "a system containing a predictive model of itself and/or its environment, which allows it to change state at an instant in accord with the model's predictions pertaining to a latter instant" (p. 339). Anticipatory systems are observable in all areas of human cognition and are not exclusive to language (Husserl, 1991). When listeners process an acoustic signal in an anticipatory system, it is done with their simultaneous anticipation of the future acoustic signal. This is called a *feed-forward anticipatory model* (Guenther & Vladusich, 2012; Kröger, Kannampuzha, & Neuschaefer-Rube, 2009; Rosen, 1987; Tatham & Morton, 2006). The anticipation occurs with regard to the signal's direction or plan, and supports listeners in binding the segment within a speech stream. Listeners use accessible information available in contextual clues and prior knowledge not only to follow speaker utterances, but also to anticipate the direction of the content in order to facilitate the interactive process. The description explains listening at the basic level.

In the feed-forward anticipatory model, Figure 3, the listener's *supervising agent* (i.e., the cognition of the listener) monitors the utterance plan of the speaker (Briscoe, 2002; d'Inverno & Luck, 2004). The listener and speaker collaborate, as the listener attempts to match the utterance plan. The utterance and matching, or

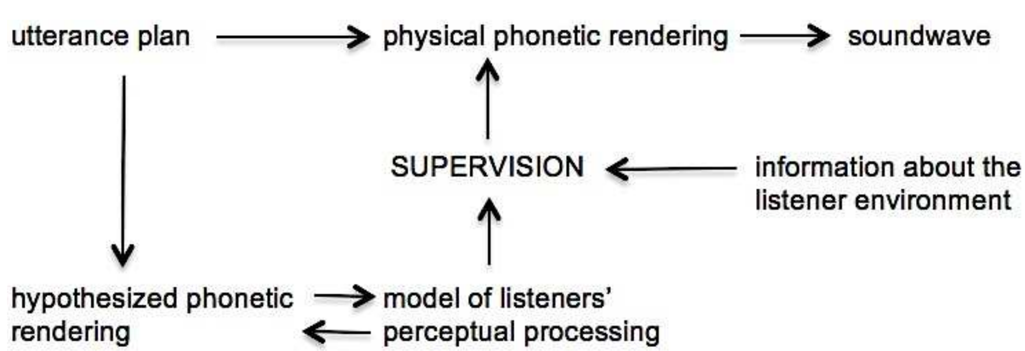

Figure 3. Feed-forward anticipatory model. Adapted from *Speech Production and Perception* (p. 217), by M. Tatham and K. Morton, 2006, New York: Palgrave MacMillan.

collaborative mapping, is dynamic, and involves multiple, partially redundant spectral and temporal cues within the signal, which can be exploited by the listener (Hickok & Poeppel, 2007). Moreover, the model in Figure 3 supports the theory that the listener, in either a first or second language, must acquire cognitive representations, or their version of the target utterance, in order to attempt to comprehend it within an acoustic stream. In the next section I look at theories of L2 spoken language processing in order to present similarities and differences to L1 theories, and to present research in L2 spoken language processing.

L2 Spoken Language Processing

In this section, I first present background information about L2 language processing. Next, I review several studies on processing conducted with L2 participants that relate to the current study.

Similar to L1 language processing, L2 language processing is theory based. Current theories are similar to those of L1 language processing, including the argument that listeners and speakers collaborate to create meaning. However, the terms used to define particular aspects of processing differ between L1 and L2. For instance, the utterance plan is not commonly used in L2 processing theory. Rost (2016) identified two principle heuristics needed for perception: recognition and categorization. He suggested that listeners should maximize recognition in order to handle the challenges inherent in dynamic speech, such as speech spoken with elision, and to minimize categorization in order to allow for speaker idiosyncrasies (e.g., utterance rate). Despite the difficulties involved in processing spoken language, Rost argued that the speech signal is redundant and therefore sampling from it, where "only limited extracts of a signal are perceived" (p. 301) is sufficient for listeners.

Cutler and Broersma (2005) hypothesized that listeners' L1 always dictates how other languages are processed, thereby adding more challenges to the role of non-native listeners than simply dealing with a speech stream and speaker differences. However, Rost (2016) has argued that the phonological disconnect between listeners' L1 and L2 can be compensated for by social reasoning and vocabulary networks. Thus, in either an L1 or an L2, listeners need to already have acquired the phonological and lexical content of the speech signal they hear.

Several investigations of L2 processing are relevant to the current study. Révész and Brunfaut (2013) looked at task factors in L2 listening comprehension to explore aural processing in advanced ESL learners. They had 77 participants in an English for Academic Purposes (EAP) summer program perform 18 versions of the same L2 listening task. After performing the listening comprehension task, which was analyzed for speed, phonological complexity, lexical complexity (i.e., lexical frequency, density, and concreteness of content words), syntactic complexity, discourse complexity, and explicitness, the participants completed a perception questionnaire. The researchers concluded that lexical complexity predicted task difficulty best because it impedes processing and blocks learners' abilities to create a version of the utterance plan, and that syntactic complexity and speed (i.e., syllables per second including and excluding pauses and pauses per second) held no significance in determining task difficulty. The participants had a much higher mean proficiency level than Henrichsen's (1984) participants, though this can only be estimated by extrapolating from two proficiency tests. Révész and Brunfaut used the International English Language Test System (IELTS) and Henrichsen used the Michigan Test of English Language Proficiency (MTELP). Ahn (1987) did not administer a proficiency test, though he did state that the participants had completed six years of English study and were English majors. The researchers also found that the participants' perceptions of task difficulty, as reflected in the perception questionnaire, were not in line with task difficulty as measured by the rate of delivery, linguistic complexity, and explicitness of the text. Révész and Brunfaut concluded that this discrepancy could have resulted in a situation where the participants allocated more processing energy to the tasks that they perceived as demanding.

In a similar study, Brunfaut and Révész (2015) investigated second language listening, task difficulty, and listener characteristics. They investigated the relationships between task difficulty and task input and response, as well as performance and working memory and listener anxiety. The researchers measured response choices, for instance, using lexical frequency measures in the multiple-choice options of the listening task. Performance in their study meant L2 listening ability as measured by the participants' overall listening score on the Select Missing Word task of the Pearson Test of English Scored Practice Test. In this study, 93 participants of mostly Asian backgrounds (65%) and the rest European were evenly divided between undergraduate and postgraduate university learners. They used a multiple-choice listening instrument whereby participants listened to a short passage that stopped before the end of the message, and then chose an appropriate ending for the missing part of the passage. They operationalized the complexity of the targeted input with measures for phonology (i.e., contractions), lexical sophistication (i.e., diversity, density, and concreteness), syntax (e.g., number of modifiers per noun phrase), and discourse. They also measured speed (syllables per second) and explicitness (native speaker judgments made by the researchers). They concluded that lexical complexity made tasks more difficult. Eight of the measured task characteristics were found to be significantly associated with task difficulty, including the frequency of contractions in the target, multiword expressions (e.g., *rely on*, p. 156), argument overlap (i.e., fewer overlapping ideas within a target passage made the passage more difficult), adjacent argument overlap (i.e., adjacent sentences linked to the same argument made the passage easier), and stem overlap (i.e., the more shared word stems in a passage, the easier it was to comprehend). The study demonstrated the complexity of a speech stream, as new variables, such as lexical sophistication and density, were added to the multitude of factors to which listeners must attend.

Kostin (2003) looked at characteristics associated with TOEFL dialog listening items to determine what made them easier or more difficult. She found that dialogs that contained two or more negatives, dialogs that required inference beyond the content of the dialog, and the pattern of utterance (items that had a statement in the second utterance, for example, statement-statement as opposed to statement-question) were significantly more difficult. Kostin also found that infrequent vocabulary did not cause item difficulty, but she qualified this by adding that this was true only when the infrequent word was not relevant to a correct response. Although not the point of her study, more detailed analyses of the target dialogs, such as the analyses conducted by Révész and Brunfaut (2013), are missing from Kostin's study.

Mathews, O'Toole, and Chen (2017) examined the link between word recognition and proficiency levels for L2 learners. They began their research by recognizing the difficulty of L2 listening by describing the "transient nature of speech" (p. 22), which requires quick processing done without rate-control from the listener (i.e., without an automated system for controlling the speed of a speaker's utterance), and that differs from listener to listener depending on a set of variables that include listening skill and depth of phonological word knowledge. The researchers consequently designed a computerized, automated delivery system for spoken utterances with the aim of allowing participants to control the aural intake of target vocabulary items. They then analyzed the interaction of the participants and the application, the impact of the application on the participants' aural proficiency, and the word learning outcomes attained from the output. The results of their experiment showed that the participants used the application to revise their responses, suggesting interaction with the system. The results also revealed significant gains for those participants who had been graded as moderate (i.e., intermediate listening proficiency). These participants recognized academic words within aural targets. Among the researchers' conclusions was that low-proficiency listeners, in particular, need spoken English texts that contain a sufficiently high number of high-frequency vocabulary to ensure reasonable comprehension. Furthermore, aligning the target aural input and the listener might be done through an automated system in the future.

In the next section, I discuss comprehension breakdowns. I first discuss the *native listener hypothesis*, and that is followed by a discussion of other variables that interfere with comprehension.

Comprehension Breakdowns

In this section I present theory focused on comprehension breakdowns by foreign language learners. I begin with an overview of the *native listening hypothesis* (Cutler, 2015), which posits that a learner's L1 interferes phonologically with any subsequent language learning target. Next, I discuss other variables that confound spoken language comprehension for L2 learners, including *syllables* and *mora timing*.

The *native listener hypothesis* is a theory about the influences of learners' first languages on subsequent language learning (e.g., L2). Cutler (2015) has contended that a language learner's L1 phonological range dominates any phonological difference in a target language. As people learn a second or new language, they must often process sounds that do not fit into and are no longer part of their conditioned sound maps (Broersma & Cutler, 2011; Brown, 2000; Tuinman & Cutler, 2011) because these L2 sounds were not part of their language development as infants (Archila-Suerte, Zevin, Bunta, & Hernandez, 2012). For example, Japanese learners of English sometimes assimilate English /r/ and, more often, the /l/ into a single category, the Japanese alveolar flap [ɾ]. Whereas L1 learners have, as children, ample opportunity to develop L1 sound maps through aural input from their parents or teachers, L2 learners are not necessarily exposed to this type of input (Carroll, Swain, & Roberge, 1992; Ellis, 2008). Learning the phonology of a new language requires learners to adjust to the phonological system of that language.

Moreover, the target language can also create confusion for learners because of its phonological range. Phonological elements of the target language, for example, the type of timing of the language, can be conducive or adverse to learning. Limitations for Japanese L2 listeners of English include making phonological distinctions in isochrony (Bybee, Chakraborti, Jung, & Scheibman, 1998), such as interpreting the L2 from the phonological range of a *mora*-based L1 (i.e., the timing of a vowel or consonant-vowel pair in Japanese) or limitations such as word-level difficulties or dialectical changes (Baese-Berk, Bradlow, & Wright, 2013). The dynamic relationship between the phonological range of the L1 and L2 could be described as a *reciprical causation*, meaning that confusion runs in both directions (Bandura, 2001).

Furthermore, the ability of learners to adjust to an L2's aural makeup is not limited to the ability to understand the semantic information in the target audio. Some languages are stress-based languages (e.g., Indo-European-based), some are tonal languages (e.g., Afro-Asiatic), and some are both stress-based and tonal languages (e.g., Sino-Tibetan). In stress-based languages such as English, stressed and unstressed sounds often demark boundaries in the speech stream. In tone-based languages, tonal placement on words can change meaning. Learners moving from one language's phonological sound system to another need to come to terms with the differing phonological range of the L2, which is one of many subtle yet powerful contrasts between an L1 and the L2.

Syllables and Mora

Another language variable that theoretically causes comprehension breakdowns due to differences between a learner's L1 and L2 is the *syllable* (Cutler, 2002). Syllables are units used to model *prosody*, which is the way that individual sounds in spoken utterances are strung together linguistically (e.g., declaritive sentence) and/or expressively (e.g., anger) (Tatham & Morton, 2006). Prosodic effects, at the cognitive level, include intonation, rhythm, and stress, and these prosodic effects manipulate, at the basic level, syllables (Tatham & Morton, 2006). Syllables are often defined by their functionality within a particular language. In English, for example, syllables are considered *stress-timed*. Stress-timed languages, including English, are defined by the number of stresses in an utterance, which determines the length of the utterance (Celce-Murcia et al., 2010/2016). In other words, English utterances give more time to stressed syllables than to unstressed syllables. The following example shows the stress-timing of a sentence in English: *The hat is on the chair* [ðə ˈhæt əz ɑn ðə ˈʧɛr]. The content words *hat* and *chair* are stressed in this example, as indicated by the " ˈ " symbol that precedes them, so they receive the most time within the utterance. In contrast, the function words in the utterance, for example *the* and *on*, receive less time in the utterance and no stress. Function words carry less information and are used to show the grammatical relationships of the content words (Celce-Murcia et. al, 2010/2016) and they receive stress only when the grammatical relationship to a content word is important. The following question and answer example shows a stressed function word:

A: Which hat is on the chair?

B: This hat is on the chair.

[ˈðɪs hæt əz ɑn ðə ˈʧɛr]

The above example shows that the demonstrative pronoun *this* is stressed instead of the adjacent content word *hat* in order for the speaker to differentiate from two or more hats, as required by the word *which* in the question.

Two more sentence stress classifications of language are *syllable-timed* languages (e.g., French) and *mora-timed* languages (e.g., Japanese), though some researchers (e.g., Celce-Murcia et al., 2010/2016) include mora-timed languages as syllable-timed. In syllable-timed languages, the rhythm of an utterance is a function of the syllable count. In Japanese, each mora, which is the minimal unit of metrical time and counts as one syllable, receives equal time as the rest of the morae in the utterance. The Japanese language, which was the participants' L1 in this study, is classified as a mora-timed language, which is different from the stress-timed classification of English, the target language in this study. *Mora* is a term used to refer to a minimal unit of metrical time (Crystal, 1997). In other words, each minimal unit of metrical time, for example, the sound か [ˈkɑː] or あ [ˈɑː], is equal between morae, so the duration of each of the example mora is equal. In theory, Japanese learners treat English syllabic targets equally without regard for language idiosyncratic timing such as stress.

A common example mora is the Japanese pronunciation of the brand *McDonald's* (restaurant) [məkˈdɑnəldz]. In order to compensate for the required mora-timing of the language, the pronunciation in Japanese is マックドナルド (*makudonarudo*), which gives equal time to each mora. Using the English pronunciation can cause a breakdown in comprehension for Japanese learners. The brand is always pronounced *makudonarudo* (with variations of the word, shortened to *makudo* マックド or *makku* マック) in Japan, so the pronunciation is learned. However, when the name was originally transferred, the Japanese spelling compensated for the mora-timing of the Japanese language.

There is evidence that Japanese learners of English segment speech at mora boundaries (Warner, Otake, & Arai, 2010), which affects their processing of English (Cutler, 2002). A learner of English whose L1 is Japanese needs to be trained to discriminate English contrasts. In theory, in order for the contrast training to have context, the utilized contrasts should be known lexical items. By partially eliminating the factors that can cause breakdowns in communication, such as unknown lexis, learners might make discriminant errors that relate more closely to the acoustic signal.

L2 learners must negotiate a dynamic speech stream in order to comprehend it. Apart from obvious lexical and syntactic differences between an L1 and an L2, phonologogical features can also create challenges to comprehension. In this study, I focus on elision and assimilation as phonological features that influence L2 learners' comprehension; I do not address other phonological features that influence L2 learners' comprehension such as stress-timing versus mora-timing.

In the next section, I discuss language aptitude. I first review language aptitude for L2 learners, and then explore ways that second language aptitude can predict success for L2 learners.

Second Language Aptitude

In this section, I review second language aptitude, and outline studies on its relation to second language listening comprehension. I first provide a definition and an overview of language aptitude and second language learning aptitude. I then look at studies that have linked second language learning aptitude and second language listening comprehension. One instrument used in this study is the Modern Language Aptitude Test–Elementary (MLAT-E), which measures language learning aptitude, so a review of how language aptitude is situated in L2 research is beneficial to this study.

Language aptitude is a cognitive variable that pertains to individual learners' personal characteristics (e.g., the ability to hear rhymes) that help achievement only in the language domain. Carroll and Diller (1981) contended that language aptitude relates to the facility of learning (when one person's facility is compared to another's), and that it cannot be trained or affected by other language learning experiences. In theory, language aptitude should remain constant throughout learners' language learning experiences. Furthermore, studies of language aptitude have shown that language aptitude links to potential language learning success (Doughty, 2013).

For L2 learners, second language aptitude is generally defined as the ability to learn a second language relative to other learners and conditions. Second language aptitude has been shown to predict second language achievement (Li, 2016) and learning rate in instructed contexts (Granena & Long, 2013); it has also been argued to be significant when learners are untutored, such as when learning occurs in an extensive listening curriculum (Li, 2016). Moreover, Li (2016) suggested that second language aptitude is "a set of abilities of central importance in the preliminary stages of L2 development" (p. 804). The suggestion of preliminary stages hinges on the lack of studies of aptitude in advanced-proficiency L2 learners. Finally, second language aptitude, because of its effect on success in second language acquisition, has been a component in decisions to invest (e.g., money) in learning at all stages of L2 development (Doughty, 2014).

Li (2016) conducted a meta-analysis of second language aptitude research involving a population of 13,035 L2 learners. His study consisted of 66 prior studies of aptitude, including 34 studies that used the MLAT as an instrument. Li used the meta-analyses to investigate how aptitude is related to individual difference variables (e.g., motivation), working memory, and general L2 proficiency. Between-group effects were examined using a fixed-effects model, and within-group effects were examined using a random-effects model. He found that aptitude predicted L2 proficiency, L2 knowledge, and L2 skills, but not vocabulary learning. Moreover, he found that aptitude for phonetic coding was associated with general L2 proficiency, but it was unimportant for listening comprehension. Li's explanation was that meaning takes precedence over form for L2 learners, so the participants relied more on contextual clues and schematic knowledge to compensate for what Juffs and Harrington (2012) described as linguistic deficiency in nonnative listeners. L2 learners can rely on top-down processing more at earlier stages of learning, which suggests that bottom-up processing is weaker and therefore needs more attention in the earlier stages of L2 development.

Granena and Long (2012) used multiple instruments to assess aptitude and language domain, which included phonology, lexis and collocations, and morphology and syntax. In their study, 65 native speakers of Chinese with high proficiency in Spanish (L2), had a length of residency (LOR) in Spain for around 10 years, and ranged in age from 8 to 31 years. They divided the participants into three ages of onset (AO) groups, with the demarcation sensitive to periods of acquisition. The first group began learning Spanish between 3 and 6 year of age, as age 6 is thought to be "peak period of sensitivity for L1 and L2 acquisition" (p. 322). The second group of learners began between 7 and 15 years of age, with 15 marking the end of the sensitive period for morphology and syntax. The third group began learning Spanish between 16 and 29 years of age. Granena and Long found significant correlation between language aptitude and the third AO group (i.e., 16-29) for phonology, and lexis and collocations. They interpreted the significant results as better monitoring ability for phonology from the older group, and better working memories for lexis and collocations.

In the following section, I discuss background and length of residency. I also explore ways that background and length of residency are used in this study to complete a profile of the participants.

Background and Length of Residency

In this section, I introduce the idea of English-language background and length of residency, which includes the length of a homestay in a country where the L2 is the native language. I first define *background*, and second, I look at theories and studies in which researchers have elicited background information from participants to review different learner profiles in relation to their L2 acquisition. Third, I discuss length of residency studies.

Tracing the outcome of learners' L2 knowledge back to its precise origins is impossible. A better understanding of the role played by learners' experiences in the acquisition of an L2 is possible, however, and such information might shed light on improving practices for learners to acquire a second language. I use the term *background* in this study to identify the language learning experiences (e.g., number of native English speaking friends) that learners had while they learned English, and that might have influenced their listening proficiency. Moreover, the correlation between background and the ability to process phonological features of sandhi-variation can inform instructional input, as learners are conditioned by the phonological input they receive (e.g., how many hours they listen to music with English lyrics).

Trofimovich (2011) contended that L2 learning success is dependent on experience with the target language. However, the experience that learners have is connected to myriad variables. Among them are age of onset of learning the L2, first exposure to the L2 (e.g., positive vs. negative experience), and willingness to assimilate linguistically (e.g., practicing the pronunciation of the target L2). Measuring L2 experience is difficult because of the volume and variety of L2 experience. Parsing out learners' L2 backgrounds—as a subsection of their experience—is also challenging, as the choice of variables used to define the term *background* is always rudimentary compared to participants' actual background experience. Moyer (2011) claimed that background has, traditionally, been measured through the quantity of input (i.e., time on task) and LOR.

Linguistic factors that shape learners' L2 background into successful language acquisition include repetition of high-frequency syllables (Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997) and more familiar sound combinations (Frisch, Large, & Pisoni, 2000). Moreover, the frequency of a pattern of phonological input determines perception and production for L2 learners (Trofimovich, 2011). That is to say, there is a possibility that larger amounts of listening enable phonological sensitivity to L2 forms (i.e., patterns), which then leads to benefits for L2 listeners. Larger amounts of listening provide opportunities for *cognitive* and *perceptual tuning* (Segalowitz, Gatbonton, & Trofimovich, 2009), which describe the engagement of learners with the language, and are based on the frequency of the target within the L2 and the similarity between the learners' L1 and the target L2.

Another challenge directly linking background and phonological effects, such as elision, is the *Matthew effect*, which is the idea that an initial advantage in, for instance, a cognitive ability, is amplified and results in differences that widen the preexisting gaps (Ceci & Papierno, 2005). Trofimovich (2011) described the effect as "self-reinforcing bidirectional processes" (p. 150) and argued for including L2 phonological learning and L2 experience as possibly interlocked, which could describe initial advantages, gained through, for instance, L2 pronunciation. Initial advantages learners might have with the ability to pronounce the target L2 give them an advantage later over those learners who do not have that skill. In theory, any variable has the potential to affect each learner with an advantage or a disadvantage, and learners' backgrounds might signal these variables, but this reciprocal causation creates challenges for researchers.

Moyer (2011) investigated the impact of quantity and quality of experience on L2 phonology for long-term attainment. For quantity, she looked at time on task, years of instruction, and LOR in the target L2 country. For quality she looked at the degree to which the participants interacted in ways that were functionally significant (i.e., with a specific language purpose) with interactive language use (e.g., listening to a native speaker). Finally, she looked at how these variables predicted phonological attainment among 50 English-language learners. Items used to measure background were delivered through a questionnaire, which included questions about time on task using the target L2. Multiple regression models were used to predict factors of AO, LOR, and time with native-speaking friends—specifically, L2 experience and degree of foreign accent, L2 experience and quantity versus quality, and L2 experience as an independent predictive value for phonological attainment. She found that the ages of onset and instruction, as well as immersion, significantly predicted the acquisition of L2 phonology. Moyer concluded that experience was an independent influence on phonological development, and that due to immediate feedback, learners can refine phonological structures. Piske, Mackay, and Flege (2001) described LOR as a "rough index" because it does not equate to input quality. Furthermore, Moyer (2011) stated that LOR is inconsistent; thus, she suggested that the LOR needs to be elaborated to clarify it as a variable. Flege and Liu (2001) also described LOR research results as ambiguous. However, Trofimovich and Baker (2006) have shown LOR to predict accuracy with the production of segmental and suprasegmental patterns.

Saito (in press) has given an overview of advanced L2 learners regarding segmental and suprasegmental acquisition in relation to process (i.e. L2 immersion environment), product (i.e., relative to native speakers), and individual differences (e.g., aptitude). He suggested (in accordance with Cook, 2002) that advanced L2 learners are better experimental baseline models than native speakers because the language acquisition for native speakers is nativelike across all linguistic domains, whereas L2 acquisition is built on an L1 foundation and therefore cannot be equated to a native speaker's L1 system. Saito went on to describe a transition period that happens as experience increases due to the need to accurately comprehend and produce the L2. A "phonetic-level restructuring" (p. 3) occurs where learners move from detecting lexical units to detecting new L2 sounds at the phonetic level.

Saito (2015) researched the effects of short (8 months to 1 year), mid (1 to 5 years), and long (5 to 13 years) LOR on late L2 learners. In his study, 39 Japanese learners ranging from 21 to 43 years old were selected using the following criteria:

More than 8 months LOR, residency after the age of 19, and frequent use of the L2 (English). He used a speaking task with picture cues to elicit spontaneous speech from the participants and then rated their use of segmentals and syllable structures that were considered difficult (e.g., the /r/-/l/ contrast) based on the participants' Japanese L1. He then investigated breakpoints in the LOR-proficiency function of the linear models, including word stress, intonation, speech rate, lexical appropriateness, and grammatical accuracy as dependent variables. Saito found that LOR was especially significant for the first three years and suprasegmental performance improved for the first five years of L2 experience. This study was not connected directly to processing, but it highlights LOR in second language research. In the next section, I review three gaps in the literature addressed in the current study.

Gaps in the Literature

This study is designed to address three gaps in the second language listening comprehension literature. The first gap is the need to better understanding L2 learners' proficiency problems as they decode a speech stream. The second gap is the need to better understand L2 learners processing abilities when dealing with sandhi-variation in aural texts. The third gap is the need to make connections between L2 learners' backgrounds and abilities to process variation in speech.

The first gap in the second language listening literature addressed in this study is the lack of research available on the possible existence of an item difficulty hierarchy of listening items that include the phonological effects of elision and assimilation. Listening comprehension requires the parsing of a speech stream, but confounding variables such as elided words complicate a speech stream in ways that affect parsing for L2 listeners but not L1 listeners (Henrichsen, 1984). Variables embedded in a speech stream (e.g., elision and assimilation) have not been explored thoroughly in second language research, and few researchers have explored the effects of aural processing using elided and assimilated targets for the purpose of providing insight into the sandhi-variation phenomena and L2 spoken language processing.

The second gap in the second language listening literature addressed in this study is the lack of research on aural processing and aural input with the phonological features of elision and assimilation. Comprehension breakdowns occur as learners negotiate lexical load, utterance rate, and length of the target utterance, among other factors. Breakdowns can be the result of one or more variable aspects of a speech stream. Sandhi-variation is a characteristic of most spoken English; it varies in degree and it is associated with an inability to comprehend content (Ahn, 1987; Rost, 2016) or with distractions caused by the content (Field, 2003). However, research exploring how L2 learners process spoken English with sandhi-variation has rarely been addressed. There is a lack of understanding about aspects of aural input that is challenging for L2 learners and how different challenges, such as elided or assimilated speech, affect aural processing.

The third gap in the second language listening literature addressed in this study concerns the paucity of research investigating learners' backgrounds and length of residency and their ability to process the acoustic signal, which includes sandhi-variation. No empirical studies have linked the type of L2 input that learners have received with the act of processing a speech stream. Furthermore, although there is a growing body of research on the relationship between learners' L2 language learning backgrounds and their performance on standardized English assessment tests, currently no researchers have explored the relationship between language learners' backgrounds and their ability to process elision and assimilation. Learners' experiences with an L2, for example, during a homestay, might modify their ability to comprehend certain acoustic signals more than others. A learner's historical input might help or hinder their comprehension of current aural input. Research exploring this connection is missing from the SLA literature.

Purposes of the Study

The first purpose of this study is to investigate the item difficulty hierarchy of elision and assimilation that learners have in relation to L2 listening proficiency. By testing the participants' comprehension of utterances with elision and assimilation, the item difficulty hierarchy can shed light on the problems L2 listeners encounter while decoding the utterances. Furthermore, the person ability estimates on the Elicited Imitation Test might correlate with the Test of English as a Foreign Language Paper-based Test (TOEFL PBT), the Listening Vocabulary Levels Test (LVLT), and the Modern Language Aptitude Test–Elementary (MLAT-E). Identifying an item difficulty hierarchy of elision and assimilation that learners follow as they develop L2 listening proficiency, and with their language aptitude, would improve researchers' and educators' current understanding of L2 learners' listening comprehension, which could lead to more useful graded L2 aural instructional input.

The second purpose is to evaluate links between second language aural processing and aural input containing elision and assimilation. Evaluating these links will provide insight into how learners deal with sandhi-variation as they process a speech stream. Prior studies of sandhi-variation and aural processing have provided a basic understanding of how processing is applied by second language listeners (Ahn, 1987; Henrichsen, 1984), and, more recently, how the brain handles the processing of utterances (Ding, Melloni, Zhang, Tian, & Poeppel, 2016). Gauging the processing ability of individual second language listeners provides additional research findings, provides instructors with insight into the needs of second language listeners, and might inform classroom content, such as aural-text type used in listening lessons.

The third purpose is to investigate the participants' historical aural input in order to explain which types of input facilitate listening comprehension. Interview data are used to reveal variations in participants' historical background—their experiences with native speakers—with English. Next, I cross-reference the participants' abilities to process language containing sandhi-variation with their background, to show the relationship between historical input type and the ability to parse utterances with the phonological features of elision and assimilation. The aural experiences that participants have historically encountered relate to participants' abilities to process sandhi-variation, and the correlation can suggest the type of input that facilitates second language aural processing.

Research Questions

This study is guided by four research questions:

1a. What is the empirical item hierarchy for the phonological features of elision, and assimilation? What are the determinants of this hierarchy?

1b. How do two input rates and phrase lengths influence the participants' comprehension of the phonological features of elision and assimilation?

2. What is the strength of the relationship between the participants' language proficiency, as established by the TOEFL PBT, the LVLT, and the MLAT-E, and the comprehension of the phonological features of elision and assimilation?

3. What is the strength of the relationship between participants' English-learning backgrounds and their comprehension of the phonological features of elision and assimilation?

Based on my understanding of how L2 learners comprehend speech streams in English, I have formulated hypotheses for the research questions. I have three overarching hypotheses for this study, two hypotheses about the interactions between elision and assimilation and the participants' abilities and aptitudes, and one hypothesis about the participants' backgrounds and length of residency abroad and their processing of elision and assimilation.

Hypothesis 1a: Items with the phonological features of assimilation and elision will have a greater negative effect on comprehension than items without those features.

Hypothesis 1b: Items with 8 syllables will have a greater negative effect on comprehension than items with 6 syllables. Items with a higher utterance rate will affect participant comprehension negatively compared to items with a lower utterance rate.

Hypothesis 2: Comprehension of items with elision, assimilation, higher utterance rate, and 8 syllables will have a positive correlation with proficiency, lexical knowledge, and aptitude as measured by the TOEFL PBT, the LVLT, and the MLAT-E, respectively.

Hypothesis 3: Learner backgrounds as established by the questionnaire and interview data will have a positive correlation with comprehension of items with elision and assimilation.

In this chapter I reviewed the relevant literature concerning sandhi-variation and comprehension of spoken English for Japanese learners. The review was divided into four initial sections: Sandhi-Variation, Spoken Language Processing in English, Language Learning Aptitude, and Background and Length of Residency. The Sandhi-Variation section served as background for the topic and covered sandhi-variation in English, in Classical Phonetics, L2 learning, and in current linguistic theory. The Spoken Language Processing in English section contained language processing theory, L2 spoken language processing theory, reasons for comprehension breakdowns of L2 learners, and an explanation of syllables and mora. The Second Language Aptitude section concerned the connection between aptitude and L2 acquisition, especially phonological processing. The Background and Length of Residency section covered background, and length of residency and L2 acquisition. I concluded the chapter by considering the gaps in the literature, my purpose in doing the study, and finally my research questions and hypotheses. In the next chapter, I describe the methodology of the current study.

CHAPTER 3

METHODS

In this chapter I describe the methods that I have used in the study. I first describe the educational setting and the participants of the study. I then describe the instrumentation, which includes (a) the Test of English as a Foreign Language Paper-Based Test (TOEFL PBT), (b) the Listening Vocabulary Levels Test (LVLT), (c) the Modern Language Aptitude Test–Elementary (MLAT-E), (d) the Pre-Listening in English questionnaire, (e) the Elicited Imitation Test (EIT), and (f) the Background and Length of Residency interview. Next, I describe the research design, procedures, and analyses. Finally, I describe the Rasch model, including the Rasch fit statistics, the principle component analysis of item residuals, and the partial-credit model.

Educational Setting

In this section, I describe the educational setting of the study, and include brief overviews of the participants in the study, the institution where the study took place, and the program in which the participants studied. The setting for the study was a private, coeducational university located in Yamanashi Japan, specifically an intensive English language program that was part of a four-year, liberal arts curriculum. Majors offered at the school were sociology, philosophy, linguistics, economics, art, music, and Japan studies. The students in the intensive English language program were expected to graduate from the program and go on to take a variety of liberal arts classes, taught in English. After one year of liberal arts courses in Japan, students were expected to study abroad. The intensive English language program of which the participants were members is an academic program that was established for the purpose of preparing them to enter the liberal arts courses and go abroad for academic purposes.

The liberal arts college had 90 students enrolled in total at the time of the study. Of the 90 students, 32 were in the intensive English program, and some of these became participants. The students in the English language program came from various regions of Japan. In the liberal arts program, the approximately 45 international students were from seven countries, which included Croatia, Ireland, France, the Netherlands, Norway, Scotland, and the United States. Students were present for the majority of the school year. Because of the relatively small size of this liberal arts college, its de facto separation from the other faculties within the university, the requirement that all students live in a dormitory system connected to the school, and the balanced, 50:40 ratio of Japanese students to international students, there was an international atmosphere at the school. Furthermore, English was the main language spoken.

Students in the intensive English language program who were targeted for participation in the current study were divided into three streams, upon entering the program, and placed into those streams based on their TOEFL PBT scores. Students scoring 459 points and lower on the TOEFL PBT were placed in the Level 1 stream (beginner to low-intermediate), those scoring between 460 and 479 points were placed in the Level 2 stream (intermediate), and those scoring between 480 and 499 points were placed in the Level 3 stream (intermediate to high-intermediate).

Students in the intensive English program were required to take five academic skills-based classes that met twice per week: writing, reading, listening and speaking, testing practice, and content-based practice. Furthermore, students were required to attend an autonomous learning language laboratory five hours per week. Each class lasted 1.25 hours and met twice a week. Including the language laboratory, students had a total of 20 hours of class time per week. Approximately 2 hours of homework were assigned per skills-based class, which gave the students around 20 hours of homework per week. I selected 25 students from this program for participation in the current study. In the next section I will describe in more detail both these participants and my selection criteria.

Participants

Twenty-five first- and second-year Japanese university students participated in the current study. The participants' ages ranged from 18 to 20 (M = 18.8). They were selected using the following criteria: (a) They had a TOEFL PBT score (as designated by the institutional class distribution constraints), and (b) they were English as a foreign language (EFL) students. Experience abroad was not a requirement for participation.

The participants had 6 to 10 years of English education, with the majority having a minimum of six years of formal English education in secondary school (junior high and high school). A few of the participants received extra schooling in the form of English conversation schools (*eikaiwa*) and cram schools (*juku*).

All 25 participants were taking university level, intensive, academic English courses that were focused primarily on English skills (e.g., academic writing), content in English (e.g., current events), and linguistics (e.g., grammar). None of the participants had decided a major as they were all in the intensive language program, a requirement prior to moving into liberal arts classes (i.e., with a major concentration).

A summary of data from a pre-listening in English questionnaire is shown in Table 1. The left column shows the four questions on the questionnaire. The middle column summarizes the participants' answers. The right column shows the percentage of participants whose answers were represented by the summary. All of the participants had been learning English for at least six years and they all saw English ability as a necessity in their lives. Furthermore, all of them used English-based media, and music was the most common example of media type.

Table 1. *Pre-Listening in English Questionnaire*

Question	Summary of answers	% of answers represented by summary
1. How do you feel about English, in general?	Necessary (e.g., career, communication)	92%
2. How do you feel when you listen to English?	Happy/excited; hearing something different; want to speak like a fluent speaker	90%
3. How do you feel about the English sound system (e.g., pronunciation)?	Different; interesting; need to practice the sound system	96%
4. What is your history of studying English? (How many years have you been studying English? Do you listen to/watch English media? Have you studied abroad in any way? etc.)	6-10 years; music/movies/YouTube; some travel to English speaking countries	88%

In the next section, I describe the six instruments used in the study: the TOEFL PBT, the LVLT, the MLAT-E, the Pre-Listening in English questionnaire, the EIT, and Background and Length of Residency interview.

Instrumentation

In this section, I describe the six instruments used in the current study. I begin with an overview of the instruments, and then describe each instrument in more detail and in the order that they were administered.

Six instruments were used in the study: (a) the TOEFL PBT, (b) the LVLT, (c) the MLAT–E, (d) the Pre-Listening in English questionnaire, (e) EIT, and (f) the Background and Length of Residency interview. Each of the instruments was used to determine a different aspect of the participants' proficiencies. Participants took the first three proficiency tests antecedent to the last three. The TOEFL PBT was used as a measure of general English proficiency and listening proficiency. The LVLT was used as a measure of aural vocabulary knowledge. The MLAT-E was used as a measure of phonemic proficiency. These tests set baseline scores for the participants' general English proficiency (and listening proficiency), listening vocabulary size, and phonemic cognition. Next, the Pre-Listening in English questionnaire was used for the purpose of establishing participant attitudes to listening to English in general and to predict performance on the proficiency tests. The EIT was used as a measure of participants' ability to process elision and assimilation. Finally, the Background and Length of Residency interview was used to gain qualitative information about the participants' experiences with English. The following sections contain descriptions of each instrument in detail, starting with the TOEFL PBT.

Test of English as a Foreign Language Paper-Based Test

The TOEFL PBT is divided into three sections: Listening, Structure and Written Expression, and Reading. The Listening Comprehension section contains 50 multiple-choice questions; the Structure and Written Expression section contains 40 multiple-choice questions; and the Reading Comprehension section contains 50 multiple-choice questions. The Listening section scores for the participants were used as a measure of listening comprehension. An example test item from the listening section of the TOEFL PBT is shown below.

A test taker hears:

(Man) Shall I lock up the computer lab now before I go home?

(Woman) Don't bother. I'm not leaving for a while; I can check it on my way out.

(Narrator) What will the woman probably do?

The test taker reads:

A. Lock the computer lab later.

B. Leave with the man.

C. Buy a new lock for the computer lab.

D. Show the man where the lab is. (ets.org, 2016)

Participants marked their answers by filling in a bubble sheet. The answer sheets were collected according to the procedure required by the Educational Testing Service (ETS), and sent to the ETS. All TOEFL PBT data used were official scores sent from ETS. The TOEFL PBT results were used as an English-language proficiency variable and a listening proficiency variable. In the next section, I discuss the LVLT.

Listening Vocabulary Levels Test

The LVLT (McLean, Kramer, & Beglar, 2015) was designed to assess knowledge of the first five 1,000-word frequency levels and the Academic Word List (Coxhead, 2000) (Appendix A). An example test item from the LVLT, translated from Japanese, is as follows:

A test taker hears:

(Man) Waited. I waited for a bus.

The test taker reads in Japanese (L1):

A. 食べた (translation: *ate*)

B. 待った (translation: *waited*)

C. 見た (translation: *saw*)

D. 寝た (translation: *slept*)

Participant LVLT raw scores for both Part 1, the first 1,000 high-frequency words, and Part 2, the second 1,000 high-frequency words, averaged 23 out of 24 possible answers, in other words 96% aural coverage of the vocabulary for the first and second 1,000 words (see Appendix A). The percentage complies with the minimum percentage of successful answer recommended by McLean, Kramer, and Beglar (2015) for the aural receptive vocabulary level of the non-native participants. The results of the LVLT were used to assess the participants' readiness to handle the lexical load of the aural targets. Utterances for imitation in an elicited imitation instrument used in this study were written using the first 2,000 high frequency words of English; thus, the purpose was to eliminate the participants' aural vocabulary knowledge as a reason for incorrect responses. McLean et al. (2015) set comprehension of aural texts made up of the first 2,000 high frequency words of English at 45-46 correct answers out of 50 items. This benchmark of 45-46 was used in this study as the minimum cut-off score for participation. The first 3,000 high frequency words were tested in order to avoid a ceiling effect. In the next section, I discuss the MLAT-E.

Modern Language Aptitude Test–Elementary

The results of the MLAT-E (Carroll & Sapon, 2010) were used to assess the participants' language aptitude (i.e., phonemic coding). Phonemic coding ability concerns the effective auditory processing of input (Skehan, 2002) and allows the participants to analyze and code auditory material for the purpose of online retention. Chan, Skehan, and Gong (2011) defined phonemic coding ability as “the ability to analyze sound so that it can be retained for more than a few seconds” (p. 56). According to Skehan (2002), MLAT-E scores can estimate aptitude for foreign language learning.

One disadvantage to using the MLAT-E was that the directions were given in the L2 instead of the participants’ L1, which presented a possible confound to the participants’ L2 proficiency (Doughty, 2014). However, instructions and tasks on the MLAT-E were worded simply, as they were designed for elementary school-aged learners. Instructions were broken into parts and practice examples were given for each part. For example, the Matching Words section of the MLAT-E had 25 examples and eight practice tasks in the instructions before the participants attempted the actual task. In addition, prior to each section, the proctor checked with the participants to be sure that they had understood the tasks.

The MLAT-E consisted of four parts designed to assess the test-takers’ ability to associate sounds and symbols, sensitivity to grammatical structure, ability to hear speech sounds, auditory alertness, and ability to remember. Participants’ raw scores were used. Part 1 of the test was labeled Hidden Words and was designed to measure English vocabulary knowledge and sound-symbol association ability. An example of a test item from the Hidden Words section of the test was as follows:

rivr	☐ large stream of water	☐ a part of the body
	☐ hill	☐ a dog’s name

For this item, the participants chose and marked the box next to the word or group of words that mean the same as the word representation on the left. For the above example, the correct answer is large stream of water, as it is a definition of river (rivr). There were 30 items in this part of the test. Part 2 of the test was labeled Matching Words, and was designed to test the participants' understanding of target word usage in example sentences. An example Matching Words test item is as follows:

Henry THREW the heavy stone.

Sally rides a bicycle.
☐ ☐ ☐ ☐

For this item, participants chose and marked the box under the word in the sentence that had the same function as the word in all capital letters in the sentence directly above it. For the above example, the correct answer was *rides* because it is a transitive verb, just as *threw* is a transitive verb. There were 30 items in this part of the test. Part 3 of the test was labeled Finding Rhymes and was designed to test the participants' ability to match speech sounds by selecting words that rhyme. An example of an item was as follows:

DOOR................ ☐ car............. ☐ four.......... ☐ mayor......... ☐ our

For this item, participants chose and marked the box next to the word that rhymes with the word on the left. For the above example, the correct answer is *four*, as it rhymes with *door*. This type of question tested the participants' familiarity with the English sound system. There were 45 items in this part of the test. Part 4 of the test was labeled Number Learning, and asked the participants to learn nonce names for numbers and then recognize the nonce names aurally. The participants wrote the nonce numbers they heard. An example of an item on the test is *chall* [ʧɑl] (spelling and transcription is mine), which was taught as equal to the number one. Participants heard the word, and wrote the number that the word represented. There were six base numbers, 1, 2, 3, 10, 20, and 30. There were 25 items in this part of the test, and all but four numbers, 3, 30, 1, 31, and 10, were repeated twice. In the next section, I discuss the Pre-Listening in English questionnaire.

Pre-Listening in English Questionnaire

The four questions on the Pre-Listening in English questionnaire were designed to gather qualitative information from the participants regarding their understanding of the English sound system as something different than their L1 sound system. I designed the questions to illuminate the complexity of understanding that the participants have. The questionnaire worked also as a precursor to a more extensive interview that followed the EIT.

The participants completed the Pre-Listening in English questionnaire in their L1 about their general feelings toward English, the English sound system, and their history with English (Appendix B). The Pre-Listening in English questionnaire had four questions:

1. How do you feel about English, in general?
2. How do you feel when you listen to English?
3. How do you feel about the English sound system (e.g., pronunciation)?
4. What is your history of studying English? (How many years have you been studying English? Have you studied abroad? Do you listen to English songs?)

A Japanese language instructor, teaching at the university level, who had native-level English proficiency, translated the questionnaire from English to Japanese. A verbal protocol was used while translating in order to verify meaning of the English as it was translated. Two bilingual university instructors then back-translated the form to check for English meaning. All of the items were confirmed for the intended meanings. The

Japanese translation for the Pre-Listening in English questionnaire was as follows:

1. 一般的に、英語についてどう思いますか？
2. あなたが英語を聞くとき、どのように感じていますか？
3. 英語のサウンドシステム（例えば、発音, イントネーション等）につ い てどう思いますか？
4. これまでどのように英語を勉強してきましたか（何年間英語を勉強してきましたか？何らかの理由で英語で授業が行われる海外の学校で勉強したことがありますか？あなたは英語の歌を聴いていますか？など）

In the next section, I discuss the EIT.

Elicited Imitation Test

Elicited imitation is a testing method that usually requires test-takers to listen to prompts—including sounds, phrases, or sentences—and then repeat the prompts verbatim. The elicited imitation testing method has been widely used in L2 research. Yan, Maeda, Lv, and Ginther (2015) cited 76 articles of theoretical and empirical studies using elicited imitation between 1970 and 2014. A database search of elicited imitation studies between 2014 and 2018 located 12 journal articles in refereed, L2 journals. Researchers that used elicited imitation in L2 research include Bley-Vroman and Chaudron (1994); Chaudron, Nguyen, and Prior (2005); Erlam (2006); Graham, McGhee, and Millard (2010); Spada, Shiu, and Tomita (2015); Suzuki and DeKeyser (2015); Tracy-Ventura, McManus, Norris, and Ortega (2014); and Wu and Ortega (2013). Galliard and Tremblay (2016) argued that elicited imitation tasks could be used to test L2 proficiency. The underlying assumption is that when participants can repeat the target prompt verbatim, they have acquired the language (e.g., phonological) features of the item or the feature associated with the item (Munnich, Flynn, & Martohardjono, 1994; Underhill, 1987). Repetition of an item is difficult if one has not acquired the phonological features (Rebuschat & Mackey, 2013). The advantage of using an elicited imitation test in the current study is that the participants can clearly demonstrate understanding of the speech stream variation, elision and assimilation. Moreover, these types of tests are not prompted with written forms of language other than in the instruction section prior to the actual items, and the participants did not write at any point during the test; thus, it is a pure listening test.

The sandhi-variation version of the Elicited Imitation Test (EIT) used in this study (Appendix C) was designed to measure the participants' abilities to parse the speech stream when elision and assimilation were present. The test used in this study had two stages. In the first stage, the test elicited elision and assimilation imitations of the initial prompt. This type of paraphrasing on elicited imitation tasks has been shown to require the participants to access implicit knowledge, which is evidence of construct-related validity (Yan et al., 2015). In the second stage, the test elicited an *enunciation version* of the initial prompt, where each word needed to be enunciated in its full form. For example, if the initial prompt contained the utterance *I don't like milk* [aɪ doʊnt˭ laɪk mɪlk], the participant would repeat the phrase in the first stage, [aɪ doʊnt˭ laɪk mɪlk], and enunciate each word in the second stage: *I do not like milk* [aɪ dʊ nɑt laɪk mɪlk]. The phoneme t is unreleased in this example, as indicated by the " ˭ " symbol that follows it, so it is spoken as an unaspirated stop. The second stage requirement of enunciating the target was an original test design for this study and was designed to signal recognition of sandhi-variation in the items by the participants.

Elicited imitation was chosen to assess elision and assimilation proficiency because it was mandatory that the test content be controlled in order to observe links to specific phonological aspects of the speech stream that characterize the English language. More precisely, it was a requirement to have test items that elicit the attributes of elision and assimilation to test participants on their mastery of these important linguistic properties, in relation to the phoneme coding and the participants' listening proficiency. It was possible to focus on such attributes because the elicited imitation test was composed of decontextualized sentences.

The EIT consisted of 48 sentences, using 24 elision targets and 24 assimilation targets. Elision occurs when sounds are dropped in spoken English, for example, *how did it go* [haʊ dɪd ɪt goʊ] can be pronounced *how dit go* [haʊdɪt$^{=}$ goʊ]. Elision is more technically described as when a phoneme is realized as zero (Roach, 2009), or when a phoneme is not realized. Assimilation occurs when adjacent sounds influence each other, especially at word boundaries. For example, *that person* [ðæt$^{=}$ pɜrsən] is often pronounced *tha person* [ðæ pɜrsən] or *thapperson* [ðæppɜrsən] because the final consonant of the first word is affected by the consonant that comes after it (i.e., regressive assimilation). I created the test items, and the recording was of my voice.

Two more aspects of the EIT, other than elision and assimilation, were syllable and speed variation, and these aspects were blended evenly into the 48 total items. First, each item on the EIT was either 6 syllables long (24 items) or 8 syllables long (24 items), not accounting for pauses between words. Elicited imitation targets that vary in length have been shown to be more sensitive when used to measures of different L2 proficiency levels (Yan et al., 2015). One study in elicited imitation that used varying utterance lengths claimed that up to 15 syllables per utterance is an acceptable prompt length for adult L2 learners with high proficiency (Munnich et al., 1994). However, Perkins, Brutten, and Angelis (1986) found that between seven and eight syllables is acceptable for high proficiency second language speakers. To be conservative due to the elision and assimilation that added another dimension of difficulty to both the target utterances and to the speaking proficiency of the participants (i.e., having to imitate the targets), utterances that varied between six to eight syllables were used in this study. Furthermore, the sandhi-variation form defined the syllable count when it was present in the sentence. Therefore, the item *Howzit going today?* was defined as having six syllables, whereas the fully enunciated form (e.g., *How is it going today?*) had seven syllables. An example of a test item with 8 syllables and produced as an audio target without elision, with each word in the target item said in complete form, was the following: *coffee is a waste of money* [kɑfi ɪz ə weɪst ʌv mʌni]. Syllable count was chosen over word count as a more precise way of determining the length of an utterance, as word length varies more than syllable length. Second, there were two different utterance rates (speeds) for items. Faster items replicated L2 native speaking rate (24 items), and slower items replicated average lecturer speaking rate for lectures given to non-native speakers (24 items). The added time pressure structure of this test was chosen to replicate authentic utterances of L1 native speaking rate. Average lecturer speaking rate for lectures given to non-native speakers is 140 words per minute or 3.17 syllables per second. This lecture rate for non-natives is slower than the average conversational speech rate for native English speakers of around 4.7 syllables per second (Tauroz & Allison, 1990). The sentence speed was increased to the native English speaker conversation rate of 4.7 syllables per second, using the edit speed function on the free software Audacity®: Free Audio Editor and Recorder, Version 2.1. This insured a consistency of the speed of all utterances under this condition.

To summarize the item types, the EIT consisted of 48 items. Half (24 items) were elision items that could be elided and half (24 items) were assimilation items that could be assimilated. Of the elision items, half (12 items) were spoken with elision, and half (12 items) were spoken with the possibility of elision ignored. Similarly, of the assimilation items, half (12 items) were spoken with assimilation, and half (12 items) were spoken with the possibility of assimilation ignored. Only words spoken with elision or assimilation were linked; otherwise, words were uttered as distinctly individual, with pauses between each word. Moreover, for each 12-item set (e.g., 12 items with elision that was pronounced in the item), 6 were 6 syllables in length and 6 were 8 syllables in length. For each of the 6-item set (e.g., 6 syllable-length), 3 items were fast (native speaker speed) and 3 items were slow (ELT lecturer speed). All items were scored the same, which means that two items, for example, one with elision realized and one with elision not realized in utterances, were scored with the same method, which is to say, dependent upon a participant's two-stage answer. Finally, the test items were randomized using online randomizing software (www.random.org).

The EIT was recorded with an Olympus Voice-Trek DS-60. The recorded WMA files were transferred to an Apple MacBook Air. The recorded items were then manipulated into a single recording, with pauses and tones synced, using Audacity®: Free Audio Editor and Recorder software, Version 2.1. The voice on the recording was mine (male, USA, native English speaker). My voice was familiar to all of the participants, which I viewed as a way of avoiding the voice in the EIT being a distraction. Trofimovich (2011) argued that voice can distract L2 listeners and cause misunderstandings, even when the content of the utterance is known.

The instructions on the EIT were written in the participants' L1, Japanese, to ensure that they understood the task. To make sure that the instructions were clear, the instructions were piloted with four students of roughly the same proficiency level as the participants; these students did not participate in the study. They showed their comprehension by explaining the procedures in their own words to a Japanese language instructor with advanced English proficiency (TOEFL 600). The test-takers might or might not have been familiar with elicited imitation as a test type; thus, providing an adequate number of practice items was necessary to insure that the participants understood the task. The test instructions contained five example types, which were chosen not to introduce elision and assimilation, but to prime the participants for Stage 1 and Stage 2 of the EIT. The examples were written in their complete form on the instruction sheet. Participants read the instructions prior to hearing the same instructions and examples represented in the written text. The instruction sheet contained the following written instructions in both English and Japanese and examples that were also heard by the participants prior to the test.

Elicited Imitation Test Instructions

口頭模倣テスト答え方向

This is an elicited imitation test.

これは口頭模倣（聞こえてきた語句や文を復唱し、次に完全な形に直してもう一度言葉に発する）テストです。

You will hear complete sentences.

初めに文が話されます。

First, repeat each sentence exactly as you hear it.

まず、聞こえてきたフレーズや文を復唱してください。

Second, say the sentence a second time, pronouncing each word in its complete form.

次にもう一度その語句や文を言ってください。その際、あなたが聞いた語句や文が短縮系であれば完全な形に直して言ってください。

Your responses will be recorded.

あなたの回答は録音されます。

Examples:

You hear:	I *don't* like milk.
You repeat:	I *don't* like milk.
<u>After the tone...</u>	[tone is heard]
You then say:	I . *do* . *not* . like . milk.
You hear:	*Watcha* doing?
You repeat:	*Watcha* doing?
<u>After the tone...</u>	[tone is heard]
You then say:	*What* . *are* . *you* . doing?
You hear:	Are *ya* happy?
You repeat:	Are *ya* happy?
<u>After the tone...</u>	[tone is heard]
You then say:	Are . *you* . happy?
You hear:	*Wouldjou* like a drink?
You repeat:	*Wouldjou* like a drink?
<u>After the tone...</u>	[tone is heard]
You then say:	*Would* . *you* . like . a . drink?
You hear:	It's a *pinkcar*.
You repeat:	It's a *pinkcar*.
<u>After the tone...</u>	[tone is heard]
You then say:	It . is . a . *pink* . *car*.

Participants could read along as they heard the examples aurally, which had a second benefit of introducing the voice used on the instrument. The first two examples contained two parts. Part 1 was a modeled explanation about how to repeat. Part 2 contained a signal tone, then a model of how to enunciate the target sentence The third, fourth, and fifth examples had a 14-second silences after each target prompt, which replicated the items (i.e., including the silences) on the actual test. After an initial 6-second silence of the 14-second total, a 0.5 second tone could be heard. The tone signaled the division of the allotted 14 seconds into the two stages of the answer, repetition (stage 1) and enunciation (stage 2). After the tone, 8 seconds were provided for the enunciation portion, as piloting showed that the enunciation section required 2 seconds more than the repetition section to complete. After the final example was given, the participants signaled understanding or lack of understanding of the test procedure prior to the beginning of the target items.

Table 2 shows the items used on the EIT. The items on the elicited imitation test were made up of vocabulary from the high-frequency 2,000 words of English according to the British National Corpus (BNC)/Corpus of Contemporary American English (COCA) list. Also, each sentence was a complete, simple sentence. There were no compound or complex sentences. Two university teachers checked the items. Both teachers held PhDs in linguistics, and were familiar with transcription. For target sentences 2, 3, 7, 10, 12, 13, 16, 23, and 24, I use the term assimilation instead of the term *linking*. As stated in Chapter 2 of this study, assimilation refers to the influence or influences that phonological segments have on one another, and according to Tatham and Morton (2006), assimilation is “voluntary or optional” (p. 23). It occurs prior to an utterance, and the speaker can manipulate the target utterance. Linking, similarly, describes the relationship at junctures between sounds (Roach, 2009), but it is a term that many L2 and L1 listening researchers, such as Rost (2016), Cutler (2015), and Tatham and Morton (2006), exclude in their studies in lieu of the term

Table 2. *Target Sentences Used on the Elicited Imitation Test.*

Item	Target sentence	IPA transcription	Type of Sandhi-variation	Description of the Sandhi-Variation
1	Coffee is a waste of money.	kɑfi ɪz ə weɪst ʌv mʌni	−Elision	None
2	He disappeared in snow.	hi ˌdɪsəpɪrd ɪ snoʊ	+Assimilation	Final *n* segment assimilates to alveolar fricative *s* & second alveolar *n*
3	My mom used to like chocolate.	maɪ mɑm juzt tu laɪg ʧɔklət	+Assimilation	Final *k* segment moves to lenis according to post-alveolar affricative ʧ
4	I have never seen your family.	aɪ hæv nɛvər sin jʊər fæməli.	−Elision	None
5	He hosts several games.	hi hoʊs sɛvrəl geɪmz.	+Elision	*ts* elides in consonant cluster
6	Send them to me by mail.	sɛnd ɛm tu mi baɪ meɪl.	+Elision	ð elides after *d*
7	The best estimate is secret.	ðə bɛs ɛstəmət ɪz sikrət.	+Assimilation	Final plosive *t* moves toward alveolar fricative *s* to create a cluster
8	East side of the city.	is saɪd ʌv ðə sɪti.	+Elision	*t* elides between alveolar fricative *s*
9	Tell him the drink you want.	tɛl hɪm ðə drɪŋk ju wɑnt.	−Elision	None
10	Somehow, we woke up very late.	sʌmˌhaʊ i woʊk ʌp vɛri leɪt.	+Elision	One approximant w elides from the pair
11	I did not recognize the spot.	aɪ dɪd nɑt rɛkəgˌnaɪz ðə spɑt.	−Assimilation	None
12	Her daughter writes quite well.	hɜr dɔtər raɪts kwaɪt⁼ wɛl.	+Assimilation	Fortis plosive *t* moves to lenis to accommodate the approximant *w*
13	Did your father used to work there?	dɪd jʊər fɑðər juzd tu wɜrg ðɛr?	+Assimilation	Plosive *k* moves to lenis *g* to alleviate the fricative ð
14	There are ten cars in the car shop.	ðɛr ɑr tɛŋ kɑrz ɪn ðə kɑr ʃɑp.	+Assimilation	*n* moves to velar nasal ŋ to alleviate the plosive *k*
15	The river runs that way.	ðə rɪvər rʌnz ðæt weɪ.	−Elision	None
16	Your joint effort to win paid off.	jʊər ʤɔɪnt⁼ɛfərt tu wɪn peɪd ɔf.	+Elision	*t* elides after *n* to remove cluster
17	We have to buy a new car soon.	wi hæv tu baɪ ə nu kɑr sun.	−Assimilation	None
18	That video shouldn’t be watched.	ðæt vɪdioʊ ʃʊdənt bi wɑʧt.	+Elision	Contraction of *not*: *o* is elided
19	He’ll be about an hour.	hil bi əbaʊt ən aʊər.	−Elision	None
20	He’s surprised about the pressure.	hiz sərpraɪzd əbaʊt ðə prɛʃər.	−Assimilation	None
21	Concert tickets are cheap.	kɑnsɜrt tɪkəts ɑr ʧip.	−Elision	None
22	What do you think is funny?	wʌt du ju θɪŋk ɪz fʌni?	−Assimilation	None
23	Can’t you meet me tonight?	kænt ju mit⁼ mi tənaɪt?	+Assimilation	Plosive *t* becomes palatalized for the nasal *m*

Table 2 (continues)

Table 2 (continued)

Item	Target sentence	IPA transcription	Type of Sandhi-variation	Description of the Sandhi-Variation
24	What do you think about?	wʌt du ju θɪŋg əbaʊt?	+Assimilation	Fortis plosive *k* moves to lenis *g* to accommodate open vowel *a*
25	My employer ran in the race.	maɪ ɛmplɔɪər æn ɪn ðə reɪs.	+Elision	One *r* elides from the pair
26	Satisfy your dreams with software.	sætəˌsfaɪ jʊər drimz wɪð sɔfˌtwɛr.	+Assimilation	Fortis fricative *s* moves to lenis *z* to accommodate the approximant *w*
27	Does your family live here still?	dʌz jʊər fæməli lɪv hir stɪl?	−Elision	None
28	I have a good neighbor.	aɪ hæv ey gʊd neɪbər.	−Assimilation	None
29	She has lived there about a year.	ʃi hæz laɪvd ðɛr əbaʊt ey jɪr.	−Elision	None
30	We are grateful for it.	wi ɑr greɪtfəɫ fɔr ɪt.	+Elision	*l* elides between pair of labiodental *f*
31	It's useful to clean it.	ɪts jusfəl tu klin ɪt.	−Elision	None
32	I don't know what class I should take.	aɪ doʊnt noʊ wʌt klæs aɪ ʃʊd teɪk.	−Assimilation	None
33	She did not like his shirt very much.	ʃi dɪd nɑt˭ laɪk hɪz ʃɜrt vɛri mʌʧ.	+Assimilation	Fortis plosive *t* moves to lenis to alleviate lateral alveolar *l*
34	Does its shape change all of the time?	dʌz ɪts ʃeɪp ʧeɪnʤ ɔl ʌv ðə taɪm?	−Assimilation	None
35	What are we doing here?	wʌt ɑr wi duɪŋ hir?	−Assimilation	None
36	When did you go to the park?	wɛn dɪd ju goʊ tu ðə pɑrk?	−Assimilation	None
37	He will be about an hour late.	hi wɪl bi baʊt ən aʊər leɪt.	+Elision	Initial weak vowel *a* elides
38	I have to go to school.	aɪ hæf tu goʊ tu skul.	+Assimilation	Lenis fricative *v* moves to fortis *f* to allow for plosive *t*
39	It was attached to the panel.	ɪt wʌz ətæʧ tu ðə pænəl.	+Elision	Final ʧt consonant cluster elides for dominant initial *t*
40	Our area is here.	aʊər ɛriə ɪz hir.	−Assimilation	None
41	We breathe in air deeply.	wi brið ɪn ɛr dipli.	+Assimilation	Initial *i* moves to schwa to accommodate ð
42	Who is your regular doctor?	hu ɪz jʊər rɛgjələr dɑktər?	−Elision	None

Table 2 (continues)

Table 2 (continued))

Item	Target sentence	IPA transcription	Type of Sandhi-variation	Description of the Sandhi-Variation
43	Will the owners of the coat be found?	wɪɫ ði oʊnərz ʌv ðə koʊt bi faʊnd?	+Elision	Double *l* elides to make room for ð
44	The king got up to make a speech.	ðə kɪŋ gɑt ʌp tu meɪk eɪ spiʧ.	−Elision	None
45	It is painted bright blue.	ɪt ɪz peɪntəd braɪt blu.	−Assimilation	None
46	They will get them in time.	ðeɪ wɪl gɛt ðɛm ɪn taɪm.	−Elision	None
47	Best side is on the left.	bɛs saɪd ɪz ɑn ðə lɛft.	+Elision	Plosive *t* elides within consonant cluster
48	The actual labor was cheap.	ði ækʧuəl leɪbər wʌz ʧip	−Assimilation	None

assimilation (as it is currently defined). Furthermore, target sentences 18 and 23 contain contractions that are not included as target forms of elision. Contractions are high-frequency items with known qualities to even fairly low-proficiency learners.

Table 3 shows the type of assimilation used for each item in the current study and the item hierarchy as Rasch item measurements. Regressive assimilation is the most common form of assimilation in English and it dominated the assimilated items in this study. In the current study I investigated the general effects of both elision and assimilation on listening comprehension and parsing, not the conditional effects of those features. The one item with progressive, +assimilation (i.e., +A8+SEIT26: drimzwɪð) on the EIT was approximately in the middle of the item difficulty hierarchy, so it is impossible to say that regressive assimilation hindered processing more than progressive assimilation from the data.

Table 3. *Assimilated Items with Regressive and Progressive Forms*

Item code	Type of assimilation	Item	IPA
+A6+SEIT41	Regressive	breathe in	briðən
+A6+SEIT2	Regressive	in snow	ɪsnoʊ
+A6-SEIT12	Regressive	quite well	kwaɪwɛl
+A8+SEIT14	Regressive	ten cars	tɛŋkɑrz
+A6-SEIT38	Regressive	have to	hæftu
+A8-SEIT33	Regressive	not like	nɑdlaɪk
+A8+SEIT26	Progressive	dreams with	drimzwɪð
+A6-SEIT23	Regressive	meet me	midmi
+A8-SEIT3	Regressive	like chocolate	laɪgʧɔklət
+A8-SEIT13	Regressive	work there	wɜrgðɛr
+A8+SEIT7	Regressive	best estimate	bɛsɛstəmət
+A6-SEIT24	Regressive	think about	θɪŋgəbaʊt

The type of scoring used in elicited imitation tests has been shown to influence the tests' reliability to distinguish between participants across proficiency levels (Yan et al., 2015). Three common scoring methods, binary, ordinal, and interval, have been used in prior elicited imitation studies. A meta-analysis by Yan et al. (2015) showed that ordinal scoring was more capable of distinguishing speakers across proficiency levels compared with binary and interval scoring. Ordinal scoring was used for stage 1 of the test, based on Erlam (2006). Scoring followed three criteria:

1. Obligatory occasion created—supplied (2 points);
2. Obligatory occasion created—not supplied (1 point);
3. No obligatory occasion created.

For stage 2 of the test, binary scoring was used, which followed two criteria:

1. Obligatory occasion created—supplied (1 point);
2. No obligatory occasion created.

The scores from Stage 1 and Stage 2 were combined to form composite scores for each item.

A scoring example for the criteria is as follows: E.g., *Coffee is a waste of money* [kɑfi ɪz ə weɪst ʌv mʌni].

1. Obligatory occasion created—supplied (2 points): kɑfi ɪz ə weɪst ʌv mʌni
2. Obligatory occasion created—not supplied (1 point): kɑfi ɪz weɪst mʌni
3. No obligatory occasion created: No answer

For stage 2 of the test, binary scoring was used, which followed two criteria:

1. Obligatory occasion created—supplied (1 point): kɑfi ɪz ə weɪst ʌv mʌni
2. No obligatory occasion created: kɑfi ɪz weɪst ʌv mʌni

The first criterion describes a response in which the participant imitated the target form correctly, recreating the elision, assimilation, or plain form of the utterance. All the syllables are judged present. The second criterion describes a response where the participant imitated the form (e.g., elision) incorrectly. The specific target was incomplete, so, for instance, the participant was able to repeat only 6 of 8 syllables in the target. The third criterion describes a non-response to the prompt. In other words, the participant was silent. All responses that did not fit into either the first or the second criterion were included in the third criterion. The grading was repeated for the initial imitation and the second stage, enunciation repetition, when the participant enunciated the words. For the second stage, the participant was either able to enunciate the entire sentence or not. The pilot was administered to three native English speakers, and each target (48 total) was responded to accurately for both the repetition and enunciation stages. In the next section, I describe the interview process.

Background and Length of Residency Interview

I created the Background and Length of Residency Interview to explore which specific types of input might have been present in the day-to-day lives of the participants (Appendix D). Research Question 3 investigated relationships that might exist between natural aspects of the language environment, including input from that environment, and the ability to parse elision and assimilation as it was presented in the EIT.

The Background and Length of Residence Interview served as an overarching variable with the following sub-components:

1. Current background
2. English education history (e.g., time learning four skills)
3. Conversation history with native English speakers

4. Length of residency abroad
5. English language media (e.g., time spent watching English language programs on television)

Questions were formulated to gather information about the learners' language experiences. The questions were contingency-based with a variety of response types required, including multiple-choice and open-ended answers. Specific questions were provided with a selection of answers that fit into the possible categories of the participants. For example, the question *What year are you at school?* had answer choices of *first*, *second*, *third*, and *fourth-year university*, which allowed for the only possible answers of the targeted participants. However, for questions with a broader scope, the answer was sometimes open to a variety of responses for example: *At what age did you first learn English in terms of speaking?* Examples of questions from each section are as follows:

1. Current background: *What year are you at school?*
2. English education history (e.g., time learning four skills): *How many years (cumulative) have you spent learning English?*
3. Conversation history with native English speakers: *Do you have a job where you use English?*
4. Length of residency abroad: *How old were you when you entered the country of your residence?*
5. English language media (e.g., time spent watching English language programs on television): *How often do you watch English language media (television, movies)?*

Finally, the interview questions were translated from English into Japanese, with several passes at the translation to gain optimal precision. A Japanese language instructor, teaching at the university level, who has native-level proficiency of English, translated the interview questions from English to Japanese. The translation was done using a verbal protocol in order to verify meaning of the English while translating. Two more Japanese language instructors reviewed the translation and made suggestions. The questions were then submitted to four Japanese learners of the same age and relative proficiency as the target participants, and who comprehended the questions without difficulty.

In the pilot stage, the participants took part in a verbal protocol as they completed the questions in their L1 in order to provide information about their understanding of the questions and why they responded in the way that they did. Such a protocol has been shown to provide valuable information and substantive validity evidence (Gass & Mackey, 2000). The translation of these questions was finalized after the pilot study.

The interviews, including questions, answers, and the verbal protocol, were recorded with an Olympus Voice-Trek DS-60 and transcribed. The interviews lasted approximately 20 minutes. The electronic WMA files were downloaded from the device onto a computer, and transferred to MP3 files. The MP3 files were used for transcribing the interviews. Finally, the interviews were translated from Japanese to English.

In summary, six instruments were used in this study. First, three instruments, the TOEFL PBT, the LVLT, and the MLAT-E, were used to gain a cross-sectional baseline of L2 proficiency, vocabulary knowledge, and L2 aptitude. Second, the Pre-Listening in English questionnaire was given to the participants to gauge their relationship with the

English sound system. Third, an EIT was used to test the participants' ability to parse speech streams with elision and assimilation. Fourth, the Background and Length or Residency interview was completed to gain deeper understanding of the participants' relationship with English input.

Research Design

This study was a one-semester, observational, panel study, using mixed methods. In all cases, the participants provided both quantitative and qualitative data. Test scores on the TOEFL PBT, the LVLT, the MLAT-E, and the EIT were correlated with the data gathered on the English language input the participants have received. The dependent variable consisted of the Rasch item difficulty estimates (i.e., comprehension) on the EIT. The independent variables were elision, assimilation, utterance rate, and syllable count.

Data for the persons and items were available for the LVLT, the MLAT-E, and the EIT. Item and person data were not available from the TOEFL PBT test, as the results of the test were from an institutional TOEFL test, which mandates that test forms and results be restricted in accordance with ETS rules. Section scores from the Listening, Structure and Written Expression, and Reading sections of the TOEFL PBT were included in this study for a standardized proficiency variable. Available persons- and items-data from the LVLT, the MLAT-E, and the EIT were analyzed using the dichotomous Rasch model.

Procedures

Participant involvement was contingent on enrollment in intensive university English courses and available TOEFL PBT results. The participants completed the TOEFL PBT twice, one day apart (April 1 and April 2, 2016), during the orientation week before the semester in which the data were gathered. After the participants had completed the test twice, the higher score was used to place them in a course level. The higher of the participants' two scores was used as the baseline TOEFL PBT proficiency score in the current study.

The LVLT was administered at the beginning of the semester, one week after the TOEFL PBT placement tests (April 9, 2016). The first through third 1,000 word frequency level (test items 1-72) was administered in one sitting. I estimated that the participants would not score higher than the minimum criterion score for the second 1,000 word frequency level, so the first and second 1,000 word frequency level items were used to verify that they knew the lexis on the EIT. The participants were given directions in their L1 (Japanese) and answer sheets for the test. The audio was played from electronic files on a computer, and through a single Bluetooth speaker (Bose Soundlink II). Participants marked their answers by circling the letter of their answer choice on the test sheet, and the completed test sheets were then collected and scored by the proctor.

I administered and scored the MLAT–E at the beginning of the semester, two days after the LVLT (April 11, 2016). The MLAT–E test packs and score sheets used in the study were available to order online from the Language Learning and Testing Foundation and were sent through the mail.

The Pre-Listening in English questionnaire was administered ten days after the MLAT-E (April 21, 2016). Students filled out the four-question, open-ended-answer questionnaire in Japanese. The questionnaire was collected and translated by a bilingual, Japanese staff member with a master's degree in teaching English.

The EIT was administered one day after the Pre-Listening in English questionnaire (April 22, 2016) using open source VideoLan software (commonly referred to as VLC) for Mac OS X, and Audacity®: Free Audio Editor and Recorder software, Version 2.1. The VLC media player presented the target utterances and Audacity®: Free Audio Editor and Recorder software recorded the participants' aural responses. The test was administered with one computer and one piece of software, as opposed to using different production, recording, and prompting devices. That said, an external Bluetooth speaker (Bose Soundlink II) was used instead of the computer's speaker to ensure adequate sound quality. Furthermore, the participants' tests were externally recorded using an Olympus Voice-Trek DS-60 and a second back up recording using Audacity®: Free Audio Editor and Recorder was made using the computer's embedded microphone. The recorded data (i.e., participants' imitation attempts) were transcribed and scored using the rubric defined in the Elicited Imitation Test section.

Finally, the participants were asked to complete the Background and Length of Residency interview in their L1, Japanese, which was administered on the same day as the EIT (April 22, 2016). Each participant was interviewed using the questions in the Background and Length of Residency interview question (Appendix D). The interviewer was a native Japanese speaker with a degree in second language teaching. The interviews were recorded, transcribed, and translated by the interviewer.

Analyses

In this section I explain how each of the four research questions were analyzed. Research Question 1a concerning the empirical item hierarchy for the phonological features of elision, and assimilation, and the determinants of this hierarchy was answered by inspecting the Rasch item difficulty estimates. Further analyses were conducted using *t*-tests to compare the different phonological features embedded in the items. The paired *t*-tests were conducted for elision and different utterance rates, elision and different utterance lengths, assimilation and different utterance rates, and assimilation and different utterances lengths. The assumptions for the paired *t*-test are as follows: (a) The variables have a normal distribution, and (b) the variables are continuous, interval variables.

Research Question 1b concerning how input rates and phrase lengths influence the participants' comprehension of the phonological features of elision and assimilation was answered by inspecting the Rasch item difficulty estimates. The paired *t*-test data from research question 1a were reused to answer the question. The assumptions for the paired *t*-test are as follows: (a) The variables have a normal distribution, and (b) the variables are continuous, interval variables.

Research Question 2 concerning the strength of the relationship between the participants' language proficiency, as established by the TOEFL PBT, the LVLT, and the MLAT-E, and the comprehension of the phonological features of elision and assimilation is answered as follows. First, the data were analyzed using the Rasch model to determine overall item quality, item difficulty, and person ability. The Rasch item and person reliability and separation estimates were reported. These analyses showed the hierarchy of item difficulty for elision and assimilation, which can be used to investigate any relationship with proficiency (e.g., aural vocabulary ability scores from the LVLT). Furthermore, hierarchical complexity has been built into the EIT in the forms of number of syllables per utterance (i.e., 6 or 8 syllables per utterance) and rate of utterance.

The EIT was correlated with the TOEFL PBT, LVLT, and MLAT-E. Raw scores on the EIT, which produced raw scores for participants' perception of elision and assimilation in utterances, faster and slower utterance rates, and longer and shorter utterances, were correlated with proficiency variables including TOEFL PBT scores, LVLT scores, and MLAT–E scores. Finally, this research question was answered with a Pearson correlation to determine the degree to which proficiency scores are related to scores on the EIT sentence length and utterance rate for elision and assimilation. The assumptions for the Pearson correlation are that (a) the variables have normal distribution, (b) the variables are continuous, interval variables, and (c) there is a linear relationship between the variables.

Research Question 3 concerning the strength of the relationship between participants' English-learning backgrounds and their comprehension of the phonological features of elision and assimilation was answered by coding themes in the questionnaire to links in the quantitative data, for example, time abroad with ability to comprehend utterances with the phonological features of elision and assimilation. The five sections include current background, English education history, conversation history with native English speakers, length of residency abroad, and English language media.

Table 4 shows that both answers, 8 years and 7 years, are coded *2*. In this example, the default code is 2 because all of the participants have had at least 7 years of English instruction, as English instruction is implemented from the elementary school level (小学

校; *shōgakkō*). A participant with less than 7 years of English education was coded 1. A participant with more than 10 years of English education was coded 3. The questionnaire takes into account different types of English education (e.g., elementary school English) and the possibility of overlap (e.g., elementary school English and a home tutor for English), and such variations receive different codes.

Table 4. Example Coding for Answers from the Background and Length of Residency Interview, Section 2, English Education History (EEH)

Participant	EEH: How many years have you spent learning English?	
1	Answer: *I have been learning English since Elementary school. Around 8 years.*	Code: 2
2	Answer: *7 years because I started in Elementary.*	Code: 2

Next, a Pearson correlation was calculated to determine how the background variables were related to the comprehension scores. The background variables were (a) current background, (b) English education history, (c) conversation history with native English speakers, (d) length of residency abroad, and (e) English language media. The Pearson correlation showed the degree to which the background variables were related to the EIT scores. In the next section, I describe the Rasch model, including the fit statistics, the principal component analysis of item residuals, and the partial-credit model.

The Rasch Model

The Rasch model allows for objectivity in measurement because it produces an equal interval scale from ordinal data (Bond & Fox, 2015). As items become more difficult on a test, for instance, the items become more difficult for all of the test-takers. The Rasch mathematical model allows for analyses of how each participant's ability on an instrument interacts with each item's difficulty to produce an observed outcome (Linacre, 2016b). The Rasch model also orders persons according to their estimated ability and items according to their estimated difficulty on the same interval measurement scale, called a logit scale (Bond & Fox, 2015).

The Rasch model is used to analyze dichotomous data, which are data that have two values, for example 1 and 0, where the 1 has greater value than 0. The model predicts that a test taker, for example, has a better probability to answer easier questions correctly and difficult questions incorrectly. The formula for the dichotomous Rasch model is $B_n – D_i = \ln(P_{ni}/1–P_{ni})$, where B_n is the ability of the test taker along the variable; Di is the difficulty of a test item; P_{ni} is the probability of the test taker correctly answering a specific test item; and $1 - P_{ni}$ is the probability of a test taker incorrectly answering a test item; ln is the natural log function. When a respondent B_n answers an item D_i, the response is expressed by the natural log of the respondent correctly answering the item (P_{ni}) divided by the probability of the respondent not correctly answering the test item ($1 - P_{ni}$). Thus, the Rasch mathematical model makes use of the location of a respondent and the test items along a single variable.

Rasch Fit Statistics

To check that the data fit the Rasch model, and thereby validate the objectivity of the instruments, the data were assessed with a variety of item and person fit statistics. Item-fit statistics indicate how item responses fit the expectations of the model, as well as whether a person's response to a single item is consistent with that person's response to another item on the same scale. Person-fit statistics indicate how an individual's responses fit the model expectations, and if there is consistency of a response to an item matched to other responses within the sample.

The Rasch model includes two unstandardized fit statistics, outfit MNSQ (Mean Square) and infit MNSQ. Outfit MNSQ indicates the average of the standardized residual variance across items and persons and is unweighted (i.e., it is affected by unexpected responses far from a particular item or person measure). Outfit MNSQ is based on the conventional sum of squared standardized residuals. The squared standardized residual is $z^2 = (X\text{-}E)^2/\sigma^2$. Outfit MNSQ is $\boldsymbol{\Sigma}\ (z^2)/N$, where N is the number of observations summed, X is an observation, E is the expected value based on Rasch parameter estimates, and σ^2 is modeled variance of E. For the current study, outfit MNSQ were not analyzed in the validation chapter, as the data are dominated by low information responses and therefore outfit MNSQ is outlier-sensitive (Linacre, 2016b).

The infit MNSQ, on the other hand, is weighted by individual variance, so it is affected by unexpected responses closer to a particular item or person measure (Bond & Fox, 2015). The range of infit MNSQ values for fit to the Rasch model is from 0.7–1.3, based on the criterion set by Bond and Fox (2015). The squared standardized residual is $z^2 = (X\text{-}E)^2/\sigma^2$. Infit MNSQ is $\boldsymbol{\Sigma}\ (z^2\sigma^2)/\ \boldsymbol{\Sigma}\ (\sigma^2) = \boldsymbol{\Sigma}\ ((X\text{-}E)^2)/\boldsymbol{\Sigma}\ (\sigma^2)$, where X is an observation, E is the expected value based on Rasch parameter estimates, and σ^2 is modeled variance of E, which is the statistical information in a Rasch observation.

The Rasch model also provides two standardized fit statistics, infit ZSTD (Z standard) and outfit ZSTD, which are useful when the sample size is small (Linacre, 2006). Data that fit the Rasch model perfectly are expected to have a value of 0.0. Data valued at less than 0.0 indicate that they are too predictable, and more than 0.0 indicate that the data lack predictability. Standardized infit or outfit values that are greater than 2.0 indicate that the item misfits the model, and that it might not be measuring the same latent variable as the other items. Such items with values exceeding 2.0 should be considered for deletion (Linacre, 2016b). The formula for the ZSTD is ZSTD = $(MnSq^{1/3} - 1)(3/q) + (q/3)$, where $q^2 = 2/df$, and where df represents the degrees of freedom.

Rasch Principal Component Analysis of Item Residuals

A Rasch principal component analysis (PCA) of item residuals was used to detect items not functioning as indicators of the constructs being measured. Measuring a single dominant dimension, *unidimensionality*, is an assumption of the Rasch model. Unidimensionality underlies accurate measurement and interpretation of person ability and item difficulty. Sets of items with explained variance by the Rasch model that exceeds 50%, with the first residual contrast accounting for less than 15% of the variance, and the eigenvalue of the first residual contrast less than 3.0, are considered unidimensional (Linacre, 2016b).

Rasch Partial Credit Model

The Rasch Partial Credit Model was used for analyzing the EIT, as scores on the instrument range from 0 to 3. The Rasch Partial Credit Model is a rating scale model in which the thresholds are not constrained, so variance in threshold estimates from item to item is accepted. The formula for the Rasch Partial Credit Model is $\ln(P_{nik} /1 - P_{ni}) = B_n - D_{ik}$. Similar to the Rasch dichotomous formula (which the Partial Credit Model becomes when there are only two items), B_n represents a respondent, but, in this formula, answers to an item were D_{ik}. D_{ik} represents the item-unique threshold estimates. The response is expressed by the natural log of the respondent correctly answering the item (P_{nik}) divided by the probability of the respondent not correctly answering the test item ($1 - P_{ni}$) (Linacre, 2016b).

Chapter Summary

In this chapter I reviewed the methods of this study. The review was divided into seven sections: (a) Educational Setting, (b) Participants, (c) Instrumentation, (d) Research Design, (e) Procedures, (f) Analyses, and (g) The Rasch Model. The Educational Setting section described where the study took place. The Participants section described the participants of this study. The Instrumentation section described each of the six instruments: (a) The TOEFL PBT, (b) the LVLT, (c) the MLAT-E, (d) EIT, (e) the Pre-Listening in English questionnaire, and (f) the Background and Length of Residency interview. The Procedures section described the steps used in the study. The Analyses section described how the data were analyzed. I concluded the chapter by describing the Rasch Model: (a) The Rasch Fit Statistics, (b) the Rasch Principal Component Analysis of Item Residuals, and (c) the Rasch Partial Credit Model. In the next chapter, I describe the validity evidence for the six instruments used in the current study.

CHAPTER 4

INSTRUMENT VALIDATION

In this chapter I describe the validity evidence for the six instruments used in the study: a standardized proficiency test, an aural vocabulary levels test, a standardized aptitude test, a Pre-Listening in English questionnaire, a background and length of residency interview, and an elicited imitation test. A series of pilot tests were administered to participants over the course of a single academic semester to gather data that could be used to provide validity evidence.

Standardized Instruments

Test scores from published tests were used to assess the participants' English language proficiency, aural vocabulary levels, and aptitude, prior to completing the Elicited Imitation Test, which was created for this study to test the participants' abilities to process Sandhi-variation. The proficiency test in this study was the Institutional Testing Program Test of English as a Foreign Language Paper-Based Test (TOEFL PBT). Aural vocabulary knowledge was assessed with the Listening Vocabulary Levels Test, and aptitude was assessed with the Modern Language Aptitude Test–Elementary.

TOEFL PBT

The TOEFL PBT results are used as an English-language proficiency variable. The test is divided into three sections: Listening, Structure and Written Expression, and

Reading. The Listening section scores for the participants are also used as one measure of listening comprehension.

The TOEFL and its variants (e.g., TOEFL Internet-Based Test) are widely used amongst academic institutions to gauge academic English proficiency (Cumming, Grant, Mulcahy-Ernt, & Powers, 2004). The Educational Testing Service (ETS) publishes validity studies to support the use of their tests and these studies can be found on the ETS website (www.ets.org) under the *Monograph Series* link. Test item scores for each test are not made available by ETS, so validity evidence cannot be assessed at the individual person or item level for the participants in this study.

The TOEFL scores were gathered from its use as a pre-semester placement test, which was given by the participating institution to place the participants into one of three English language proficiency streams in the university's English language program. All participants took the TOEFL on the same day, during the time allotted by the institution for the test. The test was proctored according to the Supervisor's Manual published by the ETS. The test booklets and answer sheets were secured, sent to an ETS testing center immediately, and scores were sent back to the institution one week later. Table 5 shows the overall, institutional results data for the participants from the TOEFL PBT. Individual total scores and Listening Section scores only were made available for this study.

Table 5. *Listening Section and Total Score for the TOEFL PBT*

	N	Minimum	Maximum	*M*	*SD*
Listening	25	43.00	68.00	50.20	5.15
Total score	25	423.00	613.00	491.72	43.64

Listening Vocabulary Levels Test

The Listening Vocabulary Levels Test (LVLT) (Appendix A) raw scores are used to assess the participants' readiness to handle the first 2,000 high-frequency words of English and the lexical load of the aural targets in the Elicited Imitation Test. The Elicited Imitation Test is made up of items selected from the first 2,000 high frequency words of English from the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA), which were chosen for the Elicited Imitation Test because both corpuses use tokens from spoken American English and because the tokens chosen for the corpuses represent current English used by native speakers of English (Webb & Sasao, 2013). The LVLT items are also from the BNC and COCA. Knowledge of the first 3,000 high frequency words of English (i.e., Levels 1-3) was tested using 24 items at each frequency level (i.e., 24 items x 3 levels = 72 items total). McLean, Kramer, and Beglar (2015) showed that the items on the test displayed good fit to the Rasch model, and they found that word frequency was a good predictor of difficulty in that higher frequency words were less difficult, so the hypothesized order of difficulty for the words on the LVLT was generally supported.

The result of a dichotomous Rasch analysis (Rasch, 1960) using WINSTEPS 3.80 (Linacre, 2016a) indicated that the Listening Vocabulary Levels Test for the current data had an item reliability (separation) estimate of .43 (.86) and a person reliability (separation) estimate of .74 (1.67). According to Fox and Jones' (1998) criteria, the item reliability estimate was poor and the person reliability estimate was fair. Because of poor reliability, the precision of the person ability estimates was checked. It showed that *SE*

= .41 to 1.06, *M* = 1.97, *SD* = 1.15. Low item reliability estimates can occur because of a narrow range of item difficulty and/or a small sample size. A low item reliability estimate for the LVLT items was expected because the participants were required to know 67% of the items on the test form used in this study. In other words, the item reliability for the LVLT is similar to the difficulty estimates.

Table 6 shows item-level Rasch statistics for the Listening Vocabulary Levels Test items. The item labels represent the frequency level (1 = the first 1,000 words; 2 = the second 1,000 words, and 3 = the third 1,000 words) and the frequency level question number. For example, 2.17 means that it is item 17 of the second 1,000 high frequency words of English. Excluding the items listed as MINIMAL MEASURE, that is, items that all participants answered correctly, the items showed reasonable fit to the Rasch model with infit MNSQ values ranging from .76 to 1.36, which is slightly outside of the 0.7-1.3 criterion recommended by Bond and Fox (2015). Item 3.23 (*constituent*) underfit with an Infit MNSQ statistic of 1.36; it was removed and the analysis was run again. A Pearson correlation of the person ability estimates with and without the item was conducted. The two sets of person measures correlated at $r = .99$ ($p < .01$), indicating that keeping item 3.23 did not degrade the measurement. Furthermore, because the item showed only slight misfit and because its inclusion in the test was purposeful, the item was retained. The point-measure correlations revealed that a number of items, such as item 1.11 (*any*), item 2.07 (*pub*), and item 3.23 (*constituent*), had low correlations, and item 3.07 had a negative correlation. Low correlations (< .12) can be the result of an item showing weak differentiation from one level of measurement to the next (Bond & Fox, 2015). The low correlation of the items did not result in item exclusion. Negative correlations can be caused by special knowledge of an item, which, for item 3.07 (*behavior*), might be due to the high frequency of use within the participants' academic setting. In other words, the item is so easy for university students, which stands in opposition to the latent variable, that it has a negative correlation. Table 6 is organized by Rasch item difficulty estimates, from the most to the least difficult.

Table 6. *Summary of the Rasch Statistics for the Listening Vocabulary Levels Test Items*

Item	Measure	*SE*	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Pt-measure correlation
2.17	2.98	.49	1.03	.2	1.52	1.3	.32
3.20	2.98	.49	.96	-.1	.99	.1	.44
3.22	2.98	.49	.85	-.6	.74	-.6	.56
3.04	2.53	.46	.87	-.6	.82	-.6	.53
3.23	2.12	.44	1.36	2.1	1.56	2.0	.07
3.12	1.93	.44	1.16	1.1	1.57	2.0	.22
3.19	1.73	.44	1.17	1.1	1.39	1.4	.22
3.21	1.54	.44	1.03	.2	.94	-.1	.39
3.24	1.34	.45	.99	.0	.91	-.2	.41
3.01	1.14	.45	1.09	.5	1.04	.2	.31
3.02	1.14	.45	1.02	.2	.97	.0	.37
3.05	1.14	.45	.83	-1.0	.74	-.7	.54
3.10	.93	.47	1.21	1.1	1.12	.4	.21
2.09	.47	.50	.92	-.2	.73	-.4	.54
2.01	.20	.53	1.13	.5	.92	.1	.24
3.08	.20	.53	.78	-.7	.86	.0	.47
2.06	-.11	.58	.82	-.4	.83	.0	.42
2.24	-.11	.58	1.19	.6	1.44	.8	.09
1.20	-.48	.64	1.15	.5	1.23	.5	.13
2.12	-.48	.64	.86	-.2	.49	-.4	.43
2.22	-.48	.64	.76	-.5	.46	-.5	.49
3.13	-.48	.64	.97	.1	.72	-.1	.32
3.16	-.48	.64	.93	.0	.71	-.1	.34
2.07	-.96	.76	1.18	.5	1.44	.7	.02
2.23	-.96	.76	1.08	.3	.89	.3	.17
1.10	-1.74	1.04	1.06	.4	.84	.4	.12
1.11	-1.74	1.04	1.10	.4	1.26	.6	.04
1.12	-1.74	1.04	.84	.1	.26	-.3	.36
1.14	-1.74	1.04	.84	.1	.26	-.3	.36
2.05	-1.74	1.04	1.06	.4	.84	.4	.12
2.14	-1.74	1.04	.88	.2	.30	-.2	.33
3.03	-1.74	1.04	.84	.1	.26	-.3	.36
3.07	-1.74	1.04	1.12	.4	1.98	1.0	-.05
3.09	-1.74	1.04	1.06	.4	.84	.4	.12
3.15	-1.74	1.04	.84	.1	.26	-.3	.36

Table 6 (continues)

Table 6 (continued)

Item	Measure	*SE*	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Pt-measure correlation
3.17	-1.74	1.04	.88	.2	.30	-.2	.33
3.18	-1.74	1.04	.84	.1	.26	-.3	.36
1.01	-2.99	1.84		MINIMUM MEASURE			.00
1.02	-2.99	1.84		MINIMUM MEASURE			.00
1.03	-2.99	1.84		MINIMUM MEASURE			.00
1.04	-2.99	1.84		MINIMUM MEASURE			.00
1.05	-2.99	1.84		MINIMUM MEASURE			.00
1.06	-2.99	1.84		MINIMUM MEASURE			.00
1.07	-2.99	1.84		MINIMUM MEASURE			.00
1.08	-2.99	1.84		MINIMUM MEASURE			.00
1.09	-2.99	1.84		MINIMUM MEASURE			.00
1.13	-2.99	1.84		MINIMUM MEASURE			.00
1.15	-2.99	1.84		MINIMUM MEASURE			.00
1.16	-2.99	1.84		MINIMUM MEASURE			.00
1.17	-2.99	1.84		MINIMUM MEASURE			.00
1.18	-2.99	1.84		MINIMUM MEASURE			.00
1.19	-2.99	1.84		MINIMUM MEASURE			.00
1.21	-2.99	1.84		MINIMUM MEASURE			.00
1.22	-2.99	1.84		MINIMUM MEASURE			.00
1.23	-2.99	1.84		MINIMUM MEASURE			.00
1.24	-2.99	1.84		MINIMUM MEASURE			.00
2.02	-2.99	1.84		MINIMUM MEASURE			.00
2.03	-2.99	1.84		MINIMUM MEASURE			.00
2.04	-2.99	1.84		MINIMUM MEASURE			.00
2.08	-2.99	1.84		MINIMUM MEASURE			.00
2.10	-2.99	1.84		MINIMUM MEASURE			.00
2.11	-2.99	1.84		MINIMUM MEASURE			.00
2.13	-2.99	1.84		MINIMUM MEASURE			.00
2.15	-2.99	1.84		MINIMUM MEASURE			.00
2.16	-2.99	1.84		MINIMUM MEASURE			.00
2.18	-2.99	1.84		MINIMUM MEASURE			.00
2.19	-2.99	1.84		MINIMUM MEASURE			.00
2.20	-2.99	1.84		MINIMUM MEASURE			.00
2.21	-2.99	1.84		MINIMUM MEASURE			.00
3.06	-2.99	1.84		MINIMUM MEASURE			.00
3.11	-2.99	1.84		MINIMUM MEASURE			.00
3.14	-2.99	1.84		MINIMUM MEASURE			.00
M	21.9	-1.45	.99	.2	.88	.2	
SD	5.0	1.87	.15	.6	.43	.7	

Note. Pt-M Corr = point-measure correlation.

Figure 4 shows that the 72 LVLT items exhibited various levels of difficulty. The most difficult item was Item 2.17, which is the word *refer* in the sentence, *She referred to*

```
---------------------------------------------------------------------
More able    |  More difficult items
participants |
  4         T+
          X  |
             |
             |
             |
             |
             |
             |T
  3       X  +  2.17: refer      3.20: personnel   3.22: devastated
            S|
         XX  |
             |
             |  3.04: paved
         XX  |
             |
       XXXX  |  3.23: constituent
  2          +
          X M|  3.12: impose
             |  3.19: mode
       XXXX  |S
          X  |  3.21: competent
             |  3.24: weave
         XX  |
             |  3.01: restore    3.02: compound    3.05: remedy
  1       X  +
         XX S|  3.10: conceive
          X  |
             |
             |  2.09: cross
             |
             |  2.01: time       3.08: fuel
          X  |
  0       X  +M
             |  2.06: bacterium 2.24: weave
            T|
             |
             |  1.20: handle    2.12: resist    2.22: seal   3.13: solution
             |  3.16: tunnel
             |
             |
 -1          +  2.07: pub       2.23: warn
             |
             |
             |
             |
             |S
             |  1.10: actual    1.11: any       1.12: far    1.14: caused
             |  2.05: jump      2.14: cause     3.03: latter 3.07: behavior
 -2          +  1.01: time      1.02: stone     1.03: poor   1.04: drive
             |  1.05: jump      1.06: shoe      1.07: test   1.08: nothing
             |  1.09: cross     1.13: game      1.15: many   1.16: where
             |  1.17: school    1.18: grow      1.19: flower 1.21: camp
             |  1.22: lake      1.23: past      1.24: round  2.02: period
             |  2.10: soldier   2.03: standard  2.04: basis
             |  2.08: circle    2.11: result    2.13: lend   2.15: speech
             |  2.16: pressure  2.18: army      2.19: knee   2.20: rope
             |  2.21: brand     3.06: bacterium 3.11: legend
             |  3.14: celebrate
Less able    |  Easier items
participants |
---------------------------------------------------------------------
```

Figure 4. Item-person map for Listening Vocabulary Levels Test. Each item label includes the frequency level (i.e., before the decimal) and the item number within the frequency level (i.e., after the decimal). X = 1 participant. M = mean, S = 1 SD, T = 2 SD.

him. One of the least difficult items was Item 3.14, which is the word *celebrate* in the sentence, *We have celebrated a lot recently*. The item difficulty hierarchy appears to be primarily caused by word frequency, as the item distribution of the first 2,000 high frequency words is bottom-heavy. Three items, 2.17, 2.09, and 2.01, from the second 1,000 high frequency words, are distributed along the vertical axis. All participants attained a minimum score of 21 out of the 24 total points (i.e., 87.5%) on both the first 1,000 high frequency words and the second 1,000 high frequency words.

Table 7 shows the PCA of item residuals results. The raw variance explained by the Rasch measures is low (36.2%), meaning that around 64% of the variance is random. The raw variance explained by measures did not meet the Rasch PCA criteria for this study, which is > 50% of total variance explained by the first component; however, this was partly due to the low variance in the person ability estimates, as shown by the person separation statistic of .72. The Rasch model explained 23.9% of the variance and that the unexplained variance in the first residual contrast accounted variance and that the unexplained variance in the first residual contrast accounted for 7.3 units (12.7%) of the total variance, so it does not appear to be unidimensional. The first residual contrast did not meet the < 2.0 criterion of the unexplained variance. The eigenvalues of unexplained variance in each of the contrasts is > 2.0, especially in the first contrast, which indicates contrasting patterns in the residuals at a strength of around 7 items. The PCA of item residuals indicates the presence of additional constructs in the data.

Table 7. *Standardized Residual Variance (in Eigenvalue Units) for Listening Vocabulary Levels Test*

	Eigenvalue	% of variance
Total raw variance in observations	58.0	100.0%
Raw variance explained by measures	21.0	36.2%
Raw variance explained by persons	7.2	12.3%
Raw Variance explained by items	13.8	23.9%
Raw unexplained variance (total)	37.0	63.8%
Unexplained variance in 1st contrast	7.3	12.7%
Unexplained variance in 2nd contrast	3.8	6.6%
Unexplained variance in 3rd contrast	3.5	6.0%
Unexplained variance in 4th contrast	3.2	5.6%
Unexplained variance in 5th contrast	2.8	4.8%

Table 8 shows the items with positive and negative residual loadings for the LVLT. The positive loading item residuals > .40 include *far* (1.12), *cause* (1.14), *latter* (3.03), *independence* (3.15), and *reviewed* (3.18). The negative residual loadings < -.40 include *resist* (2.12), *maintain* (2.01), *solution* (3.13), *refuse* (2.14), and *reward* (3.17). The salient characteristic of positive loading items is that 60% are from the third 1,000 most frequent words. All of the items with high positive loadings are extremely difficult and underfit the model, which creates the appearance of two factors. The contrast between the top five items with strong positive loadings (i.e., easy items) and the bottom five items with negative loadings (i.e., difficult items) is indicative of a secondary dimension, difficulty. No action was taken because the secondary dimension does not affect the unidimensionality of the test, as difficulty is not a legitimate dimension.

Table 8. *Item Loadings from the Rasch PCA of Residuals for the Listening Vocabulary Levels Test*

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
1.12	.96	-1.74	.84	.26
1.14	.96	-1.74	.84	.26
3.03	.96	-1.74	.84	.26
3.15	.96	-1.74	.84	.26
3.18	.96	-1.74	.84	.26
2.22	.37	-.48	.76	.46
3.16	.34	-.48	.93	.71
3.10	.27	.93	1.21	1.12
3.21	.23	1.54	1.03	.94
3.24	.22	1.34	.99	.91
2.09	.20	.47	.92	.73
3.05	.20	1.14	.83	.74
2.06	.17	-.11	.82	.83
3.08	.16	.20	.78	.86
3.20	.15	2.98	.96	.99
3.12	.14	1.93	1.16	1.57
3.22	.12	2.98	.85	.74
3.04	.08	2.53	.87	.82
2.17	-.78	2.98	1.03	1.52
3.23	-.54	2.12	1.36	1.56
3.01	-.44	1.14	1.09	1.04
3.02	-.43	1.14	1.02	.97
3.19	-.40	1.73	1.17	1.39
2.12	-.37	-.48	.86	.49
2.01	-.36	.20	1.13	.92
3.13	-.30	-.48	.97	.72
2.14	-.28	-1.74	.88	.30
3.17	-.28	-1.74	.88	.30
2.24	-.25	-.11	1.19	1.44
2.23	-.22	-.96	1.08	.89
1.20	-.13	-.48	1.15	1.23
2.07	-.10	-.96	1.18	1.44
3.07	-.09	-1.74	1.12	1.98
1.11	-.04	-1.74	1.10	1.26
3.09	-.03	-1.74	1.06	.84
1.10	-.01	-1.74	1.06	.84
2.05	-.01	-1.74	1.06	.84

Raw scores were then accepted for use to demonstrate readiness to handle the first 2,000 high frequency words of English. The minimum score set to demonstrate adequate knowledge of each frequency level was 21 correct responses (88%) out of the 24 items at each frequency level, a criterion that was met by all participants; thus, all the participants were considered to have acquired the first 2,000 high frequency words of English. The lowest score among the participants for the first 1,000 most frequent words was 23 (i.e., 96%; *M* = 23.72, *SD* = .61), and the lowest score on the second 1,000 most frequent words was 21 (i.e., 88%; *M* = 22.04, *SD* = 1.65). The lowest score on the test of the third 1,000 most frequent words was 11 (i.e., 46%; *M* = 17.44, *SD* = 3.16). Table 9 shows the minimum and maximum scores as well as the 95% confidence intervals for first through third 1,000 high frequency word levels.

Table 9. *Descriptive Statistics for the Listening Vocabulary Levels Test*

	Minimum	Maximum	*SD*	95% CI
1st 1,000 High-frequency words	22	24	.61	[23.48, 23.96]
2nd 1,000 High-frequency words	19	24	1.65	[21.39, 22.69]
3rd 1,000 High-frequency words	11	24	3.16	[16.20, 18.68]

Modern Language Aptitude Test—Elementary

The Modern Language Aptitude Test—Elementary (MLAT–E) was used to measure the participants' phonemic coding aptitude. The aim of the four-part test includes assessment of test-takers' ability to associate sounds and symbols, their sensitivity to grammatical structure, their ability to hear speech sounds, auditory alertness (i.e., how well a participant understands what is heard), and ability to remember (i.e., how well a participant processes a new word). The four parts are labeled Hidden Words (30 items), Matching Words (30 items), Finding Rhymes (45 items), and Number Learning (25 items). No minimum score was set for the MLAT because the participants' ability to complete the Elicited Imitation Test (i.e., the target) was measured by other instruments (e.g., TOEFL PBT). Table 10 shows the minimum and maximum raw scores on each section of the MLAT, as well as the standard errors and confidence intervals.

Table 10. *Descriptive Statistics for the Modern Language Aptitude Test*

	Minimum	Maximum	*M*	*SD*	*SE*	95% CI
Hidden Words	8 (27%)	30 (100%)	17.60	5.41	.99	[15.48, 19.72]
Matching Words	13 (43%)	32 (100%)	24.28	3.85	.68	[22.71, 25.79]
Finding Rhymes	10 (22%)	45 (100%)	22.48	8.75	1.30	[19.05, 25.91]
Number Learning	2 (8%)	25 (100%)	21.12	8.23	1.65	[17.89, 24.35]

A separate analysis was completed for each section of the MLAT–E. The following analyses are for the sections Hidden Words, Matching Words, Finding Rhymes, and Number Learning.

Hidden Words. The result of a dichotomous Rasch analysis using WINSTEPS 3.80 indicated that the MLAT–E section, Hidden Words, had an item reliability (separation) estimate of .89 (2.81), and a person reliability (separation) estimate of .86 (2.53). The item and person reliability estimates are considered good, as they fall within the .81-.90 range. These results indicate that the instrument is sensitive enough to distinguish between high and low performers. Analysis of the Rasch measures showed a person mean of .51 (*SD* = 2.49). The Rasch measure person mean was larger than the item mean (-.62), which indicated that the participants performed well on the Hidden Words section.

Table 11 shows the summary of the Modern Language Aptitude Test–Elementary items for section 1: Hidden Words. The item labels represent the sections (1) and the item number within that section. For example, the item 1.06 represents section 1 question 06. Several items showed a problematic fit.

Table 11. *Summary of the Modern Language Aptitude Test–Elementary Items for Hidden Words, Section 1*

Item	Measure	*SE*	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Pt-measure correlation
1.24	4.63	1.13	1.34	.7	.44	.0	.58
1.29	2.99	.75	.41	-1.3	.17	-.6	.81
1.30	2.99	.75	.41	-1.3	.17	-.6	.81
1.25	2.48	.67	.46	-1.4	.23	-.8	.80
1.26	2.48	.67	.46	-1.4	.23	-.8	.80
1.27	2.48	.67	.46	-1.4	.23	-.8	.80
1.28	2.48	.67	.46	-1.4	.23	-.6	.80
1.12	1.41	.54	1.24	.9	1.04	.3	.52
1.21	1.41	.54	.58	-1.6	.40	-1.1	.75
1.11	1.13	.52	1.20	.9	1.58	1.1	.47
1.04	.87	.50	1.87	3.2	2.43	2.3	.20
1.16	.87	.50	1.59	2.3	2.67	2.6	.28
1.20	.87	.50	.97	.0	1.15	.5	.56
1.22	.87	.50	1.10	.5	1.44	1.0	.50
1.23	.63	.49	.83	-.8	.74	-.5	.61
1.03	-1.04	.52	1.34	1.3	2.72	1.8	.22
1.08	-1.32	.54	1.03	.2	2.21	1.3	.35
1.19	-1.63	.57	1.05	.3	.78	.1	.39
1.13	-1.98	.62	.50	-1.6	.27	-.6	.56
1.18	-1.98	.62	.70	-.8	.44	-.3	.48
1.06	-2.40	.68	1.56	1.3	4.44	2.1	-.02
1.09	-2.40	.68	1.22	.6	.96	.4	.26
1.15	-2.40	.68	.59	-1.0	.27	-.5	.49
1.14	-2.94	.79	.65	-.5	.24	-.5	.41
1.17	-2.94	.79	.65	-.5	.24	-.5	.41
1.02	-18	1.06	1.01	.3	.35	-.2	.23
1.07	-3.78	1.06	1.13	.4	.64	.2	.16
1.01	-5.08	1.84		MINIMUM MEASURE			.00
1.05	-5.08	1.84		MINIMUM MEASURE			.00
1.10	-5.08	1.84		MINIMUM MEASURE			.00
M	14.70	-.51	.79	.92	-.1	.99	
SD	8.00	.39	.41	1.2	1.04	1.0	

Excluding the items that were MINIMAL MEASURE—items that all participants answered correctly—the data in Table 11, which is organized from the most to least difficult items, showed an infit MNSQ range of .41 to 1.87. Using productive fit criteria of 0.7-1.3, as recommended by Bond and Fox (2015), 3 of the 30 items (46%), Items 1.16, 1.06, and 1.04, did not display adequate fit on the infit MNSQ statistics. Items 1.04 (*nikl*), 1.16 (*frit*), and 1.06 (*nif*) underfit the Rasch model and were therefore examined. Item

1.04 (*nikl*) represents the U.S. denomination, *nickel* (i.e., 5 cents), and would not be familiar to the participants unless they had spent time in the United States. Items 1.16 (*frit*, i.e. *fright*) and 1.06 (*nif*, i.e., *knife*) both contain the [aɪ] represented by *i*, which differs from the participants' L1 (i.e., a single *i* in the L1 represents [iː]). These three items were removed and the analysis was rerun.

Table 12 shows the revised (i.e., with items removed) summary of the Modern Language Aptitude Test–Elementary items for Section 1: Hidden Words. The new item

Table 12. *Revised Summary of the Modern Language Aptitude Test–Elementary Items for Hidden Words, Section 1*

Item	Measure	*SE*	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Pt-measure correlation
1.24	5.56	1.21	1.67	1.0	.43	.0	.56
1.29	3.53	.86	.40	-1.2	.12	-.6	.81
1.30	3.53	.86	.40	-1.2	.12	-.6	.81
1.25	2.86	.77	.32	-1.6	.13	-.6	.84
1.26	2.86	.77	.32	-1.6	.13	-.6	.84
1.27	2.86	.77	.32	-1.6	.13	-.6	.84
1.28	2.86	.77	.32	-1.6	.13	-.6	.84
1.12	1.49	.60	1.72	1.8	1.60	.9	.50
1.21	1.49	.60	.54	-1.4	.32	-.8	.79
1.11	1.15	.57	1.46	1.4	2.33	1.5	.49
1.20	.84	.54	1.23	.9	2.57	1.7	.54
1.22	.84	.54	1.14	.6	2.40	1.6	.55
1.23	.56	.53	.78	-.9	.59	-.3	.68
1.03	-1.29	.54	1.45	1.5	8.08	3.3	.20
1.08	-1.60	.57	1.18	.6	6.59	2.9	.33
1.19	-1.94	.60	1.19	.6	.80	.2	.40
1.13	-2.33	.65	.42	-1.8	.22	-.6	.56
1.18	-2.33	.65	.64	-.9	.37	-.3	.51
1.09	-2.80	.72	1.33	.8	1.18	.6	.27
1.15	-2.80	.72	.50	-1.2	.21	-.5	.50
1.14	-3.39	.83	.50	-1.0	.16	-.6	.44
1.17	-3.39	.83	.74	-.4	.25	-.4	.39
1.02	-4.27	1.08	1.12	.4	.39	-.1	.22
1.07	-4.27	1.08	1.30	.6	1.27	.7	.12
1.01	-5.61	1.86		MINIMUM MEASURE			.00
1.05	-5.61	1.86		MINIMUM MEASURE			.00
1.10	-5.61	1.86		MINIMUM MEASURE			.00
M	14.70	-.62	.86	.87	-.3	1.27	—
SD	8.2	.39	.47	1.2	1.99	1.1	—

reliability (separation) estimate is .90 (3.05) and the person reliability (separation) estimate is .89 (2.83). The new infit MNSQ had a range of .32 to 1.72, which was an improvement of the prior summary. Items 1.24 (*nvit*), 1.12 (*nme*), 1.11 (*rnj*), and 1.03 (*ansr*) misfit in the revised summary. The newly misfitting items did not initially misfit, and the underfit in the revised summary might be due to the small *N*-size in this study. A small *N*-size can make item fit statistics very sensitive to unusual person responses. After removing the three items that underfit in the initial summary, no other items were removed. The point-measure correlations revealed no low correlations (< .12), and there were no negative point-measure correlations.

Figure 5 shows that the 27 items of the Hidden Words section ranged in difficulty from -5.61 (item 1.10) to 5.56 (item 1.24). The most difficult item was item 1.24, which has the letter string *nvit* (invite) and the choices of *furniture*, *jealous*, *ask*, and *to help out*. The correct answer is *ask*. Note that the first and last letters of the word *invite* are missing in the item. The letter string *nvit* consists of an initial double consonant, which is a lower frequency combination, and this particular combination (i.e., *n-v*) is not found in English. An example of one of the three least difficult items was item 1.01, which has the letter string *rivr* (river) and the choices of *large stream of water*, *a jealous person*, *hill*, and *a dog's name*. The correct answer is *a large stream of water*. Note that for this item, the first and last letters of the word *river* are intact. Moreover, *rivr* is similar to the actual word, *river*, except for the missing vowel *e*. The initial consonant-vowel (i.e., *r-i*) occurs in English more frequently than the initial double consonant of item 1.24 (i.e., *nvit*). The hierarchy of item difficulties supports the ability of the participants, as the item distribution is spread along the variable.

```
-----------------------------------------------------------------------------
 More able participants|  More difficult items
                    X  |T 1.24: nvit
                       |
                       |
   5                   +
                       |
                    X T|
                       |
                       |
   4                   +
                       |
                       |  1.29: atemt   1.3: ansr
                    X  |
                       |
   3                X  +
                       |S 1.25: ne      1.26: pensl 1.27: agre 1.28: alrm
                       |
                      S|
                       |
   2                   +
                       |
                    X  |
                       |  1.12: nme     1.21: lfnt
                       |  1.11: rnj
   1                   +
                  XXX  |  1.20: kmfrt   1.22: pasnjr
                       |  1.23: salr
                    X  |
                      M|
   0             XXXX  +M
                       |
                  XXX  |
                       |
                 XXXX  |
  -1                   +
                       |  1.03: ansr
                       |
                       |  1.08: oshn
                       |
  -2                  S+  1.19: ovn
                   XX  |
                       |  1.13: egl     1.18: buflo
                       |
                       |S 1.09: silns   1.15: shf
  -3                X  +
                       |
                    X  |  1.14: resnt   1.17: hrd
                       |
                       |
  -4                  T+
                       |  1.02: nedl    1.07: tn
                       |
                       |
                       |
  -5                   +  1.01: rivr    1.05: midl    1.10: ruf
Less able participants |  Easier items
-----------------------------------------------------------------------------
```
Figure 5. Item-person map for Modern Language Aptitude Test–Elementary Version, Hidden Words. X = 1 participant. M = mean, S = 1 SD, T = 2 SD.

Table 13 shows the PCA of item residuals results. The Rasch model explained 66.8% of the variance (eigenvalue = 48.3) and the unexplained variance in the first residual contrast accounted for 9.2% of the total variance (eigenvalue = 6.6), so it does not appear to be unidimensional. The Rasch model met the > 50% criterion for explained variance, but did not meet the < 2.0 eigenvalue criterion for the first residual contrast, meaning that the unexplained variance indicates a possible second dimension. The eigenvalues for the first three residual contrasts is high (> 2.0), which means that the unexplained variance of the first three contrasts did not meet the criteria for the contrasts to be at noise level. However, the percentages of variance in the residual contrasts are all < 10%, which indicates that the data is potentially unidemensional.

Table 13. *Standardized Residual Variance (in Eigenvalue Units) for Modern Language Aptitude Test–E, Hidden Words*

	Eigenvalue	% of variance
Total raw variance in observations	72.3	100.0%
Raw variance explained by measures	48.3	66.8%
Raw variance explained by persons	18.4	25.4%
Raw Variance explained by items	23.4	36.7%
Raw unexplained variance (total)	29.9	41.3%
Unexplained variance in 1st contrast	6.6	9.2%
Unexplained variance in 2nd contrast	3.4	4.7%
Unexplained variance in 3rd contrast	3.3	4.6%
Unexplained variance in 4th contrast	2.2	3.1%
Unexplained variance in 5th contrast	1.8	2.4%

Table 14 shows the item loadings for the PCA of item residuals for the MLAT, Section 1 (i.e., Hidden Words). The items (item numbers) with positive residual loadings > .40 include the items *ne* (1.25), *pensl* (1.26), *agre* (1.27), *alrm* (1.28), *atemt* (1.29), *jrne* (1.30), and *nvit* (1.24). The items with negative residual loadings > -.40 include *ansr* (1.03), *oshn* (1.08), *nme* (1.12). The contrast between the top five items with strong positive loadings (i.e., easy items) and the bottom five items with negative loadings (i.e., difficult items) is indicative of a difficulty dimension. The positive loading items all hold a place at the end of the instrument, after the participants had opportunity to adjust to the items on the test. Items with negative loadings come earlier in the instrument, before participants have adjusted to the difficulty of the test. No action was taken because the secondary dimension does not affect the unidimensionality of the test, as item difficulty is not a legitimate dimension.

Table 14. *Item Loadings from the Rasch PCA of Residuals for the Modern Language Aptitude Test–E, Hidden Words*

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
1.25	.91	2.86	.32	.13
1.26	.91	2.86	.32	.13
1.27	.91	2.86	.32	.13
1.28	.91	2.86	.32	.13
1.29	.67	3.53	.40	.12
1.30	.67	3.53	.40	.12
1.24	.63	5.56	1.67	.43
1.21	.21	1.49	.54	.32
1.20	.20	.84	1.23	2.57
1.19	.14	-1.94	1.19	.80
1.23	.09	.56	.78	.59
1.02	.04	-4.27	1.12	.39
1.07	.04	-4.27	1.30	1.27
1.09	.03	-2.80	1.33	1.18
1.03	-.82	-1.29	1.45	8.08
1.08	-.78	-1.60	1.18	6.59
1.12	-.56	1.49	1.72	1.60
1.11	-.35	1.15	1.46	2.33
1.15	-.24	-2.80	.50	.21
1.14	-.22	-3.39	.50	.16
1.13	-.20	-2.33	.42	.22
1.22	-.19	.84	1.14	2.40
1.18	-.10	-2.33	.64	.37
1.17	-.02	-3.39	.74	.25

Matching Words. The result of a dichotomous Rasch analysis using WINSTEPS 3.80 indicated that the MLAT–E section, Matching Words, had an item reliability (separation) estimate of .70 (1.55), and a person reliability (separation) estimate of .61 (1.63). According to the criteria used in this study, the item reliability is fair (.67-.80) and the person reliability is poor (< .67). The results indicate that the instrument is sensitive enough to distinguish between high and low performers, but that there is a poor range of ability. Analysis of the Rasch measures showed a person mean of .49 with a standard deviation of 1.99. The mean person ability estimate was higher than the mean item difficulty estimate (-.51), which indicates that the participants performed well on the instrument overall.

Table 15 shows the summary of the Modern Language Aptitude Test–Elementary items for Section 1: Matching Words. Item labels represent the sections (2) and the item number within that section. For example, item 2.30 represents question 30 in Section 2. Excluding the items labeled MINIMAL MEASURE (i.e., items that all participants answered correctly), the data in Table 15, which is organized from most to least difficult, showed an infit MNSQ range of .54 to 1.38. Using productive fit criteria of 0.7-1.3, as recommended by Bond and Fox (2015), items 2.20 (*policeman,* 1.38) and 2.09 (*cold,* 1.36) underfit the model and were therefore removed and the analysis was run again. A Pearson correlation of the person ability estimates with and without the items was calculated. The two sets of person measures showed high correlation of $r = .99$ ($p < .01$), meaning that retaining the items did not degrade the measurement. Furthermore, because the items showed only slight misfit and because the inclusion of these items in the test was purposeful, the items were kept. Items 2.20 (i.e., *The kind POLICEMAN helps all the children*) and 2.29 (i.e., *Helen WON the prize for the best flowers*) both showed low point-measure correlations (< .12), perhaps because the target syntax of each item (i.e., noun and verb) is simple. The low correlation of the items did not result in the exclusion of the items.

Table 15. *Summary of the Modern Language Aptitude Test–Elementary Items for Matching Words, Section 2*

Item	Measure	*SE*	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Pt-measure correlation
2.30	2.90	.48	.82	-.8	.69	-.7	.52
2.20	2.47	.45	1.38	2.0	1.37	1.1	.03
2.21	2.47	.45	.96	-.2	1.91	2.4	.30
2.25	2.27	.44	1.00	.0	1.74	2.2	.29
2.04	1.51	.44	1.14	.9	1.24	1.0	.26
2.29	1.31	.45	1.12	.8	1.07	.4	.31
2.23	1.11	.46	.99	.0	1.11	.5	.39
2.24	.67	.49	1.08	.4	.98	.1	.35
2.09	.14	.55	1.36	1.1	1.37	.8	.11
2.14	.14	.55	.90	-.2	.82	-.2	.47
2.16	.14	.55	.75	-.7	.61	-.8	.61
2.06	-.19	.59	.98	.1	1.22	.5	.36
2.11	-.19	.59	1.12	.4	1.31	.7	.26
2.15	-.19	.59	.84	-.3	.94	.1	.47
2.27	-.19	.59	.86	-.3	.58	-.6	.54
2.02	-.58	.67	1.00	.2	1.03	.3	.32
2.18	-.58	.67	1.07	.3	.78	-.1	.35
2.05	-1.10	.78	.86	-.1	.99	.3	.37
2.07	-1.10	.78	.79	-.2	.47	-.3	.50
2.08	-1.10	.78	.82	-.1	.58	-.2	.46
2.19	-1.10	.78	.82	-.1	.58	-.2	.46
2.28	-1.10	.78	.76	-.3	.39	-.5	.53
2.01	-1.92	1.06	.54	-.2	.14	-.5	.55
2.12	-1.92	1.06	.89	.2	.24	-.3	.42
2.13	-1.92	1.06	.89	.2	.24	-.3	.42
2.17	-1.92	1.06	.89	.2	.24	-.3	.42
2.03	-3.21	1.83		MINIMUM MEASURE			.00
2.10	-3.21	1.83		MINIMUM MEASURE			.00
2.22	-3.21	1.83		MINIMUM MEASURE			.00
2.26	-3.21	1.83		MINIMUM MEASURE			
M	.03	-.43	.82	.96	.1	.89	—
SD	.94	1.72	.44	.17	.6	.45	—

Figure 6 shows that the 30 items of the Matching Words section exhibited various levels of difficulty. The most difficult item was Item 2.30. Item 2.30 involves reading a

```
----------------------------------------------------------------------------
  More able participants    |  More difficult items
                       XXX  |
                            |
                            |
    3                       +
                            |T 2.30: birds
                        XX S|
                            |
                            |
                            |  2.20: policeman   2.21: soft
                            |
                        XX  |  2.25: drinks
                            |
                            |
    2               XXXXXX  +
                            |
                           M|
                      XXXX  |
                            |
                            |  2.04: winds
                       XXX  |S
                            |  2.29: won
                            |
                         X  |  2.23: bone
    1                       +
                            |
                         X S|
                            |  2.24: sharp
                         X  |
                            |
                            |
                            |
                            |
                            |  2.09: cold    2.14: little     2.16: sister
    0                       +M
                            |
                           T|  2.06: water   2.11: pirates    2.15: child
                         X  |  2.27: gloves
                            |
                            |
                            |  2.02: black   2.18: carry
                            |
                            |
                         X  |
   -1                       +
                            |  2.05: finger  2.07: elephants  2.08: forgot
                            |  2.19: eggs    2.28: Henry
                            |
                            |
                            |S
                            |
                            |
                            |  2.01: black   2.12: insects    2.13: plays
                            |  2.17: fresh
   -2                       +  2.03: father  2.10: shut       2.22: frog
                            |  2.26: friendly
  Less able participants    |  Easier items
----------------------------------------------------------------------------
```

Figure 6. Item-person map for Modern Language Aptitude Test–Elementary Version, Matching Words. X = 1 participant. M = mean, S = 1 SD, T = 2 SD.

target sentence with one word in all capital letters (i.e., *When winter comes the BIRDS fly south*), and then choosing a word in a new sentence (i.e., *Do men still believe the world is flat?*) that performs the same grammatical function. The correct answer for Item 2.30 is *men*. An example of one of the eight least difficult items is Item 2.01, which has the target sentence *A small BOY rang the bell*, and the sentence *Our dog never bites the mailman*, from which to chose a grammatically similar word to the one capitalized in the first sentence (i.e., BOY). The correct answer is *dog*. The range of item difficulty is acceptable for the purpose of the study because the difficulty of the items should be within the participants' abilities.

Table 16 shows the Rasch PCA of item residuals results. The variance explained by measures (34.8%) is poor (i.e., < .5) and did not meet the criteria of > 50%, though this could be explained by a ceiling effect as seen in Figure 6 above (i.e., 43% of the items are easier than the participant ability). In addition, the variance in the person ability estimates was restricted, as shown by the person separation statistic of 15.1%. The Rasch model explained 19.7% of the variance and that the unexplained variance in the first residual contrast accounted variance and that the unexplained variance in the first residual contrast accounted for 5.1 units (12.7%) of the total variance. Furthermore, the model did not meet the criteria for remaining variance of < 2.0, meaning that the unexplained variance indicated a second dimension. The eigenvalues of unexplained variance in the first three contrasts is high (> 2), which means that the unexplained variance of the first three contrasts did not meet the criteria for the contrasts to be at noise level.

Table 16. *Standardized Residual Variance (in Eigenvalue Units) for Modern Language Aptitude Test–E, Matching Words*

	Eigenvalue	% of variance
Total raw variance in observations	39.9	100.0%
Raw variance explained by measures	13.9	34.8%
Raw variance explained by persons	6.0	15.1%
Raw Variance explained by items	7.9	19.7%
Raw unexplained variance (total)	26.0	65.2%
Unexplained variance in 1st contrast	5.1	12.7%
Unexplained variance in 2nd contrast	3.2	8.1%
Unexplained variance in 3rd contrast	2.7	6.7%
Unexplained variance in 4th contrast	2.1	5.3%
Unexplained variance in 5th contrast	1.7	4.2%

Table 17 shows the content of item loadings, including the positive (> .40) and negative (< -.40) loadings for the PCA of Residuals for the MLAT, Section 2 (i.e., Matching Words). Items 2.12, 2.13, 2.17, 2.12, 2.18, and 2.25 had positive loadings. An example item with a positive loading is Item 2.12, which includes the target, *Once upon a time, fierce PIRATES sailed the seas*, and the sentence to choose a similar word from, *The sudden storm blew down many big trees*. The correct answer is *storm*. Items 2.07, 2.06, 2.24, 2.05, and 2.08 had negative loadings. An example item with a negative loading is 2.07, which includes the target, *ELEPHANTS like to eat peanuts*, and the sentence to choose a similar word from, *Gentle rain is good for flowers.* The correct answer is *rain*. The contrast between the top five items with strong positive loadings (i.e., easy items) and the bottom five items with negative loadings (i.e., difficult items) is indicative of a secondary dimension, difficulty. The easy items represent grammatical items that have clear connections between the example sentence and the target sentence. Difficult items represent a less clear connection between the example and the target sentences. Action was not taken because the secondary dimension does not affect the unidimensionality of the test, as difficulty is not a legitimate dimension.

Table 17. *Item Loadings from the Rasch PCA of Residuals for the Modern Language Aptitude Test–E, Matching Words*

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
2.12	.91	-1.92	.89	.24
2.13	.91	-1.92	.89	.24
2.17	.91	-1.92	.89	.24
2.21	.51	2.47	.96	1.91
2.18	.41	-.58	1.07	.78
2.25	.41	2.27	1.00	1.74
2.19	.36	-1.10	1.05	1.05
2.11	.34	-.19	1.12	1.31
2.14	.17	.14	.90	.82
2.15	.15	-.19	.84	.94
2.16	.13	.14	.75	.61
2.20	.07	2.47	1.38	1.37
2.23	.03	1.11	.99	1.11
2.01	-.65	-1.92	.64	.14
2.04	-.49	1.51	1.14	1.24
2.27	-.44	-.19	.86	.58
2.28	-.40	-1.10	.76	.39
2.29	-.40	1.31	1.12	1.07
2.07	-.37	-1.10	.76	.39
2.06	-.32	-.19	.98	1.22
2.24	-.30	.67	1.08	.98
2.05	-.28	-1.10	.86	.99
2.08	-.27	-1.10	.82	.58
2.02	-.14	-.58	1.00	1.03
2.09	-.05	.14	1.36	1.37
2.30	-.04	2.90	.82	.69

Finding Rhymes. The result of a dichotomous Rasch analysis using WINSTEPS 3.80 indicated that the MLAT–E section, Finding Rhymes, had an item reliability (separation) estimate of .87 (2.59), and a person reliability (separation) estimate of .88 (2.74). According to the criteria for this study, the item reliability and the person reliability were good (.81-.90). These results indicate that the instrument is sensitive enough to distinguish between high and low performers, as well as a good range of ability in the performers. Analysis of the Rasch measures showed a person mean of 1.19 with a standard deviation of 1.55. The person mean was larger than the item mean (.00), which indicated that the participants generally performed well on the instrument.

Table 18 shows the summary of the Modern Language Aptitude Test–Elementary items for section 3: Finding Rhymes. The item labels represent the sections (3) and the item number within that section. For example, the Item 3.30 represents question 30 in Section 3.

The data in Table 18, which is organized from the most to least difficult item, showed a poor infit MNSQ range of .62 to 1.55. Using the fit criteria of 0.7-1.3, as recommended by Bond and Fox (2015), 6 of the 45 items misfit the Rasch model; items 3.14, 3.25, and 3.09 underfit and items 3.10, 3.29, and 3.23 overfit. Items 3.14 (*dash*) and 3.25 (*bone*) were examined for removal from the data due to their high amount of variation. The choices given for *dash* (3.14), included *brush*, *push*, *splash*, and *wash*, with the correct answer *splash*. Both *dash* and *splash* are from the fourth 1,000 word frequency-words of English, so they are less frequent than the other words (i.e., *brush*, from the second 1,000 high-frequency words of English; *push* and *wash* from the first 1,000 high-frequency words of English). Item 3.25 (*bone*) included the multiple choices *lawn*, *moon*, *one*, and *own*. *Bone*, *one*, and *own* are all in the first 1,000 high-frequency words of English, and *lawn* and *moon* are in the second 1,000 high-frequency words of English. While the participants should be familiar with these words, perhaps the similarity of the choices caused difficulty. Rhyming is not part of the participants L1, so adjusting to phonemic mapping onto similarly spelled target words might have caused the items to underfit. The two items were removed from the data and the analysis was rerun.

Table 18. *Summary of the Modern Language Aptitude Test–Elementary Items for Finding Rhymes, Section 3*

Item	Measure	*SE*	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Pt-measure correlation
3.38	3.48	1.04	1.08	.4	.98	.5	.57
3.39	2.20	.65	1.06	.3	.91	.2	.50
3.41	2.20	.65	.76	-.5	.43	-.6	.65
3.44	2.20	.65	1.06	.3	.91	.2	.50
3.45	2.20	.65	1.02	.2	2.83	1.8	.42
3.35	1.82	.58	.72	-.8	.45	-.9	.66
3.36	1.82	.58	.78	-.6	.50	-.7	.63
3.43	1.82	.58	1.02	.2	1.26	.6	.46
3.19	1.51	.54	1.11	.5	1.54	1.0	.39
3.27	1.51	.54	1.26	1.0	1.82	1.4	.33
3.40	1.51	.54	1.11	.5	1.50	1.0	.39
3.42	1.51	.54	.95	-.1	1.10	.4	.50
3.17	1.23	.51	1.25	1.0	1.44	1.0	.34
3.37	1.23	.51	1.03	.2	1.01	.2	.47
3.14	.98	.49	1.55	2.2	2.17	2.4	.14
3.15	.98	.49	.98	.0	.85	-.3	.50
3.33	.98	.49	1.00	.1	1.34	.9	.45
3.16	.74	.48	1.30	1.4	1.35	1.1	.31
3.25	.30	.46	1.45	2.2	1.40	1.4	.24
3.31	.30	.46	.72	-1.6	.66	-1.3	.60
3.34	.30	.46	1.24	1.2	1.61	2.0	.29
3.24	.09	.46	.89	-.6	.82	-.7	.52
3.06	-.12	.46	1.34	1.7	1.38	1.5	.26
3.20	-.12	.46	.81	-1.0	.94	-.2	.52
3.32	-.33	.46	.87	-.6	.83	-.6	.50
3.05	-.54	.47	.87	-.6	.80	-.7	.49
3.09	-.54	.47	1.38	1.8	1.36	1.3	.22
3.10	-.54	.47	.67	-1.8	.61	-1.6	.59
3.11	-.54	.47	.70	-1.6	.69	-1.2	.57
3.30	-.54	.47	1.01	.1	.98	.0	.41
3.01	-.76	.47	.77	-1.1	.71	-1.0	.53
3.03	-.99	.48	.92	-.3	.83	-.4	.44
3.28	-.99	.48	1.03	.2	1.06	.3	.37
3.29	-.99	.48	.62	-2.0	.51	-1.7	.60
3.23	-1.23	.50	.62	-1.8	.48	-1.5	.58
3.04	-1.48	.52	.74	-1.0	.53	-1.1	.50
3.21	-1.48	.52	1.03	.2	.87	-.1	.35
3.26	-1.76	.55	1.08	.4	1.27	.6	.26
3.08	-2.08	.59	.85	-.4	.60	-.5	.39
3.12	-2.08	.59	1.01	.2	.69	-.3	.32
3.13	-2.46	.65	.89	-.1	.60	-.3	.33
3.18	-2.46	.65	.91	-.1	.56	-.4	.33
3.02	-2.96	.77	.89	.0	.45	-.3	.30
3.07	-2.96	.77	.93	.1	.54	-.2	.27
3.22	-2.96	.77	1.16	.5	1.06	.4	.12
M	-.04	.55	.99	.0	1.01	.1	—
SD	1.28	.11	.22	1.0	.49	1.0	—

Table 19. *Revised Summary of the Modern Language Aptitude Test–Elementary Items for Finding Rhymes, Section 3*

Item	Measure	*SE*	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Pt-measure correlation
3.38	3.61	1.05	1.12	.4	1.10	.5	.55
3.39	2.30	.66	1.11	.4	.98	.3	.50
3.41	2.30	.66	.71	-.6	.38	-.7	.68
3.44	2.30	.66	1.11	.4	.98	.3	.50
3.45	2.30	.66	1.03	.2	3.55	2.1	.41
3.35	1.91	.59	.69	-.9	.41	-.9	.68
3.36	1.91	.59	.78	-.6	.48	-.7	.64
3.43	1.91	.59	1.03	.2	1.41	.8	.46
3.19	1.58	.55	1.15	.6	1.83	1.4	.39
3.27	1.58	.55	1.34	1.2	1.92	1.5	.32
3.40	1.58	.55	1.13	.6	1.59	1.1	.40
3.42	1.58	.55	.94	-.1	1.09	.4	.52
3.17	1.30	.52	1.36	1.4	1.71	1.4	.31
3.37	1.30	.52	1.05	.3	1.07	.3	.48
3.15	1.04	.50	1.03	.2	.93	.0	.50
3.33	1.04	.50	1.00	.1	1.50	1.2	.47
3.16	.79	.49	1.38	1.7	1.46	1.2	.30
3.31	.34	.47	.72	-1.4	.66	-1.2	.62
3.34	.34	.47	1.28	1.3	1.78	2.3	.30
3.24	.12	.47	.94	-.2	.87	-.4	.51
3.06	-.10	.47	1.39	1.8	1.44	1.5	.27
3.20	-.10	.47	.82	-.9	1.08	.4	.53
3.32	-.32	.47	.88	-.5	.83	-.6	.52
3.05	-.54	.47	.88	-.5	.81	-.6	.51
3.09	-.54	.47	1.47	2.0	1.46	1.5	.22
3.10	-.54	.47	.71	-1.5	.65	-1.3	.59
3.11	-.54	.47	.70	-1.5	.73	-.9	.58
3.30	-.54	.47	1.03	.2	1.00	.1	.43
3.01	-.76	.48	.77	-1.1	.69	-.9	.54
3.03	-1.00	.49	.93	-.3	.83	-.3	.45
3.28	-1.00	.49	1.06	.3	1.16	.5	.37
3.29	-1.00	.49	.61	-2.0	.48	-1.6	.61
3.23	-1.24	.50	.61	-1.9	.46	-1.4	.59
3.04	-1.50	.52	.77	-.9	.53	-.9	.50
3.21	-1.50	.52	1.06	.3	.88	-.1	.35
3.26	-1.78	.55	1.05	.3	1.43	.8	.28
3.08	-2.11	.59	.86	-.3	.60	-.4	.39
3.12	-2.11	.59	1.04	.2	.69	-.2	.32
3.13	-2.49	.65	.88	-.2	.58	-.3	.34
3.18	-2.49	.65	.86	-.2	.49	-.4	.36
3.02	-2.99	.77	.88	.0	.46	-.2	.30
3.07	-2.99	.77	.90	.0	.50	-.1	.29
3.22	-2.99	.77	1.14	.4	.98	.4	.14
M	0.00	.56	.98	.0	1.03	.1	—
SD	1.72	.11	.22	.9	.58	1.0	—

The new item reliability (separation) estimate was .88 (2.67) and the person reliability (separation) estimate was .89 (2.80). The new infit MNSQ had a range of .61 to 1.47, which was an improvement of the prior analysis. Items 3.17 (*prince*), 3.06 (*all*), and 3.09 (*low*) misfit in the revised analysis. The newly misfitting items did not underfit in the first analysis, and the underfit of these items in the revised analysis might be related to the small *N*-size used in this study. A small *N*-size can make the item fit very sensitive to person responses that do not fit the Rasch model. After removing the two items that underfit in the initial analysis, no other items were removed. The point-measure correlations revealed no low correlations (< .12).

Figure 7 shows that the 43 items of the Finding Rhymes section exhibited various levels of difficulty, ranging from -2.99 (e.g., item 3.22) to 3.61 (item 3.38). The most difficult item was item 3.38, which asks the participant to find a rhyme with the word *fire* among the choices of *choir*, *here*, *more*, and *pear*. The correct answer is *choir*. An example of one of the least difficult items is item 3.02, which has the target word *beef*, and the choices, *calf*, *if*, *knife*, *leaf*. The correct answer is *leaf*. The overall range of participants along the variable is acceptable for the purpose of the study as the difficult items at the top of the map match the difficulty of the rhyming targets in relation to the representative concepts. For example, for some second language speakers, it is more difficult to understand the word *choir* (item 3.38) as a rhyme of *fire* than it is to understand that *leaf* is a rhyme of *beef* because the *ea* vowel compound (i.e., long vowel) is more similar to the *ee* vowel compound as the sound [iː] than is the *oir* vowel-consonant string (i.e., diphthong) similar to the *ire* sound as the sound [aɪər].

```
-------------------------------------------------------------------
More able participants |  More difficult items
    4               X  +
                       |
                       |
                       |  3.38: fire
                       |
                       |T
                       |
                       |
    3                  +
                       |
                       |
                       |
                       |
                    X  |
                      T|  3.39: knees  3.41: food   3.44: size
                       |  3.45: put
    2                  +
                       |  3.35: cow    3.36: pillar 3.43: course
                       |S
                       |  3.19: shoot  3.27: warm   3.40: after
                       |  3.42: eight
                       |
                    X  |  3.17: prince 3.37: of
                    X  +  3.15: hole   3.33: folk
                    X  |
                       |  3.16: pull
                    X  |
                    X  |
                   XX  |  3.31: hum    3.34: bead
                       |
                       |  3.24: fist
    0                  +M
                   XX M|  3.06: all    3.20: maid
                       |
                    X  |  3.32: tough
                    X  |  3.05: rain   3.09: low    3.10: say
                       |  3.11: mine   3.30: coast
                       |  3.01: door
                    X  |
   -1              XX  +  3.03: pie    3.28: best   3.29: climb
                    X  |
                      S|  3.23: pass
                       |
                   XX  |  3.04: make   3.21: school
                    X  |
                       |S 3.26: pale
                    X  |
   -2               X  +
                       |  3.08: mix    3.12: mend
                      T|
                       |  3.13: you    3.18: name
                       |
                       |
                       |
   -3                  +  3.02: beef   3.07: face   3.22: rose
Less able participants |  Easier items
--------------------------------------------------------------------
```

Figure 7. Item-person map for Modern Language Aptitude Test–Elementary Version, Finding Rhymes. Person codes are frequency level and frequency level test item. X = 1 participant. M = mean, S = 1 SD, T = 2 SD.

Table 20 shows the Rasch PCA of item residuals results. The raw variance explained by the Rasch measures was low (42.5%), meaning that around 58% of the variance was random. The raw variance explained by the Rasch measures did not meet the Rasch PCA criteria for this study, which is > 50% of total variance explained by the first component. The Rasch model explained 28.0% of the variance and that the unexplained variance in the first residual contrast accounted variance and that the unexplained variance in the first residual contrast accounted for 6.5 units (8.5%) of the total variance. Furthermore, the variance explained by the Rasch measures (42.5%) was poor (i.e., < 50%). The eigenvalues of unexplained variance in each of the contrasts is high (> 2.0), especially in the first contrast (8.5%), which indicates contrasting patterns in the residuals at the strength of around 8 items. The unexplained variance of the five contrasts did not meet the criteria for the contrasts to be at noise level. However, the percentages of variance in the residual contrasts are all < 10%, which indicates that the data is possibly unidemensional.

Table 20. *Standardized Residual Variance (in Eigenvalue Units) for Modern Language Aptitude Test–E, Finding Rhymes*

	Eigenvalue	% of variance
Total raw variance in observations	76.3	100.0%
Raw variance explained by measures	33.3	42.5%
Raw variance explained by persons	11.9	15.7%
Raw Variance explained by items	21.3	28.0%
Raw unexplained variance (total)	43.0	56.4%
Unexplained variance in 1st contrast	6.5	8.5%
Unexplained variance in 2nd contrast	4.6	6.0%
Unexplained variance in 3rd contrast	4.0	5.3%
Unexplained variance in 4th contrast	3.8	5.0%
Unexplained variance in 5th contrast	3.5	4.6%

Table 21 shows the residual loadings, including the positive (> .40) and negative (< -.40) loadings for the MLAT, Section 3: Finding Rhymes. The following items and target/answer rhyme pairs had positive loadings: 3.37 (*of/love*), 3.43 (*course/force*), 3.40 (*after/laughter*), 3.39 (*knees/these*), 3.44 (*size/wise*), 3.32 (*tough/stuff*), 3.42 (*eight/late*), and 3.41 (*food/rude*). The following items and target/answer rhyme pairs had negative loadings: 3.21 (*school/rule*), 3.08 (*mix/kicks*), 3.24 (*fist/kissed*), 3.26 (*pale/sail*), 3.28 (*best/guest*), 3.07 (*face/case*), and 3.06 (*all/wall*). The contrast between the top five items with strong positive loadings (i.e., easy items) and the bottom five items with negative loadings (i.e., difficult items) is indicative of a secondary dimension, difficulty. The easy items represent items that have clear phonetic representations. Difficulty items represent less clear phonetic representations. Action was not taken because the secondary dimension does not affect the unidimensionality of the test, as difficulty is not a legitimate dimension.

Number Learning. The result of a dichotomous Rasch analysis using WINSTEPS 3.80 (Linacre, 2016a) indicated that the MLAT–E section, Number Learning, had an item reliability (separation) estimate of .00 (.00), and a person reliability (separation) estimate of .70 (1.53). According to the criteria used in this study, the item reliability is poor (< .67) and the person reliability is fair (.67-.80). The analysis showed a person mean of 2.35 with a standard deviation of 1.59. The person mean was larger than the item mean (1.25), which indicated that the participants generally performed well on the instrument. Perhaps the poor item reliability is due to the small *N*-size of 25

Table 21. *Item Loadings from the Rasch PCA of Residuals for the Modern Language Aptitude Test–E, Finding Rhymes*

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
3.37	.87	1.30	1.05	1.07
3.43	.79	1.91	1.03	1.41
3.40	.66	1.58	1.13	1.59
3.39	.56	2.30	1.11	.98
3.44	.56	2.30	1.11	.98
3.32	.55	-.32	.88	.83
3.42	.51	1.58	.94	1.09
3.41	.43	2.30	.71	.38
3.15	.38	1.04	1.03	.93
3.35	.34	1.91	.69	.41
3.33	.32	1.04	1.00	1.50
3.03	.25	-1.00	.93	.83
3.38	.20	3.61	1.12	1.10
3.34	.19	.34	1.28	1.78
3.13	.18	-2.49	.88	.58
3.45	.13	2.30	1.03	3.55
3.31	.12	.34	.72	.66
3.18	.11	-2.49	.86	.49
3.16	.10	.79	1.38	1.46
3.29	.07	-1.00	.61	.48
3.20	.05	-.10	.82	1.08
3.21	-.62	-1.50	1.06	.88
3.08	-.50	-2.11	.86	.60
3.24	-.49	.12	.94	.87
3.26	-.47	-1.78	1.05	1.43
3.28	-.47	-1.00	1 .06	1.16
3.07	-.42	-2.99	.90	.50
3.06	-.41	-.10	1.39	1.44
3.11	-.40	-.54	.70	.70
3.30	-.33	-.54	1.03	1.00
3.17	-.32	1.30	1.36	1.71
3.01	-.31	-.76	.77	.69
3.22	-.31	-2.99	1.14	.98
3.10	-.29	-.54	.71	.65
3.27	-.28	1.58	1.34	1.92
3.02	-.24	-2.99	.88	.46
3.23	-.24	-1.24	.61	.46
3.19	-.22	1.58	1.15	1.83
3.09	-.21	-.54	1.47	1.46
3.12	-.20	-2.11	1.04	.69
3.04	-.18	-1.50	.77	.53
3.05	-.17	-.54	.88	.81
3.36	-.10	1.91	.78	.48

coupled with the small number of items (i.e., 25) and the type of test required of the participants. Number Learning requires the participants to memorize nonce names for numbers. The participants then listen to recorded versions of the nonce words randomly repeated, and then attempt to write down the number numerically. The memorization requirement is small (6 words). Of the 25 participants, 18 participants (72%) answered without error, 4 participants (16%), seemingly not paying attention, did not answer, and 3 participants (12%) had 4-6 errors each. While the majority of the participants were able to complete the test without errors, a very few participants were unable to complete the test without making many errors.

Table 22 shows the summary of the Modern Language Aptitude Test–Elementary items for section 4: Number Learning. The item labels represent the sections (4) and the item number within that section. For example, the Item 4.01 represents question 1 within Section 4.

The data in Table 22, which is organized from the most to least difficult, showed a poor infit MNSQ range of .46 to 4.55. Using productive fit criteria of 0.7-1.3, 19 of the 25 items misfit (i.e., items 4.01, 4.08, 4.10, 4.02, 4.03, 4.04, 4.05, 4.07, 4.09, 4.11, 4.12, 4.13, 4.15, 4.16, 4.17, 4.19, 4.21, and 4.14). Four items, 4.01, 4.02, 4.12, and 4.14, underfit the Rasch model and were examined for removal. The test for Number Learning asks participants to learn new names for numbers in the range of 1 to 39. If a participant does not learn one or all of the new names (e.g., not paying attention to the section of the audio where the new names are introduced), then they would do poorly on this section.

Table 22. *Summary of the Modern Language Aptitude Test–Elementary Items for Number Learning, Section 4*

Item	Measure	*SE*	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Pt-measure correlation
4.01	3.11	.97	1.59	1.4	.90	.4	.81
4.06	2.04	1.16	.46	-.7	.19	-.4	.94
4.08	2.04	1.16	.46	-.7	.19	-.4	.94
4.10	2.04	1.16	.46	-.7	.19	-.4	.94
4.02	.20	1.56	4.55	1.9	6.44	2.3	.72
4.03	.20	1.56	.08	-.8	.05	-.8	.97
4.04	.20	1.56	.08	-.8	.05	-.8	.97
4.05	.20	1.56	.08	-.8	.05	-.8	.97
4.07	.20	1.56	.08	-.8	.05	-.8	.97
4.09	.20	1.56	.08	-.8	.05	-.8	.97
4.11	.20	1.56	.08	-.8	.05	-.8	.97
4.12	.20	1.56	4.16	1.8	3.65	1.6	.77
4.13	.20	1.56	.08	-.8	.05	-.8	.97
4.15	.20	1.56	.08	-.8	.05	-.8	.97
4.16	.20	1.56	.08	-.8	.05	-.8	.97
4.17	.20	1.56	.08	-.8	.05	-.8	.97
4.19	.20	1.56	.08	-.8	.05	-.8	.97
4.21	.20	1.56	.08	-.8	.05	-.8	.97
4.14	-1.56	1.13	2.95	2.4	9.90	4.8	.52
4.18	-1.56	1.13	.90	.0	.41	-.1	.82
4.20	-1.56	1.13	.90	.0	.41	-.1	.82
4.22	-1.56	1.13	.90	.0	.41	-.1	.82
4.24	-1.56	1.13	.90	.0	.41	-.1	.82
4.25	-1.56	1.13	.90	.0	.41	-.1	.82
4.23	-2.63	1.00	.88	-.7	.40	-.1	.67
M	0.00	1.36	.84	-.2	.98	-.1	—
SD	1.36	.23	1.22	1.0	2.29	1.3	—

If they learn the new names, they would attain maximum measures for most items. The reason for the extreme underfit of the four items is unclear but is possibly related to both participants' minimum or, equally, maximum scores for those items. The four items were removed from the data.

Table 23 shows the revised summary of the Modern Language Aptitude Test–Elementary items for section 4: Number Learning with items 4.01, 4.02, 4.12 and 4.14 removed. The new item reliability (separation) estimate was .00 (0.00) and the person reliability (separation) estimate was .00 (0.00). The new infit MNSQ had a range of .71 to 1.24, which was an improvement of the prior summary for underfit items. Items 4.03 to

4.17, 4.19, and 4.21 had a MAXIMUM MEASURE reading, which means that no participants answered these items correctly. The point-measure correlations revealed no low correlations (< .12) and no negative point-measure correlations.

Table 23. *Revised Summary of the Modern Language Aptitude Test–Elementary Items for Number Learning, Section 4*

Item	Measure	*SE*	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Pt-measure correlation
4.03	1.74	1.92		MAXIMUM	MEASURE		.96
4.04	1.74	1.92		MAXIMUM	MEASURE		.96
4.05	1.74	1.92		MAXIMUM	MEASURE		.96
4.06	1.74	1.92		MAXIMUM	MEASURE		.96
4.07	1.74	1.92		MAXIMUM	MEASURE		.96
4.08	1.74	1.92		MAXIMUM	MEASURE		.96
4.09	1.74	1.92		MAXIMUM	MEASURE		.96
4.10	1.74	1.92		MAXIMUM	MEASURE		.96
4.11	1.74	1.92		MAXIMUM	MEASURE		.96
4.13	1.74	1.92		MAXIMUM	MEASURE		.96
4.15	1.74	1.92		MAXIMUM	MEASURE		.96
4.16	1.74	1.92		MAXIMUM	MEASURE		.96
4.17	1.74	1.92		MAXIMUM	MEASURE		.96
4.19	1.74	1.92		MAXIMUM	MEASURE		.96
4.21	1.74	1.92		MAXIMUM	MEASURE		.96
4.18	.24	1.24	1.24	.6	1.27	.7	.84
4.20	.24	1.24	.71	-.6	.90	.87	.90
4.22	.24	1.24	1.24	.6	1.27	.7	.84
4.24	.24	1.24	1.24	.6	1.27	.7	.84
4.25	.24	1.24	.71	-.6	.67	-.6	.90
4.23	-1.21	1.24	.89	-.1	.84	-.2	.79
M	1.24	1.25	1.72	-.2	.98	-.1	—
SD	.86	.23	1.22	1.0	2.29	1.3	—

Figure 8 shows that the 21 items of the Number Learning section ranged in difficulty from -1.21 (e.g., item 4.23) to 1.74 (item 4.03). The most difficult item was Item 4.03, which asks the participant to equate a learned nonce word (i.e., *chilkacholl*) with a number (i.e., 11). An example of one of the least difficult items is Item 4.23, which has the nonce target word *chilkacholl*, and the equivalent numerical answer 11.

```
----------------------------------------------------------------------------
More able participants |  More difficult items
      1   .##########  +  4.03: chilkacholl 4.04: rosskaross
                       |  4.05: bimkacholl  4.06: ross        4.07: chilkabame
                       |  4.08: rosska      4.09: bimkabame   4.10: choll
                       |  4.11: rosskacholl 4.13: rosskabame
                       |  4.15: chilkaross  4.16: rosskaross
                       |  4.17: bimkacholl  4.19: bimkabame
                       |  4.21: rosskabame
                       |
                       |S
                       |
                       |
                       |
                       |
                       |
                      T|  4.18: chilkabame  4.20: chilkaross
                       |  4.22: bimkaross   4.24: bimka       4.25: bame
                       |
                       |
                       |
      0             .  +M
                       |
                      S|
                       |
                       |
                       |
                       |
                       |
                       |
                       |
                      M|
                       |S
                       |
                       |
                       |
                    #  |
                       |
                      S|
                       |
                       |
     -1                +
                       |
                       |T
                       |
                      T|  4.23: chilkacholl
                       |
Less able participants |  Easier items
---------------------------------------------------------------------
```

Figure 8. Item-person map for Modern Language Aptitude Test–Elementary Version, Number Learning. Choll = 1; bame = 2; ross = 3; chilka = 10; bimka = 20; rosska = 30. # = 2 participants. '.' = 1 participant. M = mean, S = 1 SD, T = 2 SD.

That both the most difficult and the least difficult items were exactly the same suggests that the participants learned or remembered numbers as they took the test. The item distribution of the learned numbers is bottom-heavy, which means that the items bunch up at the top of the Wright map (Figure 8). Only 3 of the 25 participants were unable to learn the numbers.

Table 24 shows the PCA of item residuals results. The variance explained by measures (8.9%) was poor (i.e., < 50%). This could be the result of the small *N*-size and the way that the participants either memorized or did not memorize the new numbers. The Rasch model explained 8.9% of the variance and that the unexplained variance in the first residual contrast accounted for 3.9 eigenvalue units (91.1%) of the total variance.

Table 24. *Standardized Residual Variance (in Eigenvalue Units) for Modern Language Aptitude Test–E, Number Learning*

	Eigenvalue	% of variance
Total raw variance in observation	6.6	100.0%
Raw variance explained by measures	.6	8.9%
Raw variance explained by persons	.1	.9%
Raw Variance explained by items	.5	8.0%
Raw unexplained variance (total)	6.0	91.1%
Unexplained variance in 1st contrast	3.9	58.7%
Unexplained variance in 2nd contrast	2.1	32.4%
Unexplained variance in 3rd contrast	.0	.0%

Table 25 shows the content of the positive (> .40) and negative (< -.40) loadings for the PCA of item residuals for the MLAT, Section 4 (i.e., Number Learning). The following items had positive loadings (answers): 4.18 (12) and 4.24 (20). The following items had negative loadings (answer): 4.23 (11), 4.20 (13), and 4.25 (2). The contrast between the top five items with strong positive loadings > .40 and the bottom five items with negative loadings > -.40 (i.e., difficult items) is indicative of a secondary dimension, difficulty. The difficult items were difficult for participants who failed to learn the representative number-sounds. No items were deleted because the secondary dimension of difficulty does not affect the unidimensionality of the test.

Table 25. *Item Loadings from the Rasch PCA of Residuals for the Modern Language Aptitude Test–E, Number Learning*

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
4.18	.96	.24	1.24	1.27
4.24	.96	.24	1.24	1.27
4.23	-.89	-1.21	.89	.84
4.20	-.78	.24	.71	.67
4.25	-.78	.24	.71	.67
4.22	.05	.24	1.24	1.27

Non-Standardized Instruments

Three non-standardized instruments—the Listening in English Questionnaire, the Elicited Imitation Test(s), and the Background and Length of Residency Interview —were created to assess aspects of the participants' comprehension of Sandhi-variation and their English learning backgrounds. These non-standardized instruments were created for this study.

Pre-Listening in English Questionnaire

The purpose of the Pre-Listening in English Questionnaire (Appendix B) was to elicit the participants' general feelings about English, the English sound system, and their history with English. Validity evidence for the Pre-Listening in English Questionnaire was gathered using the following procedures. A Japanese language instructor, teaching at the university level, who has native-level English proficiency, translated the questions into Japanese from English. The translation was done using a verbal protocol (i.e., the translation was done aloud) in order to verify meaning of the English while translating. Two bilingual university administration staff members then back-translated the form to check the original translation. All of the items were confirmed to express the intended meanings. The questionnaire was then given to two university students with similar backgrounds to the participants. The two university students used a verbal protocol in their first language, Japanese, while answering the questions. No changes were made to the Pre-Listening in English Questionnaire based on the verbal protocol. The questionnaire was then piloted along with the other instruments.

Elicited Imitation Tests

The Elicited Imitation Test consisted of 48 sentences (Appendix C), using both robust elision and assimilation and the absence of elision and assimilation (i.e., each word within the target utterance is enunciated). All of the sentence sets vary in length between six and eight syllables. Perkins, Brutten, and Angelis (1986) found that between seven and eleven syllables is acceptable for second language speakers to repeat. Due to the inclusion of the Sandhi-variation phenomenon, which adds another dimension of difficulty to both the target utterances and to the speaking proficiency of the participants, utterances on the test vary between six to eight syllables. Furthermore, in pre-pilot tests, sentences with nine or more syllables were difficult for non-native, bilingual speakers.

The Sandhi-variation form defines the syllable count when it is present in the sentence. Therefore, the item *Howzit going today?* is defined as having six syllables, whereas the fully enunciated form (e.g., *How is it going today?*) has seven syllables.

The participants repeat the sentence exactly as they hear it and then repeat the phrase again, being careful to enunciate each word.

The first target set of utterances for imitation consisted of 24 sentences, 12 with Sandhi-variation and 12 without Sandhi-variation. The 12 sentences with Sandhi-variation are divided further with the inclusion of either *elision* (6 syllable sentences = 6, 8, & 47; 8 syllable sentences = 10, 37, & 43) or *assimilation* (6 syllable sentences = 23, 24, & 38; 8 syllable sentences = 3, 13, & 33), two common aspects of Sandhi-variation in English. The length of each utterance in the first elicited imitation instrument also varied. To be conservative due to the Sandhi-variation phenomenon, which adds another dimension of difficulty to both the target utterances and to the speaking proficiency of the participants (i.e., having to imitate Sandhi-variation), utterances that varied between six to eight syllables were used in this study.

The second set of target utterances (randomly distributed with the first set of target utterances) is controlled by syllable length, but increases the speed of the utterance to replicate native conversation. Again there were 24 total sentences. The added time pressure structure of this test was chosen to replicate L1 native speaking rate in order to replicate authentic utterances. The average speaking rate for lectures given to non-native speakers is 140 words per minute or 3.17 syllables per second. This lecture rate for non-natives is slower than the average conversational speech rate for native English speakers of around 4.7 syllables per second (Tauroza & Allison, 1990). For this second test, there are 24 sentences that varied in length between six and eight syllables, include elision or assimilation (Table 26), and are produced at 4.7 syllables per second.

The instructions on the Elicited Imitation Test are written in the participants' L1 to insure they understand the task. To make sure that the instructions were clear, the instructions were shown to four students who did not participate in the study; these students showed comprehension by explaining the procedures in their own words. The test-takers might or might not be familiar with elicited imitation as a test type; thus, providing an adequate number of practice items is necessary to insure that the participants understand what they are being asked to perform. The first pilot test used five example types. The examples were written in their complete form on the instruction sheet. Participants read the instructions prior to hearing the same instructions and examples. The examples were also repeated on the audio recording. Participants can read along as they hear the examples aurally, which had a second benefit of introducing the voice used on the instrument. The first two examples included a modeled explanation about how to repeat and then enunciate the target sentence after a signal tone. The third, fourth, and fifth examples had a 14-second silence generation after each target prompt, which replicated the items on the test. After an initial 6 seconds of the 14-second total, a 0.5 second tone generation was heard. The tone generation signals the division of the allotted 14 seconds into the two parts of the answer, repetition and enunciation. After the tone, 8 seconds was provided for the enunciation portion as early attempts showed that the enunciation section required 2 seconds more than the repetition section to complete. After the final example, the participants signaled understanding or lack of understanding prior to the beginning of the target items.

Table 26 shows the randomized distribution of Sandhi-variation within the Elicited Imitation Test targets. The letters A and E represent assimilation (A) and elision (E). The positive (+) and negative (-) symbols before the letters A and E represent the inclusion of assimilation and elision (+) or the absence of potential assimilation and elision (-) within each target utterance. The numbers 6 and 8 represent the length of the utterance in syllables. Finally, the letter S represents the rate of the utterance: the +S means that the sentence is uttered at native speed (4.7 syllables per second), while the –S means that the rate of the utterance is less than native speed (3.17 syllables per second).

Table 26. *Distribution of Sandhi-variation*

Sentence type	Sentence number
+A6–S	23, 24, 38
-A6–S	28, 35, 45
+A6+S	2, 12, 41
-A6+S	22, 40, 36
+E6–S	6, 8, 47
-E6–S	15, 31, 21
+E6+S	5, 18, 30
-E6+S	9, 19, 46
+A8–S	3, 13, 33
-A8–S	48, 17, 11
+A8+S	7, 14, 26
-A8+S	20, 32, 34
+E8–S	10, 37, 43
-E8–S	1, 27, 29
+E8+S	39, 25, 16
-E8+S	4, 42, 44

The items on the Elicited Imitation Test are made up of vocabulary from the high-frequency 2,000 words of English according to the BNC/COCA list. Also, each sentence is a complete, simple sentence. There are no compound or complex sentences.

The type of scoring of elicited imitation tests has been shown to influence the tests' reliability to distinguish between participants across proficiency levels (Yan et al., 2015). Three common scoring methods, binary, ordinal, and internal or mixed scoring, have been used in prior elicited imitation studies. Ordinal scoring was used for stage 1 of the test, based on Erlam (2006). Scoring followed three criteria:

1. Obligatory occasion created—supplied (2 points);

2. Obligatory occasion created—not supplied (1 point);

3. No obligatory occasion created.

For stage 2 of the test, binary scoring was used, which followed two criteria:

1. Obligatory occasion created—supplied (1 point);

2. No obligatory occasion created.

The scores from stage 1 and stage 2 were combined to form composite scores for each item. The pilot was given to three native English speakers, and each target was responded to accurately for both the repetition and enunciation parts of each item.

The results of a partial credit Rasch analysis indicated that the Elicited Imitation Test had an item reliability (separation) estimate of .83 (2.24), and a person reliability (separation) estimate of .88 (2.71), which are both good (.81-.90) according to the criteria used in this study. These results indicated that the instrument is sensitive enough to distinguish between high and low performers. The analysis showed a person mean of .00 with a standard deviation of .99. The person mean was larger than the item mean, which indicated that the participants performed well on the instrument.

Table 27 shows the outcome of each category of the partial credit analysis. A fourth score of zero (0 or no credit) was given for no response (i.e., no obligatory occasion created). As described further in Chapter 3, the ratings for each item were based on each participant's ability to repeat the target correctly two times, which garnered a total possible rating of 3. Partial credit was given to the participant if errors were made in the repetition. *Partly correct* signifies an utterance (i.e., an attempt) from the participant at the initial repeat (there are two) that is incorrect. *More partly correct* signifies either an utterance attempt at the initial repeat that is incorrect and a correct second repeat. Another possibility for *More partly correct* is a correct initial repeat and an incorrect second repeat. *Fully correct* signifies a correct initial repeat and a correct second repeat.

Table 27. *Item Difficulty Measure for Rasch Partial-Credit*

Category	Label	Observed count %	Infit MNSQ	Outfit MNSQ	Andrich threshold	Category measure
1	Partly correct	4 (16%)	.95	.90	None	-2.17
2	More partly correct	12 (48%)	.80	.60	-.83	-.12
3	Fully correct	9 (36%)	.96	.97	.83	1.94

Table 28 shows the summary of Elicited Imitation Test items. The items showed an infit MNSQ range of .54 to 1.87, with criteria, as recommended by Bond and Fox (2015), set between .7 and 1.3. Items +E6+SEIT30 (Infit MNSQ = 1.87) and +E8+SEIT25 (Infit MNSQ = 1.87) underfit and were removed and the analysis was run again. A Pearson correlation of the person ability estimates for both inclusion and exclusion of the items in the data was calculated. The two sets of person measures showed a high correlation ($r = .99$, $p < .01$), meaning that keeping both items would not degrade the measurement. Furthermore, the items inclusion in the test was not a random choice, so the items were kept. Three items had low point-measure correlations (< .12), which can indicate an item with weak differentiation from one level to the next. Items –A8+SEIT34 (i.e., *Does its shape change all of the time?*), –A6–SEIT28 (i.e., *I have a good neighbor*), and –E8+SEIT4 (i.e., *I have never seen your family*) were easy because the low lexical complexity of each sentence and absence of Sandhi-variation. One item

Table 28. *Summary of the Rasch Statistics for the Elicited Imitation Test Items*

Item	Measure	*SE*	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Pt-measure correlation
+E8+SEIT16	1.65	.34	.78	-.6	.83	-.4	.58
+A6+SEIT41	1.60	.36	.98	.1	.68	-.3	.56
+E8–SEIT43	1.45	.33	.81	-.5	.84	-.4	.58
-A6–SEIT45	1.28	.30	.76	-.7	.58	-1.0	.69
-A8–SEIT48	1.24	.30	1.12	.5	1.04	.2	.40
+A6+S EIT2	1.20	.30	1.11	.5	.99	.1	.40
-E8+SEIT44	.91	.44	.54	-.9	.50	-.8	.82
-E8–SEIT29	.79	.24	.84	-.6	.66	-.8	.60
+E6+SEIT18	.75	.26	1.07	.4	1.10	.4	.39
-E6+SEIT46	.75	.26	.88	-.5	.78	-.7	.54
+A6+SEIT12	.62	.25	.70	-1.6	.63	-1.4	.68
+A6–SEIT38	.57	.25	1.06	.4	1.01	.1	.39
+A8+SEIT14	.57	.25	.85	-.8	.87	-.4	.54
+E6–S EIT6	.57	.39	.83	-.3	.90	-.1	.54
+E6+SEIT30	.55	.36	1.87	1.6	2.44	2.1	-.14
+E8+SEIT25	.55	.36	1.87	1.6	2.44	2.1	.14
-E8–S EIT27	.53	.23	.92	-.4	.82	-.4	.50
+E8+SEIT39	.47	.26	1.15	.8	1.11	.5	.30
-A8+SEIT20	.39	.25	1.02	.2	.97	.0	.40
+A8–SEIT33	.33	.41	.83	-.1	.77	-.1	.63
+A8+SEIT26	.28	.37	.96	.1	1.08	.3	.46
+E8–S EIT37	.27	.35	.68	-.7	.64	-.7	.73
-A8+SEIT34	.22	.33	1.38	1.1	1.45	1.3	.08
+E6–SEIT8	.15	.24	1.08	.5	1.01	.1	.34
-A8+SEIT32	.07	.31	1.30	.9	1.17	.5	.28
-E6+SEIT19	-.09	.21	1.14	.7	1.49	1.5	.31
+A6–SEIT23	-.12	.31	.94	-.2	.90	-.3	.42
-E6+SEIT9	-.18	.26	.95	-.2	.84	-.4	.39
+E6+SEIT5	-.22	.22	.99	.0	.89	-.3	.46
+E6–SEIT47	-.28	.25	1.09	.5	.99	.1	.28
-A8–SEIT11	-.35	.27	.93	-.3	.98	.1	.37
+A8–SEIT3	-.44	.29	1.10	.5	1.36	1.2	.18
-E6–SEIT31	-.44	.26	.94	-.2	.83	-.1	.36
+A8–SEIT13	-.50	.24	.96	-.1	.94	-.2	.44
+A8+SEIT7	-.51	.21	1.24	1.1	1.20	.5	.25
+E8–SEIT10	-.52	.22	.94	-.3	.86	-.3	.50
-E6–SEIT21	-.52	.28	.93	-.2	.76	-.4	.38
-E8–SEIT1	-.71	.28	.97	.0	.85	-.2	.35
-A8–SEIT17	-.75	.30	1.02	.2	.85	-.1	.29
-A6+SEIT22	-.81	.31	1.06	.3	1.01	.2	.21
-E6–SEIT15	-.81	.31	.98	.1	.90	.2	.26
-A6–SEIT35	-.96	.34	1.05	.3	.98	.1	.20
-A6–SEIT28	-1.02	.35	1.11	.4	1.20	.5	.11
+A6–SEIT24	-1.09	.38	1.03	.3	.75	.3	.19
-E8+SEIT4	-1.41	.44	1.05	.3	1.01	.3	.11
-A6+SEIT40	-1.47	.52	.96	.3	.47	.0	.20
-A6+SEIT36	-2.21	.74	1.03	.3	.98	.2	.08
-E8+SEIT42	-2.96	1.03	.97	.3	.65	.1	.19
M	.00	.31	1.02	.1	.99	.1	—
SD	.84	.09	.24	.6	.37	.7	—

showed negative correlation, +E6+SEIT30: *We are grateful for it*; the item was difficult because of the quantity and complexity of the [r] sounds within the target. Table 28 is organized from the most to the least difficult item.

As shown in Figure 9, the 48 items of the Elicited Imitation Test exhibited various levels of difficulty. The most difficult item was Item 16 (+E8+SEIT16), which is positive for elision, is 8 syllables in length, and is spoken at native speed (*Your joint effort to win paid off*; elision = dʒɔɪn e.fərt). The least difficult item was Item 42 (–E8+SEIT42), which is negative for elision, is 8 syllables in length, and is spoken at native speed (*Who is your regular doctor?* elision potential = huz). Several items appeared misplaced, for example, Item 45 (–A6–SEIT45), which is negative for assimilation, is 6 syllables in length, and is spoken at less than native speed (making it presumably less challenging than the least difficult item, –E8+SEIT42, which was spoken at native speed). Item 45 (*It is painted bright blue*) is the fourth most difficulty item on the test. Participants had difficulty parsing *bright* and *blue*, despite both words being in the first 1,000 high frequency words of English. There are gaps in the item difficulties of > 0.5 logits; thus, the test does not measure all participant abilities equally well. Overall, the range of participants along the variable is acceptable for the purpose of the study because the difficult items at the top of the map match the difficulty of the Sandhi-variation targets in relation to the representative concepts (e.g., targeted inclusion of elision). For example, for second language speakers, a target positively elided, with 8 syllables, spoken at native speed, such as item +E8+SEIT16 (e.g., *Your joint effort to win paid off*) was more difficult than an item such as –E8+SEIT42 (e.g., *Who is your regular doctor?*), which was not elided, was spoken at native speed, and contained 8 syllables.

```
----------------------------------------------------------------------------
More able       |  More difficult items
participants    |
    3           +
                |
                |
            X   |
                |
                |
                |
                |
    2           +
                |
                |
              T|T +E8+SEIT16
                |  +A6+SEIT41
            X   |  +E8-SEIT43
                |  -A6-SEIT45
                |  +A6+SEIT2    -A8-SEIT48
    1         S+
            X   |S -E8+SEIT44
           XX   |  +E6+SEIT18   -E6+SEIT46  -E8-SEIT29
            X   |
           XX   |  +A6-SEIT38   +A6+SEIT12  +A8+SEIT14  +E6-SEIT6    +E6+SEIT30
                |  +E8+SEIT25  -E8-SEIT27
          XXX   |  +E8+SEIT39  -A8+SEIT20
           XX M|  +A8-SEIT33   +A8+SEIT26  +E8-SEIT37
            X   |  +E6-SEIT8    -A8+SEIT34
    0     XXX  +M -A8+SEIT32
            X   |  +A6-SEIT23   -E6+SEIT19
          XXX   |  +E6-SEIT47   +E6+SEIT5   -E6+SEIT9
         XXXX S|  +A8-SEIT3    -A8-SEIT11
                |  +A8-SEIT13   +A8+SEIT7   -E6-SEIT31
                |  +E8-SEIT10   -E6-SEIT21  -E8-SEIT1
                |  -A6+SEIT22   -A8-SEIT17
                |S -E6-SEIT15
   -1         T+  -A6-SEIT35
                |  +A6-SEIT24   -A6-SEIT28
                |
                |
                |  -A6+SEIT40  -E8+SEIT4
                |T
                |
                |
   -2           +
                |
                |  -A6+SEIT36
                |
                |
                |
                |
                |  -E8+SEIT42
   -3           +
Less able       | Easier items
Participants    |
----------------------------------------------------------------------------
```

Figure 9. Item-person map for Elicited Imitation Test instrument. Person codes are ±assimilation (A) or elision (E), n = native speaking rate, EIT = Elicited Imitation Test, # = test item number, X = 1 participant. M = mean, S = 1 SD, T = 2 SD.

Table 29 shows that the raw variance explained by the Rasch measures is low (41.6%), meaning that around 60% of the variance is random. The raw variance explained by measures (41.6%) did not meet the Rasch PCA criteria for this study, which is > 50% of the total variance. The PCA of item residuals indicated that the unexplained variance in the first residual contrast accounted for 4.8 eigenvalue units (6.0%) of the total variance. The eigenvalues of unexplained variance in each of the contrasts is high (> 2.0), especially in the first contrast, which indicates contrasting patterns in the residuals at the strength of around 5 items. The unexplained variance of the five contrasts did not meet the criteria for the contrasts to be at noise level. The PCA of item residuals indicates the presence of additional constructs in the data; however, the percentages of variance in the residual contrasts are all < 10%, which indicates that the data is unidemensional.

Table 29. *Standardized Residual Variance (in Eigenvalue Units) for the Elicited Imitation Test*

	Eigenvalue	% of variance
Total raw variance in observations	80.9	100.0%
Raw variance explained by measures	32.9	41.6%
Raw variance explained by persons	9.1	11.6%
Raw Variance explained by items	23.7	29.4%
Raw unexplained variance (total)	48.0	59.4%
Unexplained variance in 1st contrast	4.8	6.0%
Unexplained variance in 2nd contrast	4.7	6.0%
Unexplained variance in 3rd contrast	4.3	5.5%
Unexplained variance in 4th contrast	4.2	5.3%
Unexplained variance in 5th contrast	3.5	4.5%

Table 30 shows the item loadings, including the positive (> .40) and negative (< -.40) loadings for the PCA of Residuals for the Elicited Imitation Test. Items +E8+SEIT16, +E6–SEIT6, -A8+SEIT34, +E8+SEIT39, and +E8–SEIT43 had positive loadings, which

Table 30. *Item Loadings from the Rasch PCA of Residuals for the Elicited Imitation Test*

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
+E8+SEIT16	.59	1.58	.78	.83
+E6–SEIT6	.53	.50	.83	.90
-A8+SEIT34	.46	.15	1.38	1.45
+E8+SEIT39	.46	.40	1.15	1.11
+E8–SEIT43	.44	1.38	.81	.84
-A6+SEIT36	.37	-2.28	1.03	.98
+E6–SEIT47	.32	-.29	1.09	.99
-E6–SEIT31	.32	-.50	.94	.83
-A6–SEIT28	.31	-1.09	1.11	1.20
-E8+SEIT44	.31	.84	.54	.50
+A8+SEIT14	.29	.51	.85	.87
-A6+SEIT22	.28	-.81	1.06	1.01
-E6–SEIT15	.26	-.88	.98	.90
+A8–SEIT13	.23	-.51	.96	.94
-E8–SEIT1	.23	-.58	.97	.85
+A6–SEIT38	.22	.51	1.06	1.01
-A8–SEIT48	.15	1.18	1.12	1.04
+E6+SEIT30	.15	.48	1.87	2.44
+E8+SEIT25	.15	.48	1.87	2.44
+A8+SEIT26	.13	.22	.96	1.08
-E8–SEIT27	.13	.46	.92	.82
+E8–SEIT10	.09	-.57	.94	.86
-A8–SEIT11	.04	-.34	.93	.98
-A6–SEIT35	.03	-1.02	1.05	.98
-E6+SEIT19	.00	-.15	1.14	1.49
-E6–SEIT21	-.59	-.58	.93	.76
-A8–SEIT17	-.56	-.77	1.02	.85
-E6+SEIT46	-.54	.69	.88	.78
-A6–SEIT45	-.49	1.22	.76	.58
-E8–SEIT29	-.49	.73	.84	.66
+A6–SEIT2	-.44	1.13	1.11	.99
+A6–SEIT24	-.41	-1.15	1.03	.75
-A6+SEIT40	-.37	-1.53	.96	.47
+A6+SEIT12	-.34	.56	.70	.63
-A8+SEIT32	-.34	.01	1.30	1.17
+E6+SEIT18	-.26	.69	1.07	1.10
+A8–SEIT33	-.25	.26	.83	.77
+E6–SEIT8	-.22	.08	1.08	1.01
-E6+SEIT9	-.21	-.24	.95	.84
-A8+SEIT20	-.20	.33	1.02	.97
-E8+SEIT4	-.12	-1.47	1.05	1.01
+E8–SEIT37	-.11	.20	.68	.64
+A6–SEIT23	-.10	-.18	.94	.90
+A6+SEIT41	-.08	1.53	.98	.68
+A8+SEIT7	-.08	-.56	1.24	1.20
+A8–SEIT3	-.02	-.41	1.10	1.36
+E6+SEIT5	-.02	-.24	.99	.89

were predominantly positive for Sandhi-variation or, when not positive for Sandhi-variation, are 8 syllables in length. Items -E6–SEIT21, -A8–SEIT17, -E6+SEIT46, -A6–SEIT45, -E8–SEIT29, +A6–SEIT2, and +A6–SEIT24 had negative loadings, which are predominantly absent of Sandhi-variation or shorter in length (i.e., 6 syllables). The contrast between the top five items with strong positive loadings (i.e., easy items) and the bottom five items with negative loadings (i.e., difficult items) is indicative of a secondary dimension, difficulty. The easy items represent items that are absent of Sandhi-variation. The difficult items represent positive Sandhi-variation. No items were removed.

Background and Length of Residency Interview (Verbal Protocol)

The Background and Length of Residency Interview (Appendix C) was created to explore specific types of English language input that might have been more readily present in the day-to-day lives of the learners, and therefore might become more quickly acquired. Questions for the Background and Length of Residency Interview were formulated to gather information about the learners' English language experiences, and were contingency-based with a variety of question types posed, including multiple-choice and open-ended. The initial list of questions in English was submitted to six second language lecturers for review. This review generated revisions to the original questions, for example, by adding the possibility of a dormitory stay to their background choices.

The finalized interview questions were translated from English into Japanese, with several passes at the translation to gain optimal precision. A Japanese language instructor, teaching at the university level, who has native-level proficiency of English, translated the interview questions from English to Japanese. The translation was done using a verbal protocol (i.e., done aloud) in order to verify meaning of the English while translating. Two other Japanese language instructors reviewed the translation and made suggestions. The questions, translated into Japanese, were then submitted to four learners of the same age and relative proficiency as the target participants. The questions were comprehended without difficulty. The translation of these questions was finalized after the pilot study.

In the pilot stage, the participants took part in a verbal protocol (i.e., completed aloud) as they completed the questions in their L1 in order to provide information about their understanding of the questions and why they responded in the way that they did. At this stage of the pilot, no further changes were made to the questions.

The interview process, including questions, answers, and the verbal protocol (i.e. as the participants thought through their answers orally), were recorded and transcribed before being translated from Japanese to English. Recordings were done using an Olympus Voice-Tek DS-60 as the primary recording device, and Audacity 2.1.0 for Mac and a MacBook Air as the backup recording device. Due to large number of interview questions and the possibility that the participants might feel tired or demotivated during the interview, the interview process was tested as a divided interview, as two parts, one week apart. The delay was not an issue to the data gathering because the information is not time sensitive. However, participant feedback during the verbal protocol showed that they were not tired or demotivated. Therefore, it was decided that the interview would be one part.

A Japanese L1 speaker with advanced English (TOEFL score = 610) interviewed the participants in Japanese. Prior to the interview, two norming interviews were completed as trial runs. These norming interviews were done with two students within the English language program who were not a part of the study. The norming interviews provided practice for types of follow-up questions asked by the interviewer. All interviews were done in an empty university classroom during a school holiday.

CHAPTER 5

RESULTS

In this chapter I present the results of the analyses for the four research questions that guide this study. The first research question is divided into two parts: Question 1a concerns the empirical hierarchy for the phonological features of elision and assimilation, and question 1b how input rates, combined with the phonological features of elision and assimilation, affected the participants' decoding ability. Question 2 concerns how the participants' English language proficiency, operationalized by the TOEFL PBT, the LVLT, and the MLAT–E, correlated with their decoding of elision and assimilation on the Elicited Imitation Test (EIT). Finally, question 3 concerns the strength of the relationship between the participants' English-learning backgrounds, and decoding of elision and assimilation on the EIT. In the sections that follow, I explain how research question 1a was explored. I begin with an overview of the analysis, and then look at items with and without elision, items with and without assimilation, items with faster and slower utterance rates, and shorter and longer items.

Research Question 1a: Empirical Item Hierarchy for Elision and Assimilation

Research question 1a asked: What is the empirical hierarchy for the phonological features of elision and assimilation, and what are the determinants of this hierarchy? This research question was answered by first inspecting the Rasch item-person maps for the EIT instrument and observing the item difficulty measures of the phonological features displayed in the figures. Next, the descriptive statistics for each of the phonological features was shown in order to clarify the item distribution. Finally, four paired-sample *t*-tests were conducted to compare four conditions ($N = 25$):

- items with elision (+E) and without elision (–E);
- items with assimilation (+A) and without assimilation (–A);
- items spoken at two utterance rates (+S [4.7 syllables per second] and –S [3.17 syllables per second]); and
- items with two syllable counts (6-syllable and 8-syllable words).

There were 24 items per sandhi-variation phenomenon, so there was limited statistical power in the analyses. The independent variable was the embedded phonological effect: elision, assimilation, utterance rate, and syllable count. The dependent variable in each of the following cases was the various Rasch item difficulty estimates for decoding on the EIT. Paired-samples *t*-tests were conducted to evaluate whether the scores differed significantly. Assumptions for the *t*-test were that difference scores are normally distributed and that the data were measured at the interval level at the ±1.96 criterion for a normal distribution. A Kolmogorov-Smirnov test of normality was used to test for normality of the different scores. The percentages on the tests were $p > .05$, indicating that the data were normally distributed. Furthermore, the data were the participants' scores, which were measured at the interval level using Rasch logits. Cohen's *d* is reported as a measure of effect size for the comparisons, where values up to .40 are small, .70 are medium, and 1.00 are large (Plonsky & Oswald, 2014). The descriptive statistics are paired with the respective *t*-test data for each phonological feature. In the sections that follow I will discuss the descriptive statistics output and Wright maps exploring this research question, including analyses of items with and without elision and assimilation, items containing faster and slower utterance rates, and items that were shorter or longer.

Items With and Without Elision

The Rasch item-person map for the items with and without elision on the EIT instrument (Figure 10) show a general tendency for items with elision (+E) to have higher difficulty estimates than items without elision (–E). However, this is not a simple comparison of +E items and –E items because all of the items also have utterance-rate and syllable-length parameters. The utterance-rate and syllable-length parameters are balanced between all of the elision items; there are 12 items for each of the two utterance rates and 12 items for each of the two syllable lengths in the EIT. I use the word *tendency* to describe the distribution of items in the item-person map for the EIT, because items with elision were more difficult for participants than items without elision. Nonetheless, this tendency was sometimes contradicted. For example, Item 39 (+E8+SEIT39) was hypothesized to be more difficult than Item 44 (-E8+S EIT44) because Item 39 is an 8-syllable item with the phonological effect of elision, while Item 44 has 8 syllables without the phonological effect of elision. Moreover, Item 44 had higher difficulty estimates than two of the three 8-syllable +E items, Items 39 and 25. Overall, four items without elision had high difficulty estimates when compared to items with elision: (a) Item 44 (-E8+SEIT44) had a higher difficulty estimate than Items 25 (+E8+SEIT25) and 39 (+E8+SEIT39), (b) Item 29 (-E8-SEIT29) had a higher difficulty estimate than Items 37 (+E8-SEIT37) and 10 (+E8-SEIT10), (c) Item 46 (-E6+SEIT46) had a higher

```
------------------------------------------------------------------------------
More able      |  More difficult items
participants   |
    3          +
               |
               |
            X  |
               |
               |
               |
               |
    2          +
               |
               |
              T|T  +E8+SEIT16
               |
            X  |   +E8-SEIT43
               |
               |
    1         S+
            X  |S             -E8+SEIT44
           XX  |                            +E6+SEIT18   -E6+SEIT46
            X  |              -E8-SEIT29
           XX  |                            +E6-SEIT6
               |                            +E6+SEIT30
               |   +E8+SEIT25  -E8-SEIT27
          XXX  |   +E8+SEIT39
           XX M|   +E8-SEIT37
            X  |
    0     XXX  +M                           +E6-SEIT8
            X  |                                         -E6+SEIT19
          XXX  |                            +E6-SEIT47   -E6+SEIT9
         XXXX S|                            +E6+SEIT5
               |                                         -E6-SEIT31
               |   +E8-SEIT10                            -E6-SEIT21
               |              -E8-SEIT1
               |S                                        -E6-SEIT15
   -1         T+
               |
               |
               |
               |              -E8+SEIT4
               |T
               |
               |              -E8+SEIT42
   -2          +
Less able      |  Easier items
participants   |
------------------------------------------------------------------------------
```

Figure 10. Item-person map for Elicited Imitation Test instrument, elision. E = Elision (+E includes an elided form, –E does not include elision), S = utterance rate (+S = faster, –S = slower), EIT = Elicited Imitation Test, and X = 1 participant. M = mean, S = 1 SD, T = 2 SD.

difficulty estimate than Items 30 (-E6+SEIT30) and 5 (-E6+SEIT5), and (d) Item 19 (-E6+SEIT19) had a higher difficulty estimate than Item 5 (-E6+SEIT5). The four columns of items in Figure 10 are arranged from left to right; +E8 (i.e., 8-syllable items with elision and ±speed), –E8 (i.e., 8-syllable items without elision and ±speed), +E6 (i.e., 6-syllable items with elision and ±speed), and –E6 (i.e., 6-syllable items without elision and ±speed).

Table 31 shows the descriptive statistics for items with and without elision, each embedded with two variations of speed (i.e., 3.17 syllables per second and 4.7 syllables per second) and two variations of word length (i.e., 6 and 8 syllables), as measured by Rasch logits. Items with elision (*M* = .45; *SD* = .65) were .80 logits more difficult than the items without elision (*M* = -.35; *SD* = 1.10). The first assumption of the paired-samples *t*-test concerns the normality of the distribution. The data were normally distributed for the +E items (skewness = .46, kurtosis = -.03) and the –E items (skewness: -1.14, kurtosis: 1.83), as the statistics were within the ±1.96 range (Field, 2013). Furthermore, *z*-scores for skewness (+E = .72; –E = 1.78) and kurtosis (+E = .84; –E = 1.46) were within the acceptable range of ±1.96 (Field, 2013). The second assumption was met because the participants' scores were measured at the interval level using Rasch logits.

Table 31. *Descriptive Statistics for +Elision and –Elision Items*

	+ Elision (+E)	– Elision (–E)
M	.45	-.35
SE	.19	.32
95% CI	[.04, .86]	[-1.04, .35]
SD	.65	1.10
Skewness	.46	-1.14
SES	.64	.64
Kurtosis	-.03	1.83
SEK	1.23	1.23

Note. All statistics were based on Rasch logits.

A paired-samples *t*-test was conducted to evaluate whether the +E and –E items differed significantly. The results showed a significant difference in the item difficulty estimates, $t(11) = 2.59$, $p = .03$, $d = .35$; thus, the items with the phonological feature of elision were significantly more difficult than the items without elision, although the effect size was small. The 95% confidence interval (CI) for the mean difference between +E and –E was .29 to 1.31, so the estimate was not precise.

Items With and Without Assimilation

The Rasch item-person map for the items with and without assimilation on the EIT instrument (Figure 11) show a general tendency for items with assimilation (+A) to have higher difficulty estimates than items without assimilation (–A). However, the comparison of +A items and –A items also includes the utterance-rate and syllable-length parameters. The utterance-rate and syllable-length parameters were balanced between all of the assimilation items; there are 12 items for each of the two utterance rates and 12 items for each of the two syllable lengths in the EIT. I again use the word *tendency* to describe the distribution of items in the item-person map for the EIT, because items with assimilation were more difficult for participants than items without assimilation. However, again the tendency was sometimes contradicted by different difficulty estimates. For example, Item 7 (+A8+SEIT7) was hypothesized to be more difficult than Item 20 (–A8+SEIT20) because Item 7 is an 8-syllable item with assimilation, while Item 20 has 8 syllables without assimilation. Moreover, Item 20 had higher difficulty estimates than two of the three 8-syllable +A items (i.e., Items 7 and 26). Overall, three

```
-----------------------------------------------------------------------------
More able      |   More difficult items
participants   |
    3          +
               |
               |
            X  |
               |
               |
               |
               |
    2          +
               |
               |
              T|T
               |                               +A6+SEIT41
            X  |
               |                                            -A6-SEIT45
               |                  -A8-SEIT48   +A6-SEIT2
    1         S+
            X  |S
           XX  |
            X  |
           XX  |                               +A6+SEIT12
               |                               +A6-SEIT38
          XXX  |   +A8+SEIT14
           XX M|   +A8-SEIT33      -A8+SEIT20
            X  |   +A8+SEIT26
    0     XXX  +M                 -A8+SEIT34
            X  |                  -A8+SEIT32   +A6-SEIT23
          XXX  |
         XXXX S|   +A8-SEIT3       -A8-SEIT11
               |   +A8-SEIT13
               |   +A8+SEIT7
               |                  -A6+SEIT22
               |S                 -A8-SEIT17
   -1         T+                                            -A6-SEIT35
               |                  +A6-SEIT24                -A6-SEIT28
               |
               |
               |                                            -A6+SEIT40
               |T
               |
               |
   -2          +
               |
               |                                            -A6+SEIT36
               |
               |
               |
               |
               |
   -3          +
Less able      |  Easier items
Participants   |
-----------------------------------------------------------------------------
```

Figure 11. Item-person map for Elicited Imitation Test instrument. A = Assimilation (+A items include assimilation, – A items do not include assimilation), 6 = 6 syllables, 8 = 8 syllables, S = utterance rate (+S = faster, –S = slower), EIT = Elicited Imitation Test, and X = 1 participant. M = mean, S = 1 SD, T = 2 SD.

items without assimilation had high difficulty estimates when compared to items with assimilation: (a) Item 48 (-A8-SEIT48) had a higher difficulty estimate than Items 33 (+A8-SEIT33), 3 (+A8-SEIT3) and 13 (+A8-SEIT13), (b) Item 20 (-A8+SEIT20) had a higher difficulty estimate than Items 26 (+A8+SEIT26) and 7 (+A8+SEIT7), and (c) Item 45 (-A6-SEIT45) had a higher difficulty estimate than Item 24 (+A6-SEIT24). The four columns of items in Figure 11 are arranged from left to right; +A8 (i.e., 8-syllable items with assimilation), –A8 (i.e., 8-syllable items without assimilation), +A6 (i.e., 6-syllable items with assimilation), and –A6 (i.e., 6-syllable items without assimilation).

Next, the items with and without assimilation were analyzed. Table 32 shows the descriptive statistics for the assimilation items as measured by Rasch logits. Items with assimilation (M = .21; SD = .77) had a higher Rasch mean difficulty estimate than the items without assimilation (M = -.36; SD = 1.05). The first assumption concerned the normality of the distribution. The skewness was .14 and kurtosis was -.39 for the +A items, and the skewness was .09 and kurtosis was -.43 for the –A items; thus, the data met the criteria of < ±1.96 (Field, 2013). Furthermore, the z-scores for skewness (+A = .22; –A = .14) and kurtosis (+A = .32; –A = .35) were within the acceptable range for the assumption of normal distribution, ±1.96 (Field, 2013). The second assumption was met because the participants' were measured at the interval level using Rasch logits.

Table 32. *Descriptive Statistics for + Assimilation and – Assimilation Items*

	+ Assimilation (+A)	– Assimilation (–A)
M	.21	-.36
SE	.22	.30
95% CI	[-.28, .70]	[-1.03, .30]
SD	.77	1.05
Skewness	.14	.09
SES	.64	.64
Kurtosis	-.39	-.43
SEK	1.23	1.23

Note. All statistics were based on Rasch logits.

A paired-samples *t*-test was conducted to evaluate whether there was a significant difference between the +A items and –A items. The results indicated a non-significant difference in the scores, $t(11) = 1.33$, $p = .21$, $d = -.33$, though the difference was in the hypothesized direction and the effect size was approximately the same as that in the previous analysis. The 95% CI for the mean difference between +A and –A was .05 to 1.09 logits; thus, the estimate was not precise.

Items with Faster and Slower Utterance Rates

The third analysis concerned possible differences in items with faster and slower utterance rates. The comparison of +S items and –S items also included the elision, assimilation, and syllable-length parameters; however, the two parameters were balanced between all of the items, as there were 24 items of each phonological effect and 24 items of each syllable-length on the EIT. Figure 12 shows the Rasch item-person map for the EIT instrument, elision items, with separate columns for faster items (+S) and slower items (–S). The column on the left in Figure 12 contains items with a faster utterance rate (i.e., 4.7 syllables per second), and the column on the right contains items with a slower

```
-----------------------------------------------------------------------------
More able      |  More difficult items
participants   |
    3          +
               |
               |
            X  |
               |
               |
               |
               |
    2          +
               |
               |
              T|T  +E8+SEIT16
               |
            X  |                 +E8-SEIT43
               |
               |
    1         S+
            X  |S -E8+SEIT44
           XX  |   +E6+SEIT18
            X  |   -E6+SEIT46    -E8-SEIT29
           XX  |                 +E6-SEIT6
               |   +E6+SEIT30
               |   +E8+SEIT25    -E8-SEIT27
          XXX  |   +E8+SEIT39
           XX M|
            X  |                 +E6-SEIT8
    0     XXX  +M
            X  |   -E6+SEIT19    +E8-SEIT37
          XXX  |   -E6+SEIT9     +E6-SEIT47
         XXXX S|   +E6+SEIT5
               |                 -E6-SEIT31
               |                 -E6-SEIT21
               |                 +E8-SEIT10
               |                 -E8-SEIT1
               |S                -E6-SEIT15
   -1         T+
               |
               |
               |
               |  -E8+SEIT4
               |T
               |
               |  -E8+SEIT42
   -2          +
Less able      |  Easier items
Participants   |
-----------------------------------------------------------------------------
```

Figure 12. Item-person map for Elicited Imitation Test instrument, elision, with separate columns for utterance rates. E = Elision (+E includes elided form, –E does not include elision), 6 = 6 syllables, 8 = 8 syllables, S = utterance rate (+S = faster, –S = slower), EIT = Elicited Imitation Test, and X = 1 participant. M = mean, S = 1 SD, T = 2 SD.

utterance rate (i.e., 3.17 syllables per second). The figure shows a slight tendency for items with the faster utterance rate to have higher difficulty estimates than items with the slower utterance rate. For example, Item 16 (+E8+SEIT16) was hypothesized to be more difficult than Item 43 (+E8–SEIT43) because of the faster (+S) utterance rate, and the hypothesized outcome was confirmed. However, the tendency was sometimes contradicted by different difficulty estimates. For example, Item 25 (+E8+SEIT25) was hypothesized to be more difficult than Item 43 (+E8–SEIT43) because both items were made up of 8 syllables and elision was present. Item 25 had a faster utterance rate (+S), while Item 43 had a slower utterance rate (–S). Moreover, Item 43 had higher difficulty estimates than two of the three 8-syllable +S items (i.e., Items 25 and 39). Overall, three items, Items 43, 6, and 29, within balanced parameters (e.g., +E8+S versus +E8-S) had higher difficulty estimates than hypothesized.

Figure 13 shows the Rasch item-person map for the EIT instrument, assimilation items, with separate columns for items with faster items (+S) and slower items (–S). The column on the left in Figure 13 contains items with a faster utterance rate (i.e., 4.7 syllables per second), and the column on the right contains items with a slower utterance rate (i.e., 3.17 syllables per second). The figure shows a slight tendency for items with the faster utterance rate to have higher difficulty estimates than items with the slower utterance rate. For example, Item 41 (+A6+SEIT41) was hypothesized to be more difficult than Item 38 (+A6-SEIT2) because of faster (Item 41) utterance rate, and the hypothesized outcome was confirmed. However, the tendency was sometimes contradicted by different difficulty estimates. For example, Item 7 (+A8+SEIT7) was hypothesized to be more difficult than Item 33 (+A8-SEIT33) because both items were

```
-----------------------------------------------------------------------------
More able      |  More difficult items
participants   |
    3          +
               |
               |
            X  |
               |
               |
               |
               |
    2          +
               |
               |
              T|T
               |  +A6+SEIT41
            X  |
               |                -A6-SEIT45
               |                -A8-SEIT48
    1         S+  +A6+SEIT2
            X  |S
           XX  |  +A6+SEIT12
            X  |
           XX  |
               |  +A8+SEIT14    +A6-SEIT38
          XXX  |
           XX M|  -A8+SEIT20
            X  |  +A8+SEIT26    +A8-SEIT33
    0     XXX  +M -A8+SEIT34
            X  |  -A8+SEIT32
          XXX  |                +A6-SEIT2
         XXXX S|                -A8-SEIT11
               |                +A8-SEIT3
               |  +A8+SEIT7     +A8-SEIT13
               |  -A6+SEIT22
               |S               -A8-SEIT17
   -1         T+                -A6-SEIT35
               |                +A6-SEIT24
               |                -A6-SEIT28
               |
               |  -A6+SEIT40
               |T
               |
               |
   -2          +
               |
               |  -A6+SEIT36
               |
Less able      |  Easier items
participants   |
-----------------------------------------------------------------------------
```

Figure 13. Item-person map for Elicited Imitation Test instrument, assimilation, with separate columns for utterance rates. A = Assimilation (+A items include assimilation, –A items do not include assimilation), 6 = 6 syllables, 8 = 8 syllables, S = utterance rate (+S = faster, –S = slower), EIT = Elicited Imitation Test, and X = 1 participant. M = mean, S = 1 SD, T = 2 SD.

+A and were made up of 8 syllables, but Item 7 had a faster utterance rate, while Item 33 had a slower utterance rate. Overall, three items, Items 45, 48, 33, within balanced parameters (e.g., +A8+S versus +A8-S) had higher difficulty estimates than hypothesized.

Table 33 shows the descriptive statistics for the items with different utterance rates. Assimilation was balanced, with 12 items for the two speeds. The mean of the items with the faster utterance rate (*M* = .04; *SD* = 1.12) was .26 logits higher than that for the items with the slower utterance rate (*M* = -.22; *SD* = .76). The skewness was -1.20 and kurtosis was 1.15 for the +S items, and the skewness was .56 and kurtosis was -.75 for the –S items. All statistics met the ±1.96 criterion for a normal distribution (Field, 2013); however, the *z*-scores for skewness (+S = -2.34; –S = 1.19) and kurtosis (+S = 1.25; –S = .35), revealed a non-normal z-score for skewness (i.e., > -1.96) for the +S items, which implies reduced reliability of the sample (Tabachnick & Fidell, 2013). The non-normal skewness for this sample likely occurred because of the small number of items in the analysis. The second assumption that data were measured at the interval level was met because the item scores were measured at the interval level using Rasch logits.

Table 33. *Descriptive Statistics for +Speed and –Speed Items*

	+Speed (+S)	– Speed (–S)
M	.03	-.06
SE	.23	.15
95% CI	[-.45, .51]	[-.38, .26]
SD	1.12	.76
Skewness	-1.10	.56
SES	.47	.47
Kurtosis	1.15	-.75
SEK	.92	.92

Note. All statistics were based on Rasch logits.

A paired-sample *t*-test was conducted to evaluate whether there was a significant difference between the +S and –S items. The results showed that there was a non-significant difference in scores, $t(23) = -.33$, $p = .74$, $d = .15$, though the difference was in the hypothesized direction. The 95% CI for the mean difference between +S and –S was -.45 to .63 logits; thus, the estimate was not precise.

Shorter and Longer Items

The fourth analysis concerned shorter and longer items. Figure 14 shows the item-person map for the EIT, elision items, with separate columns for shorter items with 6 syllables and longer items with 8 syllables. The column on the left in Figure 14 contains items with 6 syllables, and the column on the right contains items with 8 syllables. However, the comparison of 8-syllable items and 6-syllable items is balanced as they also included the elision and assimilation, and utterance-rate parameters. There was a general tendency for longer items to have higher difficulty estimates than shorter items. However, this tendency was sometimes contradicted by different difficulty estimates. For example, Item 10 (+E8-SEIT10) was hypothesized to be more difficult than Item 47 (+E6-SEIT47); although both items are +E and –S, Item 10 is an 8-syllable item, while Item 47 has 6 syllables. Overall, six items within balanced parameters (e.g., +A8+S versus +A6+S) had higher difficulty estimates than hypothesized.

```
-----------------------------------------------------------------------------
More able      |  More difficult items
participants   |
    3          +
               |
               |
            X  |
               |
               |
               |
               |
    2          +
               |
               |
              T|T                +E8+SEIT16
               |
            X  |                 +E8-SEIT43
               |
               |
    1         S+
            X  |S                -E8+SEIT44
           XX  |  +E6+SEIT18
            X  |  -E6+SEIT46     -E8-SEIT29
           XX  |  +E6-SEIT6
               |  +E6+SEIT30     +E8+SEIT25
               |                 -E8-SEIT27
          XXX  |                 +E8+SEIT39
           XX M|                 +E8-SEIT37
            X  |
    0     XXX  +M +E6-SEIT8
            X  |  -E6+SEIT19
          XXX  |  -E6+SEIT9
         XXXX S|  +E6-SEIT47
               |  +E6+SEIT5
               |  -E6-SEIT31
               |  -E6-SEIT21     +E8-SEIT10
               |                 -E8-SEIT1
               |S -E6-SEIT15
   -1         T+
               |
               |
               |
               |                -E8+SEIT4
               |T
               |
               |                -E8+SEIT42
   -2          +
Less able      |  Easier items
participants   |
-----------------------------------------------------------------------------
```

Figure 14. Item-person map for Elicited Imitation Test instrument, elision, with separate columns for 6 syllables and 8 syllables. E = Elision (+E includes elided forms, –E does not include elided forms), 6 = 6 syllables, 8 = 8 syllables, S = utterance rate (+S = faster, –S = slower), EIT = Elicited Imitation Test, and X = 1 participant. M = mean, S = 1 SD, T = 2 SD.

Figure 15 shows the item-person map for the EIT, assimilation items, with separate columns for items with 6 syllables and 8 syllables. The column on the left in Figure 15 contains items with 6 syllables, and the column on the right contains items with

```
-----------------------------------------------------------------------------
More able      |  More difficult items
participants   |
            X  |
               |
               |
               |
               |
    2          +
               |
               |
              T|T
               |  +A6+SEIT41
            X  |
               |  -A6-SEIT45
               |                   -A8-SEIT48
    1         S+  +A6-SEIT2
            X  |S
           XX  |
            X  |
           XX  |  +A6+SEIT12
               |  +A6-SEIT38     +A8+SEIT14
          XXX  |                 +A8-SEIT33
           XX M|                 -A8+SEIT20
            X  |                 +A8+SEIT26
    0     XXX  +M                -A8+SEIT34
            X  |  +A6-SEIT23     -A8+SEIT32
          XXX  |
         XXXX S|                 +A8-SEIT3
               |                 -A8-SEIT11
               |                 +A8-SEIT13
               |  -A6+SEIT22     +A8+SEIT7
               |S                -A8-SEIT17
   -1         T+  -A6-SEIT35
               |  +A6-SEIT24
               |  -A6-SEIT28
               |
               |  -A6+SEIT40
               |T
               |
               |
   -2          +
               |
               |  -A6+SEIT36
Less able      |  Easier items
participants   |
-----------------------------------------------------------------------------
```

Figure 15. Item-person map for Elicited Imitation Test instrument, assimilation, with separate columns for 6 syllables and 8 syllables. A = Assimilation (+A items include assimilation, – A items do not include assimilation), 6 = 6 syllables, 8 = 8 syllables, S = utterance rate (+S = faster, –S = slower), EIT = Elicited Imitation Test, and X = 1 participant. M = mean, S = 1 SD, T = 2 SD.

8 syllables. The difficulty estimates are similar for both the 6 and 8 syllable items. There are items that are hypothesized to be more difficult, for instance, Item 14 [+A8+SEIT14], that have lower difficulty estimates than items that are hypothesized to be easier (e.g., Item 41 [+A6+SEIT41]), but the pattern is that +A is more difficult than -A.

Table 34 shows the descriptive statistics for both sets of items. Items with 8 syllables (*M* = .07; *SD* = .98) were .16 logits more difficult than items with 6 syllables (*M* = -.09; *SD* = .93). The skewness was -1.15 and kurtosis was 2.77 for the 8-syllable items, and the skewness was -.20 and kurtosis was -.55 for the 6-syllable items. Moreover, the *z*-scores for skewness (8-syllable items = 2.34; 6-syllable items = -.43) and kurtosis (8-syllable items = 3.04; 6-syllable items = -.27) revealed both non-normal skewness (2.34) and kurtosis (3.04) for the 8-syllable items. A Kolmogorov-Smirnov test of normality was used again to check for normality of the 6-syllable and 8-syllable items. The percentage on the test 6-syllable test, $D(24) = .13$, $p = .20$, and the 8-syllable test, $D(24) = .15$, $p = .20$, were normal, indicating that the data were normally distributed. The second assumption that data were measured at the interval level was met because the participants' scores were measured at the interval level using Rasch logits.

Table 34. *Descriptive Statistics for Syllable Count: 6 Syllables and 8 Syllables*

	6-syllable items	8-syllable items
M	-.09	.07
SE	.19	.20
95% CI	[-.49, .30]	[-.35, .48]
SD	.93	.98
Skewness	-.20	-1.15
SES	.47	.47
Kurtosis	-.25	2.77
SEK	.92	.92

Note. All statistics were based on Rasch logits.

A paired-sample *t*-test was conducted to evaluate whether there was a significant difference between the 6-syllable items and 8-syllable items. The results showed that there was a non-significant difference in scores, $t(23) = -.58$, $p = .57$, $d = .04$, though the difference was in the hypothesized direction, meaning towards 8-syllable items. The participants' performances related to the number of syllables in the targets were similar to chance, and there was considerable overlap (see Figure 16) in the distribution of the scores. The 95% CI for the mean difference between 8-syllable and 6-syllable items was -.38 to .70 logits, indicating that the estimate was not precise.

Figure 16 summarizes the findings for the four variables analyzed above. When elision or assimilation was present, the participants' decoding was adversely affected and the items tended to become more difficult. Furthermore, a faster utterance rate adversely affected the participants' decoding at times. Finally, the 8-syllable items were slightly more difficult to decode than the 6-syllable items; however, the difference was not statistically significant, $p = .20$. Overall, there was considerable overlap within the variables, especially with utterance rates and syllable counts.

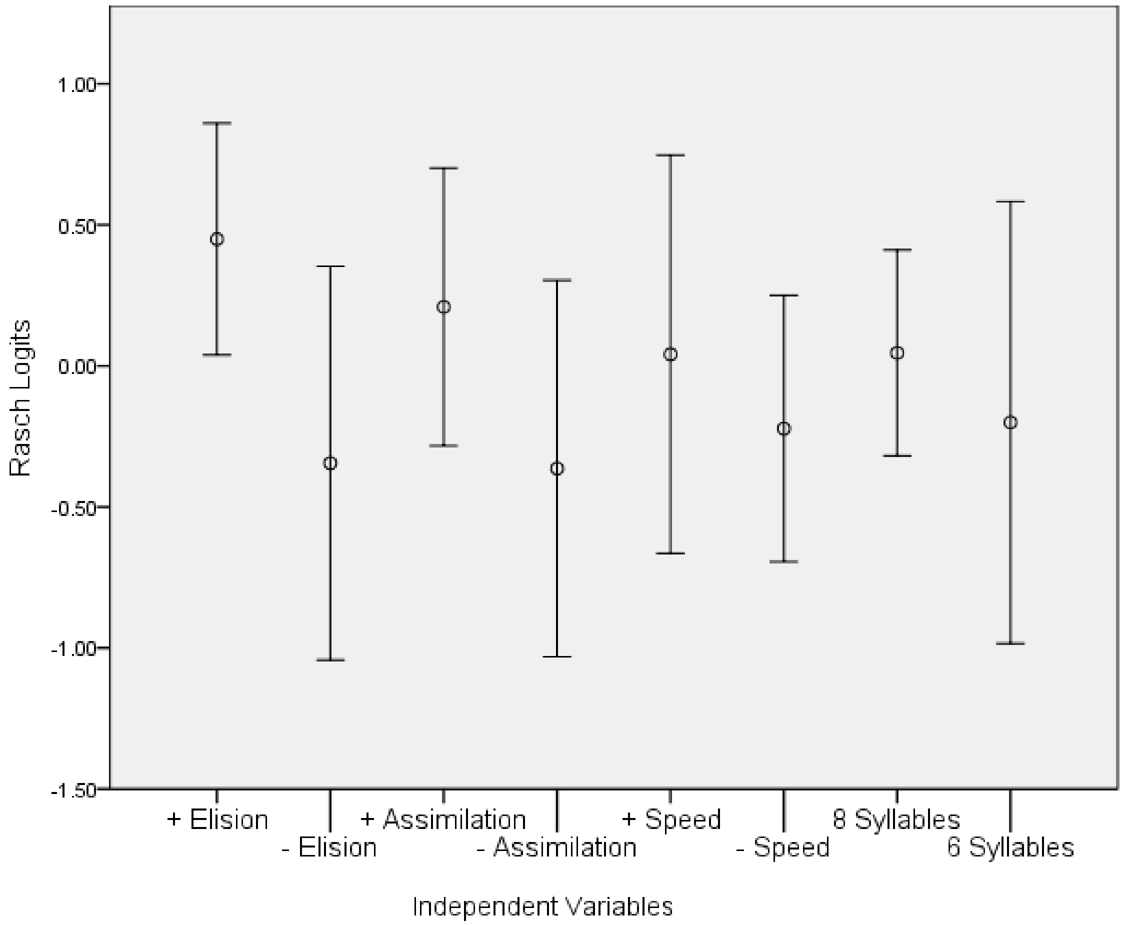

Figure 16. Line graph of 95% CI error bars for Elicited Imitation Test results.

Research Question 1B: Input Rates Combined with Elision and Assimilation

Research question 1b asked how the two input rates combined with the phonological features of elision and assimilation affected the participants' decoding of the items on the EIT. This research question was answered by first making assumptions about the results of the descriptive statistics for the EIT, elision and assimilation, and the two utterance rates of 4.7 syllables per second and 3.17 syllables per second. The effects on the participants' ability to decode the speech stream of the EIT can be observed by reviewing the descriptive statistics for each of the results.

Twelve paired-sample *t*-tests were used to analyze combinations of elision, assimilation, and utterance rates from the items used in the EIT. The 48 EIT items were made up of the following sandhi-variation types: 12 items with elision (+E items), 12 items without elision (–E items), 12 items with assimilation (+A items) and 12 items without assimilation (–A items). Furthermore, each set of +E and –E items and + A and –A items were also embedded with two different utterance rates, +S (4.7 syllables per second) and –S (3.17 syllables per second). Each of the combinations, +E with +S, +E with –S, –E with +S, and –E with –S, were made up of 6 items, for a total of 24 items, or half of the total number of items on the EIT. The other 24 items were the combinations of assimilation, +A with +S, +A with –S, –A with +S, and –A with –S, also with 6 items for each combination and a total of 24 items. Six *t*-tests were run to cover each combination.

Table 35 shows the descriptive statistics for combinations of items with and without elision (i.e., +E and –E), and items with faster and slower utterance rates (i.e., +S and –S). Items that had +E with +S (*M* = .62; *SD* = .60) had a higher mean than the items that had +E with –S (*M* = .28; *SD* = .70), and items that had –E with +S (*M* = -.50; *SD* = 1.46), and items that had –E with –S (*M* = -.19; *SD* = .68). The skewness ranged from -1.04 (–E with +S) to .89 (+E with –S), and kurtosis ranged from -1.51 (–E with –S) to 1.99 (+E with +S); thus, the data met the ±1.96 (Field, 2013). The *z*-scores for skewness ranged from -1.22 (–E with +S) to 1.04 (+E with –S) logits. The *z*-scores for kurtosis ranged from -.87 (–E with –S) to 1.14 (+E with +S). The *z*-scores for skewness and kurtosis were within the acceptable range for the assumption of normal distribution, ±1.96 (Field, 2013). The second assumption that data were measured at the interval level was met because the scores were measured at the interval level using Rasch logits.

Table 35. *Descriptive Statistics for Combinations of Elision and Speed*

	+ Elision and + Speed	+ Elision and – Speed	– Elision and + Speed	– Elision and – Speed
M	.63	.28	-.50	-.19
SE	.25	.29	.60	.28
95% CI	[-.01, 1.26]	[-.46, 1.01]	[-2.03, 1.04]	[-.91, .52]
SD	.60	.70	1.46	.68
Skewness	.65	.89	-1.04	.87
SES	.85	.85	.85	.85
Kurtosis	1.99	.98	.46	-1.51
SEK	1.74	1.74	1.74	1.74

Note. All statistics were based on Rasch logits.

Six paired sample *t*-tests were conducted between the following combinations items to explore differences among participants' decoding: (a) +E and +S versus +E and –S, (b) +E and +S versus –E and +S, (c) +E and +S versus –E and –S, (d) +E and –S versus –E and +S, (e) +E and –S versus –E and –S, and (f) –E and +S versus –E and –S. Using the Holm's sequential Bonferroni procedure and controlling for familywise error rate across the six tests at the .05 level, none of the six pairwise comparisons were found to be significant.

The first combination compared elision at faster and slower speeds. The results indicated that the mean for the +E with +S items (M = .63, SD = .60) was not significantly different than the mean of the +E and –S items (M = .28, SD = .70), $t(5)$ = 1.35, p = .24, d = .54. The 95% CI for the mean difference was -.32 to 1.02 logits.

The second combination compared elision and absence of elision at a faster utterance rate. The results of +E with +S items (M = .63, SD = .60) were not significantly different than the mean of –E with +S items (M = -.50, SD = 1.46), $t(5)$ = 1.98, p = .10, d = 1.01. The 95% CI for the mean difference was -.33 to 2.58 logits.

The third combination compared elision and a faster utterance rate with the absence of elision and slower utterance rate. The mean of +E with +S items (*M* = .63, *SD* = .60) was significantly different than the mean of –E with –S items (*M* = -.19, *SD* = .68), *t*(5) = 4.43, *p* = .01, *d* = 1.27. The 95% CI for the mean difference was .34 to 1.29 logits.

The fourth combination compared elision at slower speed with the absence of elision at a faster speed. For combination four, the results of +E with –S items (*M* = .28, *SD* = .70) was not significantly different than the mean of –E with +S items (*M* = -.50, *SD* = 1.46), *t*(5) = 1.36, *p* = .23, *d* = .68. The 95% CI for the mean difference was -.69 to 2.23 logits.

The fifth combination compared elision at a slower speed with the absence of elision at a slower speed. The results of +E with –S items (*M* = .28, *SD* = .70) was not significantly different than the mean of –E with –S items (*M* = -.19, *SD* = .68), *t*(5) = 2.06, *p* = .09, *d* = .68. The 95% CI for the mean difference was -.12 to 1.05 logits.

Finally, the sixth combination compared the absence of elision with faster and slower utterance rates. For combination six, the results of –E with +S items (*M* = -.50, *SD* = 1.46) was not significantly different than the mean of –E with –S items (*M* = -.19, *SD* = .68), *t*(5) = -.44, *p* = .68, *d* = .27. The 95% CI for the mean difference was -2.08 to 1.47 logits. Overall, the results showed that the differences became smaller as elision was removed and utterance rate became slower.

Table 36 shows the summary of *p*-values for the combinations of elision and speed. The Holm’s Bonferroni significant *p*-value column contains the value required by the adjustment to claim a significant *p*-value. There were no significant *p*-values for any of the combinations.

Table 36. *Summary of t-Test p-values for Combinations of Elision and Speed*

Combination	*p*-value	Holm's Bonferroni significant *p*-value
(a) +E and +S versus –E and –S	.01	.008
(b) +E and –S versus –E and –S	.09	.010
(c) +E and +S versus –E and +S	.10	.012
(d) +E and –S versus –E and +S	.23	.017
(a) +E and +S versus +E and –S	.24	.025
(f) –E and +S versus –E and –S	.68	.050

$p < .05$ for target α

Looking more closely at the comparison of items on the EIT with and without elision, while holding the added confounds of syllable length and utterance rate constant, shows the effect of the sound changes. Figure 17 shows the Rasch item-person map for faster, longer items with and without elision. Although there were only six items to compare, the items with elision are more difficult for the participants than the items without elision, with one exception. Item -–E8+SEIT44 (*The king got up to make a speech*) contains a final plosive *g*, which is voiced in this instance, followed by an initial plosive *g*. Nothing was elided. Possibly the context of the item caused confusion, as the idea of a *king making speeches* is not historically relevant to the L1 culture of the participants.

```
----------------------------------------------------------------------------
More able      |  More difficult items
participants   |
            X  |
               |
               |
               |
               |
    2          +
               |
               |
              T|T +E8+SEIT16
               |
            X  |
               |
               |
    1         S+
            X  |S               -E8+SEIT44
           XX  |
            X  |
           XX  |
               |
               |   +E8+SEIT25
          XXX  |   +E8+SEIT39
           XX M|
            X  |
    0     XXX  +M
            X  |
          XXX  |
         XXXX S|
               |
               |
               |
               |S
   -1         T+
               |
               |
               |
               |                -E8+SEIT4
               |T
               |
               |                -E8+SEIT42
   -2          +
Less able      |  Easier items
participants   |
----------------------------------------------------------------------------
```

Figure 17. Item-person map for Elicited Imitation Test instrument, ±elision with 8 syllables and faster utterance rate. E = Elision (+E includes an elided form, –E does not include elision), S = utterance rate (+S = faster, –S = slower), EIT = Elicited Imitation Test, and X = 1 participant. M = mean, S = 1 SD, T = 2 SD.

Figure 18 shows the Rasch item-person map for faster, shorter items with and without elision. Items with and without elision are equally difficult, perhaps because the length of the items (6 syllables) rendered all of the items relatively similar.

```
-----------------------------------------------------------------------------
More able      |  More difficult items
participants   |
    3          +
               |
               |
            X  |
               |
               |
               |
               |
    2          +
               |
               |
              T|T
               |
            X  |
               |
               |
    1         S+
            X  |S
           XX  |  +E6+SEIT18      -E6+SEIT46
            X  |
           XX  |
               |  +E6+SEIT30
               |
          XXX  |
           XX M|
            X  |
    0     XXX  +M
            X  |                  -E6+SEIT19
          XXX  |                  -E6+SEIT9
         XXXX S|  +E6+SEIT5
               |
               |
               |
               |S
   -1         T+
               |
Less able      |  Easier items
Participants   |
-----------------------------------------------------------------------------
```

Figure 18. Item-person map for Elicited Imitation Test instrument, ±elision with 6 syllables and faster utterance rate. E = Elision (+E includes an elided form, –E does not include elision), S = utterance rate (+S = faster, –S = slower), EIT = Elicited Imitation Test, and X = 1 participant. M = mean, S = 1 SD, T = 2 SD.

Figure 19 shows the Rasch item-person map for slower, longer items with and without elision. Items with and without elision have similar difficulty estimates, perhaps because the slower utterance rate of the items made them relatively similar. The least difficult item with elision (+E) is +E8-SEIT10, which has a slower utterance rate (-S) and is elided as, *Somehow we woke up very late*. In this case, the elision (i.e., [sʌm haʊ i]) only slightly alters the utterance by removing the voiceless fricative *w*.

```
----------------------------------------------------------------------------
More able      |  More difficult items
participants   |
    3          +
               |
               |
            X  |
               |
               |
               |
               |
    2          +
               |
               |
              T|T
               |
            X  |  +E8-SEIT43
               |
               |
    1         S+
            X  |S
           XX  |
            X  |                -E8-SEIT29
           XX  |
               |
               |                -E8-SEIT27
          XXX  |
           XX M|  +E8-SEIT37
            X  |
    0     XXX  +M
            X  |
          XXX  |
         XXXX S|
               |
               |  +E8-SEIT10
               |                -E8-SEIT1
               |S
   -1         T+
               |
Less able      |  Easier items
participants   |
----------------------------------------------------------------------------
```

Figure 19. Item-person map for Elicited Imitation Test instrument, ±elision with 8 syllables and slower utterance rate. E = Elision (+E includes an elided form, –E does not include elision), S = utterance rate (+S = faster, –S = slower), EIT = Elicited Imitation Test, and X = 1 participant. M = mean, S = 1 SD, T = 2 SD.

However, Figure 20 shows that the Rasch item-person map for slower, shorter items with elision were more difficult than slower, shorter items without elision. For example, Item 6 (+E6-SEIT6) elides *send them* as [sɛndɛm], which could be confused as a new, two-syllable word.

```
-----------------------------------------------------------------------------
More able      |  More difficult items
participants   |
    3          +
               |
               |
            X  |
               |
               |
               |
               |
    2          +
               |
               |
              T|T
               |
            X  |
               |
               |
    1         S+
            X  |S
           XX  |
            X  |
           XX  |  +E6-SEIT6
               |
               |
          XXX  |
           XX M|
            X  |
    0     XXX  +M +E6-SEIT8
            X  |
          XXX  | +E6-SEIT47
         XXXX S|
               |                 -E6-SEIT31
               |                 -E6-SEIT21
               |
               |S                -E6-SEIT15
   -1         T+
               |
Less able      |  Easier items
Participants   |
-----------------------------------------------------------------------------
```

Figure 20. Item-person map for Elicited Imitation Test instrument, ±elision with 6 syllables and slower utterance rate. E = Elision (+E includes an elided form, –E does not include elision), S = utterance rate (+S = faster, –S = slower), EIT = Elicited Imitation Test, and X = 1 participant. M = mean, S = 1 SD, T = 2 SD.

Furthermore, the results suggest that input rates (i.e., utterance rates) might have affected decoding when combined with assimilation. Again, in general, the 95% CI revealed that +A, for example, was more difficult than –A. Moreover, +S was more difficult than –S. Therefore, the combination of +A with +S should be more difficult, in general, for the participants of this study to decode than –A with +S or –A with –S.

Table 37 shows the descriptive statistics for combinations of items with and without assimilation (i.e., +A and –A), and items with faster and slower utterance rates (i.e., +S and –S). The combinations were +A with +S, +A with –S, –A with +S, and –A with –S. Items that had +A with +S (*M* = .63; *SD* = .73) had a higher mean than the items that had +A with –S (*M* = -.21; *SD* = .60), and items that had –A with +S (*M* = -.64; *SD* = 1.05), and items that had –A with –S (*M* = -.19; *SD* =.89). The skewness ranged from -.64 (–A with +S) to .88 (–A with –S), and kurtosis ranged from -1.30 (–A with +S) to .25 (+A with +S); thus, the data met the ±1.96 (Field, 2013). Furthermore, the *z*-scores for skewness ranged from - .75 (–A with +S) to 1.04 (–A with –S). The *z*-scores for kurtosis ranged from -.75 (–A with +S) to .14 (+A with +S). The *z*-scores for skewness and kurtosis were within the acceptable ±1.96 range for a normal distribution, (Field, 2013). The second assumption that data were measured at the interval level was met because the participants' scores were measured at the interval level using Rasch logits.

Table 37. *Descriptive Statistics for Combinations of Assimilation and Speed*

	+ Assimilation with + Speed	+ Assimilation with – Speed	– Assimilation with + Speed	– Assimilation with – Speed
M	.63	-.21	-.64	-.19
SE	.30	.30	.43	.36
95% CI	[-.14, 1.40]	[-.84, .42]	[-1.73, .46]	[-1.12, .75]
SD	.73	.60	1.05	.89
Skewness	-.30	-.13	-.64	.88
SES	.85	.85	.85	.85
Kurtosis	.25	-.62	-1.30	.01
SEK	1.74	1.74	1.74	1.74

Note. All statistics were based on Rasch logits.

Six paired sample *t*-tests were conducted between the following combinations items: (a) +A and +S versus +A and –S, (b) +A and +S versus –A and +S, (c) +A and +S versus –A and –S, (d) +A and –S versus –A and +S, (e) +A and –S versus –A and –S, and (f) –A and +S versus –A and –S. None of the six pairwise comparisons were significant, controlling for familywise error rate across the six tests at the .05 level, using the Holm's sequential Bonferroni procedure.

For the first combination, the Rasch difficulty estimates of the +A and +S items were compared with those of the +A and –S items. The results indicated that the mean for the +A and +S items (M = .63, SD = .73) was not significantly higher than the mean of the +A and –S items (M = -.21, SD = .60), $t(5) = 2.79$, $p = .04$, $d = 1.26$. The 95% CI for the mean difference was .06 to 1.61 logits.

For the second combination, the Rasch difficulty estimates of the +A and +S items were compared with those of the –A and +S items. The +A and +S items (M = .63, SD = .73) were not significantly different from the mean of –A and +S items (M = -.64, SD = 1.05), $t(5) = 1.82$, $p = .13$, $d = 1.40$; however, the effect size was large. The 95% CI for the mean difference was -.52 to 3.04 logits.

For the third combination, the Rasch difficulty estimates of the +A and +S items were compared with those of the –A and –S items. The mean of the +A with +S items (M = .63, SD = .73) was not significantly higher than the mean of the –A and –S items (M = -.19, SD = .89), $t(5) = 2.27$, $p = .07$, $d = 1.01$; however, once again, the effect size was large. The 95% CI for the mean difference was -.11 to 1.73 logits.

For the fourth combination, the Rasch difficulty estimates of the +A and –S items were compared with those of the –A and +S items. The results indicated that the mean of the +A and –S items (*M* = -.21, *SD* = .60) was not significantly higher than that of the –A and +S items (*M* = -.64, *SD* = 1.05), *t*(5) = .81, *p* = .45, *d* = .50. The 95% CI for the mean difference was -.92 to 1.78 logits.

For the fifth combination, the Rasch difficulty estimates of the +A and –S items were compared with those of the –A and –S items. The results indicated that the mean of the +A and –S items (*M* = -.21, *SD* = .60) was not significantly different than that of the –A and –S items (*M* = -.19, *SD* = .89), *t*(5) = -.09, *p* = .93, *d* = .03. The 95% CI for the mean difference was -.64 to .60 logits.

For the sixth combination, the Rasch difficulty estimates of the –A and +S items were compared with those of the –A and –S items. The results indicated that the mean of the –A and +S items (*M* = -.64, *SD* = 1.05) was not significantly different than that of the –A and –S items (*M* = -.19, *SD* = .89), *t*(5) = -.70, *p* = .52, *d* = .46, though the effect size was small. The 95% CI for the mean difference was -2.11 to 1.21 logits.

Overall, the results showed that the differences became weaker as assimilation was removed and the utterance rate became slower. Table 38 shows the summary of the *p*-values for the combinations of assimilation and speed. The Holm’s Bonferroni significant *p*-value column contains the values required by the adjustment to claim a significant *p*-value. None of the combinations differed significantly.

Table 38. *Summary of t-test p-values for Combinations of Assimilation and Speed*

Combination	*p*-value	Holm's Bonferroni significant *p*-value
(a) +A and +S versus +A and –S	.04	.008
(b) +A and +S versus –A and +S	.13	.010
(c) +A and +S versus –A and –S	.07	.017
(d) +A and –S versus –A and +S	.45	.020
(e) +A and –S versus –A and –S	.93	.025
(f) –A and +S versus –A and –S	.52	.050

$p < .05$ for target α

Second, the participants' comprehension of targets with and without assimilation was not significantly different, and therefore not clearly hierarchical. Figure 11 shows, for example, that some items with assimilation are more difficult than items without assimilation, while others are not.

Figure 21 shows the Rasch item-person map for faster, longer items with and without assimilation while holding syllable length and utterance rate constant; this figure shows the effect of the sound changes on the item difficulty. The figure shows that items with and without assimilation had similar difficulty estimates.

Figure 22 shows the Rasch item-person map for faster and shorter items with and without assimilation. Faster and shorter items with assimilation were more difficult than faster, shorter items without assimilation.

```
----------------------------------------------------------------------------
More able      |  More difficult items
participants   |
    3          +
               |
               |
            X  |
               |
               |
               |
               |
    2          +
               |
               |
              T|T
               |
            X  |
               |
               |
    1         S+
            X  |S
           XX  |
            X  |
           XX  |
               |
          XXX  |  +A8+SEIT14
           XX M|                 -A8+SEIT20
            X  |  +A8+SEIT26
    0     XXX  +M                -A8+SEIT34
            X  |                 -A8+SEIT32
          XXX  |
         XXXX S|
               |
               |  +A8+SEIT7
               |
               |S
   -1         T+
Less able      |  Easier items
participants   |
----------------------------------------------------------------------------
```

Figure 21. Item-person map for Elicited Imitation Test instrument, ±assimilation with 8 syllables and faster utterance rate. A = Assimilation (+A items include assimilation, – A items do not include assimilation), 6 = 6 syllables, 8 = 8 syllables, S = utterance rate (+S = faster, –S = slower), EIT = Elicited Imitation Test, and X = 1 participant. M = mean, S = 1 SD, T = 2 SD.

```
-----------------------------------------------------------------------------
More able      |  More difficult items
participants   |
               |
            X  |
               |
               |
               |
               |
    2          +
               |
               |
              T|T
               |  +A6+SEIT41
            X  |
               |
               |  +A6+SEIT2
    1         S+
            X  |S
           XX  |
            X  |
           XX  |  +A6+SEIT12
               |
          XXX  |
           XX M|
            X  |
    0     XXX  +M
            X  |
          XXX  |
         XXXX S|
               |
               |
               |                 -A6+SEIT22
               |S
   -1         T+
               |
               |
               |
               |                 -A6+SEIT40
               |T
               |
               |
   -2          +
               |
               |                  -A6+SEIT36
               |
               |
               |
               |
               |
   -3          +
Less able      |  Easier items
participants   |
-----------------------------------------------------------------------------
```

Figure 22. Item-person map for Elicited Imitation Test instrument, ±assimilation with 6 syllables and faster utterance rate. A = Assimilation (+A items include assimilation, – A items do not include assimilation), 6 = 6 syllables, 8 = 8 syllables, S = utterance rate (+S = faster, –S = slower), EIT = Elicited Imitation Test, and X = 1 participant. M = mean, S = 1 SD, T = 2 SD.

Figure 23 shows the Rasch item-person map for slower, longer items with and without assimilation. Items with assimilation tended to be more difficult than items without assimilation.

```
-----------------------------------------------------------------------------
More able      |  More difficult items
participants   |
            X  |
               |
               |
               |
               |
    2          +
               |
               |
              T|T
               |
            X  |
               |
               |                  -A8-SEIT48
    1         S+
            X  |S
           XX  |
            X  |
           XX  |
               |
          XXX  |
           XX M|  +A8-SEIT33
            X  |
    0     XXX  +M
            X  |
          XXX  |
         XXXX S|  +A8-SEIT3       -A8-SEIT11
               |  +A8-SEIT13
               |
               |
               |S                 -A8-SEIT17
   -1         T+
               |
               |
               |
               |
               |T
               |
               |
   -2          +
Less able      |  Easier items
Participants   |
-----------------------------------------------------------------------------
```

Figure 23. Item-person map for Elicited Imitation Test instrument, ±assimilation with 8 syllables and slower utterance rate. A = Assimilation (+A items include assimilation, – A items do not include assimilation), 6 = 6 syllables, 8 = 8 syllables, S = utterance rate (+S = faster, –S = slower), EIT = Elicited Imitation Test, and X = 1 participant. M = mean, S = 1 SD, T = 2 SD.

Figure 24 shows Rasch item-person map for slower, shorter items with and without assimilation. Items with assimilation tended to be more difficult than items without assimilation.

```
-----------------------------------------------------------------------------
More able       |  More difficult items
participants    |
            X   |
                |
                |
                |
                |
    2           +
                |
                |
               T|T
                |
            X   |
                |               -A6-SEIT45
                |
    1          S+
            X   |S
           XX   |
            X   |
           XX   |
                |  +A6-SEIT38
          XXX   |
           XX  M|
            X   |
    0     XXX   +M
            X   |  +A6-SEIT23
          XXX   |
         XXXX  S|
                |
                |
                |
                |S
   -1          T+               -A6-SEIT35
                |   +A6-SEIT24    -A6-SEIT28
                |
                |
                |
                |T
Less able       |  Easier items
participants    |
-----------------------------------------------------------------------------
```

Figure 24. Item-person map for Elicited Imitation Test instrument, ±assimilation with 6 syllables and slower utterance rate. A = Assimilation (+A items include assimilation, – A items do not include assimilation), 6 = 6 syllables, 8 = 8 syllables, S = utterance rate (+S = faster, –S = slower), EIT = Elicited Imitation Test, and X = 1 participant. M = mean, S = 1 SD, T = 2 SD.

Research question 1 asked about the empirical hierarchy for the phonological features of elision and assimilation, and about how input rates, combined with the phonological features of elision and assimilation, affected the participants' decoding ability. Only two hierarchy patterns (rather than tendencies) can be retrieved from the analyses: (a) +E6–S > –E6–S, and (b) +A6+S > –A6+S. In the next section, I look at research question 2.

Research Question 2: The Relationship Among Language Proficiency and Elision and Assimilation

Research question 2 asked about the strength of the relationships among the participants' language proficiency as measured by the TOEFL PBT, the TOEFL PBT listening section, and the LVLT, and the decoding of the phonological features elision and assimilation. Furthermore, the question asked about the strength of the relationships among the participants' language aptitude as represented by the MLAT-E, sections 1-4, and the decoding of the phonological features of elision and assimilation.

The first part of research question 2 was answered by conducting a Pearson correlation. The assumptions for the Pearson correlation are that scores are normally distributed and that the data are measured at the interval level. A Kolmogorov-Smirnov test of normality was used to test for normality of the different scores. The percentages on the test were $p > .05$, indicating that the data were normally distributed. Each variable was continuous as they were all interval variables measuring language proficiency, listening proficiency, aural vocabulary, and elision and assimilation measured by the EIT. Finally, linearity between the variables was observed in the scatterplots.

Correlation coefficients were computed among two proficiency measures—the total TOEFL PBT score and the TOEFL PBT listening section score—and the Listening Vocabulary Levels Test, and elision, and assimilation. Using the Bonferroni approach to control for Type I error across the ten correlations, a p-value of less than .005 (.05 ÷ 10 = .005) was required for significance. The results presented in Table 39 show that all 10 correlations were statistically significant and were greater than .57. In general, the results indicate a strong link between language proficiency, aural vocabulary knowledge, and the decoding of elision and assimilation.

Table 39. *Correlations Among the Ability and the Elicited Imitation Test (N = 25)*

	TOEFL PBT	TOEFL Listening	LVLT	Elision
1. TOEFL	—			
2. TOEFL Listening	.81*	—		
3. LVLT	.75*	.58*	—	
4. Elision	.78*	.77*	.64*	—
5. Assimilation	.67*	.69*	.57*	.74*

Note. TOEFL = Test of English as a Foreign Language; LVLT = Listening Vocabulary Levels Test.
*p < .005

The second part of research question 2 was answered by computing correlation coefficients among the four aptitude scales from the MLAT-E (i.e., Hidden Words, Matching Words, Finding Rhymes, and Number Learning), and the EIT scores for elision and assimilation. Rasch person measures from each of the four sections of the MLAT-E and the elision and assimilation were used in the analysis. Using the Bonferroni approach to control for Type I error across the 10 correlations, a p-value of less than .003 (.05 ÷ 15

= .003) was required for significance. The results of the correlational analyses presented in Table 40 show that 10 out of 15 correlations were statistically significant and were greater than .67. The correlations of Number Learning and elision and assimilation were lower and not significant. In general, the results indicate that language aptitude for Matching Words, Finding Rhymes, and Hidden Words (grammar) correlate with the decoding of elision and assimilation.

Table 40. *Correlations Among the Language Aptitude and Elicited Imitation Test Scores (N = 25)*

	Elision	Assimilation	Hidden Words	Matching Words	Finding Rhymes
Elision	—				
Assimilation	.74*	—			
Hidden Words	.67*	.69*	—		
Matching Words	.70*	.74*	.99*	—	
Finding Rhymes	.70*	.73*	.96*	.97*	—
Number Learning	-.14	.05	.07	.05	.05

*$p < .003$

Research Question 3: The Relationships Among Background Variables, Elision, and Assimilation

Research question 3 asked about the strength of the relationship between the participants' English-learning backgrounds, and the decoding of elision and assimilation. Correlation coefficients were computed among an interval scale—scores for elision and assimilation—and the participants' background information from the Pre-Listening Questionnaire and the Background and Length of Residency Interview (i.e., Educational background, length of residency abroad, conversations with native speakers, and media). The scores were correlated with the data from the Background and Length of Residency Interview (e.g., Media) using Kendall's tau-b.

Table 41 shows the descriptive statistics for the Pre-Listening in English questionnaire and the Background and Length of Residency Interview raw data. Pre-Listening in English questionnaire answers showed *M* = 4.28; *SD* = .54. The skewness was .15 and kurtosis was -.35. The 95% CI for the mean difference was 4.06 to 4.50, which showed small variation in the participants' answers. A small variation in the answers might have been due to the homogeneity of the participants (e.g., Japanese, first- and second-year, university students).

Table 41. *Descriptive Statistics for Pre-Listening in English Questionnaire and Background and Length of Residency Interview (N = 25)*

	Pre-Listening in English	English Education	Conversations in English	Length of Residency	Media
M	4.28	16.24	3.84	3.84	1.12
SE	.11	.61	.27	.82	.13
95% CI	[4.06, 4.50]	[14.97, 17.51]	[3.29, 4.39]	[2.14, 5.54]	[.85, 1.39]
SD	.54	3.07	1.34	4.12	.67
Skewness	.15	-.07	-1.14	1.05	-.13
SES	.46	.46	.46	.46	.46
Kurtosis	-.35	-.19	1.26	1.12	-.58
SEK	.90	.27	.90	.13	.90

Note. All statistics were based on raw data.

Using the Bonferroni approach to control for Type I error across the 21 correlations, a *p*-value of less than .002 (.05 ÷ 21 = .002) was required for significance. The results of the correlational analyses presented in Table 42 show that 4 out of 21 correlations were statistically significant and were greater than .42. However, the significant correlations only signaled obvious similarities. For example, the Pre-Listening in English questionnaire had similar questions to the English education questions in the Background and Length of Residency interview. In general, the results suggest that elision and assimilation did not correlate with learners' backgrounds. Each variable was continuous as they were all interval variables measuring questionnaire data, interview data, and elision and assimilation as measured by the EIT.

Table 42. *Correlations among the Interval Scores and the Listening in English Questionnaire and the Background and Length of Residency Interview (N = 25)*

	1	2	3	4	5	6
1. Elision	—					
2. Assimilation	.74*	—				
3. Pre-Listening questionnaire	.19	.42*	—			
4. Conversations with NES	-.11	.15	.06	—		
5. Length of residency	.36	.36	.16	.33	—	
6. English education	.08	.36	.53*	.36	.26	—
7. Media	.12	.27	.25	.30	.34	.47*

Note. NES = Native English Speakers.
**p* < .002

Participants who performed better on the EIT had varied backgrounds and lengths of residency. Table 43 shows the three most successful participants on the EIT for elision and assimilation, and their scores on the TOEFL PBT, the LVLT, and the MLAT-E. *Mak* (pseudonym) was most successful for both elision and assimilation on the EIT.

Table 43. *Three Most Successful Participants on the Elicited Imitation Test for Elision and Assimilation*

Participant	Elision	Assimilation	TOEFL PBT	TOEFL PBT Listening	LVLT	MLAT-E
Mak	67	68	613	68	47	127
Kot	63	—	583	56	48	106
Aya	53	—	497	51	46	74
Kah	—	57	457	46	46	108
Gak	—	57	507	53	47	93

Note. All statistics were based on raw data. TOEFL PBT = Test of English as a Foreign Language Paper Based Test, LVLT = Listening Vocabulary Levels Test, MLAT-E = Modern Language Aptitude Test – Elementary.

These most successful participants had varying backgrounds and lengths of residency according to their interview data. Table 44 shows these participants' summated answers to selected interview questions.

Table 44. *Summated Answers to Background and Length of Residency Interview Questions for Three Most Successful Participants on the EIT for Elision and Assimilation*

Participant	English Education History	Conversations with Native Speakers of English	LOR	Media
Mak	13 years; 5-6 hours per day listening to English	Daily meetings with native English speakers; 20 native English speaking friends	6-11 years old, lived in Europe, including Britain; speaking only English	4-6 hours per week watching English media
Kot	6 years; 1 hour per day listening to English	Weekly interaction with native English speakers; no native English speaking friends, but 6 non-native English speaking friends	5 weeks in New Zealand during junior high school; only dormitory so little contact with native speakers	Japanese news in English three times per week
Aya	8 years; 2 hours per day listening to English	Daily meetings with native English speakers; 2 native English speaking friends	Not applicable	No English media, but loves to sing in English
Kah	11 years; 5 hours per day listening to English	Daily meetings with native English speakers; 3 native English speaking friends;	1 week homestay in the U.S. during high school; high amounts of contact with native English speakers	English language movies once every three weeks
Gak	11 years; 3 hours per day listening to English	Daily conversations with native English speakers; 2 native English speaking friends	Not applicable	Watches 2-3 English language movies every week

Note. All answers translated from Japanese. EIT = Elicited Imitation Test, LOR = Length of Residency.

All of these participants had at least 6 years of English. Furthermore, the two participants without LOR experience listened in English 2-3 hours per day. Although not significant, answers to the Background and Length of Residency interview questions suggested success on the EIT because these participants had enough interaction with the L2 to develop some ability to decode elision and assimilation as represented by the EIT.

Participants' who were least successful on the EIT also had varied backgrounds and lengths of residency. Table 45 shows the three least successful participants on the EIT for elision and assimilation, and their scores on the TOEFL PBT, the LVLT, and the MLAT-E. Participant *Kaz* was least successful for elision, and participant *Nao* was least successful for assimilation on the EIT.

Table 45. *Three Least Successful Participants on the Elicited Imitation Test for Elision and Assimilation*

Participant	Elision	Assimilation	TOEFL PBT	TOEFL PBT Listening	LVLT	MLAT-E
Kaz	38	—	460	43	46	81
Yui	39	—	489	50	45	81
Yur	39	—	463	49	44	70
Nao	—	35	513	50	47	51
Moe	—	36	480	51	45	83
Dai	—	37	473	44	47	80

Note. All statistics were based on raw data. TOEFL PBT = Test of English as a Foreign Language Paper Based Test, LVLT = Listening Vocabulary Levels Test, MLAT-E = Modern Language Aptitude Test – Elementary.

These least successful participants had varying backgrounds and lengths of residency according to their interview data. Table 46 shows these participants' summated answers to selected interview questions.

Table 46. *Summated Answers to Background and Length of Residency Interview Questions for Three Least Successful Participants on the EIT for Elision and Assimilation*

Participant	English education history	Conversations with native speakers of English	LOR	Media
Kaz	9 years; 5 hours per day listening to English	Daily office visits with native English instructor; 4 native English speaking friends	Not applicable	3 hours per week listening to music with English lyrics
Yui	9 years; 1-2 hours per day listening to English; English conversation school from 8-17 years old	Daily interaction with native English speakers; no native English speaking friends	Not applicable	2 hours per weeks spent watching an English language drama
Yur	6 years; 1 hour per day listening to English	Daily meetings with native English speakers; 10 native English speaking friends	Not applicable	English language movie twice per week
Nao	8 years; 2 hours per day listening to English	Daily meetings with native English speakers; 2 native English speaking friends	Not applicable	English language videos on YouTube
Moe	15 years; 2 hours per day listening to English	Daily meetings with native English speakers; 3 native English speaking friends	Two homestays: 2 weeks in the U.S., 2 weeks in N.Z.	One English language movie per week.
Dai	6 years; 3 hours per day listening to English	Does not usually meet native speakers; 6 native English speaking friends	Not applicable	Reads an English newspaper daily for 30 minutes

Note. All answers translated from Japanese. EIT = Elicited Imitation Test, LOR = Length of Residency.

All of these participants had studied English at least 6 years. Five participants had not gone abroad or not been to a country where the native language was English.

In this chapter I presented the results of the analyses for the four research questions that guided this study. In the next chapter, I discuss the findings of the four research questions and the links to the research questions posed in Chapter 2 and the subsequent hypotheses.

CHAPTER 6

DISCUSSION

In this chapter I discuss the findings of the study. To review, my objectives in this study were to (a) investigate any empirical item difficulty hierarchy of elision and assimilation, and the determinants of this hierarchy, (b) investigate the strength of the relationships among the participants' language proficiency as measured by the TOEFL PBT, the TOEFL PBT listening section, and the Listening Vocabulary Levels Test, and the decoding of the phonological features of sandhi-variation, elision and assimilation, and (c) investigate the strength of the relationship between the participants' English-learning backgrounds, and the decoding of elision and assimilation. The study was underpinned by the creation of an Elicited Imitation Test (EIT) that was used to test comprehension of elision and assimilation. The following discussion links to the research questions posed in Chapter 2 and the subsequent hypotheses.

Research Question 1a: Empirical Item Hierarchy for Elision and Assimilation

Research question 1a asked what the empirical item hierarchy is for two phonological features of sandhi-variation, elision and assimilation. The results showed that there was a tendency for items that include elision and assimilation to be more difficult for the participants than items without sandhi-variation. A general conceptual understanding of the item differences was possible through an inspection of the Rasch item-person map for the EIT instrument (Figure 9) and by inspecting the item difficulty measures of the phonological features displayed in the figure. The results lent support to the initial hypotheses that items with elision and assimilation would have higher difficulty estimates than items without.

The Rasch item-person maps for both the items with and without elision (Figure 10) and with and without assimilation (Figure 11) on the EIT instrument showed a general tendency for items with elision and assimilation to have higher difficulty estimates than those without. These results are broadly consistent with two previous studies related to sandhi-variation by Henrichsen (1984) and Ahn (1987). Henrichsen (1984) hypothesized that, for native speakers of English, no significant difference would exist between items with and items without elision and assimilation. He further hypothesized that such a difference would, in fact, exist for non-native English speakers. His results supported the hypothesis. Henrichsen’s hypothesis held true in the current study, as the native English speakers who were tested with the EIT completed the assessment with 100% success, regardless of the presence or absence of sandhi-variation in the target items. Also consistent with Henrichsen’s hypothesis was the finding that none of the native Japanese speakers in this study were able to complete the EIT with 100% success.

Ahn (1987) conducted a study similar to Henrichsen’s by using items with and without sandhi-variation (i.e., ± elision and ± assimilation) to discover the filter effect to the comprehension of English, which he defined as limitations in aural processing ability that create a “barrier to comprehension” (p. 5). The filter effect, which was assumed to be a part of speech comprehension in English as a foreign language, was broken down into five factors: affective factors, sandhi-variation, syntactic complexity, semantic familiarity, and cultural difference. He found results similar to Henrichsen’s: Non-native English speaking participants' scores were low for test items with sandhi-variation compared to the test items without sandhi-variation. Ahn also investigated how attitude and motivation influence speech comprehension, but he concluded that these variables exhibited only a minor influence when compared to the effect of sandhi-variation. Like Ahn's conclusion, the findings of the current study support the notion that test items with sandhi-variation negatively affect aural comprehension for non-native English speakers (Table 26).

The current study differs from the Henrichsen (1984) and Ahn (1987) studies in several ways, due, in part, to how I have, in this study, assessed several variables, such as language proficiency levels. Moreover, both Henrichsen (1984) and Ahn (1987) used cloze-like written instruments to assess the participants' comprehension of the aural targets rather than listening assessments. The participants completing the cloze-like assessments could read the sentence minus the missing target, thus gaining written information from the context that might have helped them fill in the missing target. The EIT used in the current study, on the other hand, tested the aural abilities of each participant without using written text prompts, that is, the test involved only listening and an oral response.

A second difference between the current study and the previous studies is that in the current study, validity evidence was provided for the EIT using Rasch analyses; the test was shown to be reliable and valid for the Japanese university students in this study. This approach stands in contrast to the instruments used in previous studies that were both borrowed from other sources and not validated by the researchers.

A third difference between the current study and the Henrichsen (1984) and Ahn (1987) studies is that, in those studies, lexical baselines for their instruments were not set.

Without lexical baselines for their instruments, lexical items used as cloze targets or within the target sentences (i.e., cloze prompts) might have had a filtering effect on comprehension in lieu of the target if, for example, a participant did not know a particular word or phrase in a target sentence, or if a particular word or phrase was a low-frequency item. In addition, the participants in previous studies were not assessed for baseline listening proficiency and language learning aptitude. The participants in both Henrichsen and Ahn's studies were chosen and grouped based on class proficiency levels (e.g., "*low* level" [Henrichsen, 1984, p. 108]) within the respective institutions where the studies were conducted.

Although recent studies about the effects of sandhi-variation on second language aural processing are sparse, many have been focused on distinct phonological effects. However, these studies differ somewhat from the current study in their specificity. For example, researchers have investigated how specific aspects of assimilation (Mitterer & Blomert, 2003; Weber, 2001) and reduced word forms (Kohler & Niebuhr, 2011) affect processing in German and Dutch. Both German and Dutch, like English, are Germanic languages, so the findings are heuristically informative. The findings for these studies are similar to the studies with English as the L2, and they contain a common thread: Phonological adjustments to the speech stream affect second language processing. Furthermore, Kohler and Niebuhr (2011) have argued that regressive assimilation hinders processing more than other forms of assimilation, such as progressive assimilation, and that articulatory prosodies help listeners process reduced word forms. In the current study I investigated the general effects of both elision and assimilation on listening comprehension and parsing, not the conditional effects of those features. The one item with progressive assimilation (i.e., +A8+SEIT26: drimzwɪð) on the EIT was approximately in the middle of the item difficulty hierarchy, so it is impossible to say whether regressive assimilation hindered processing more than progressive assimilation from the data.

The current study can only be considered a link to recent studies that have been focused on the influence of L1 phonology and vocabulary in determining the ways that learners process L2 speech. In the current study, the participants' L1, Japanese, is a morae-based language, whose syllable-weight system is different from the stress-timed rhythm of English. Moreover, the inventory of phonemes in Japanese differs from those in English. In other words, the phonemic building blocks of Japanese are not only constructed differently, but also combined in ways that are not used or heard in English. (See, for example, Munro & Bohn, 2007, for insight into the vast amount of literature on this topic.) For both English and Japanese, assimilation, for example, is a phenomena that has been shown to be affected by language differences, such as differences in detecting place-assimilated words (Darcy, Peperkamp, & Dupoux, 2007), by distinguishing L1 phonemic contrasts that are non-existent in the L2 (Broersma, 2005), and by differences in individual choices for placing assimilation on words when it is optional (Ellis & Hardcastle, 2002). Of the three examples shown above, an example of an L2 phonemic contrast from the current study is most pertinent in contrasting with the morae-base in the L1 (Japanese): Item 13 (Rasch item difficulty = -.50), has assimilation (+A), 8 syllables, and a slower utterance rate (-S): *Did your father used to* *work there* [wɜrg ðɛr]. In this example, two sounds [k] and [ð], from the underlined words—neither of which have phonemic equals in the L2—assimilate. Furthermore, participants had to comprehend the subtle assimilation in order to parse the two words *work* and *there* for the EIT. The target L1 sound system is a filter of phonological processing for non-native listeners of English, which is evidenced in the current study by the specific outcomes that ran counter to the general outcomes. I discuss this further later on in the second part of the research question 1a section.

The results of the current study are broadly consistent with one recent study with English as the L2. Révész and Brunfaut (2013) investigated several task factors on ESL listening comprehension, including input rate, phonological complexity (e.g., elision), and lexical complexity. Similar to the current study, Révész and Brunfaut did not find significant effects for utterance rates. Their finding was explained by the similarity of the two listening speeds used in the targets: a mean rate of 2.26 syllables per second and 3.04 syllables per second. The difference in means of .78 syllables per second for the two speeds was not significant. The method used by the researchers to control the speed of the targets for two exact rates was not stated. Furthermore, not setting exact rates with very similar syllable per second rates risks an overlap of rates. However, unlike in the current study, Révész and Brunfaut found no significant effects for elision, which they suggested might have gone undetected by the participants as evidenced by the low standard deviations in the phonological complexity variables, and because the participants had advanced general English proficiency and advanced listening proficiency; as a result, the targets did not cause comprehension difficulty. That said, phonological variables in the current study, such as the presence or absence of assimilation, tended to show differences, which might indicate that the participants detected them.

Like Révész and Brunfaut (2013), I have, in the current study, provided a support for understanding second language comprehension, by analyzing such variables as aural vocabulary knowledge, which is commonly associated with second language listening ability, using the LVLT. Moreover, the LVLT served as a way to ascertain that the participants likely knew the vocabulary on the EIT. However, the participants' vocabulary knowledge could have been influenced by the EIT targets in other ways. For instance, they might not have recognized words they had acquired due to phonological changes within the utterance, which might have impacted their lexical processing.

Cutler (2015) explains one such impact on lexical processing as *phonemic confusion*, whereby one activated phoneme lingers into another and creates ambiguity. Consider item 45 on the EIT (– assimilation, 6 syllables, and – speed), *It is painted bright blue*, which is one possible example of phonemic confusion in the current study. This item contains the alliteration *bright blue*, but the combination of the [br] in *bright*, which is activated, potentially lingers into the [bl] sound in *blue*, creating ambiguity in the target not related to assimilation (i.e., a possible confound). In other words, the physical and cognitive transition, that is, beginning with a voiced bilabial stop and rhotic consonant and moving into a voiced bilabial stop and the alveolar lateral approximant, from the [br] into the [bl], is confusing (i.e., [braɪblu]). This is especially true when neither the [br] nor the [bl] consonant cluster is in the participants' L1, and the [l] and [r] distinction is not made in Japanese.

Despite this study's similarity to prior studies, to the best of my knowledge, no prior researchers have investigated a possible hierarchy of comprehension for the phonological features of elision and assimilation. Determining the hierarchy in the current study required an inspection of the Rasch item-person maps for the EIT instrument. The item difficulty measures of the phonological features of elision are shown in Figure 9 and those for assimilation are shown in Figure 10; they show little in the way of a defined hierarchy for positive (e.g., +E) versus negative (e.g., -E) items. First, there was a tendency for the participants' comprehension of targets with and without elision to be different but not clearly hierarchical, due, in part, to the embedded confounds, utterance rate and syllable length. The difference does show that there is a hierarchy overall. Figure 9 shows, for example, that some items with elision are more difficult than items without elision, and that others are not. Moreover, the 95% CI (Figure 16) show the trend in the data for the items with elision and assimilation to be more difficult than those without, which leads back to the second part of research question 1a.

The second part of research question 1a concerns the strength of the hierarchical placement of items with and without elision and assimilation; the answer to this part of the question is somewhat unclear in the data. The results of four dependent *t*-tests were used to further explore the strength of the empirical hierarchy. Embedded in the items with and without elision and assimilation were two speeds—4.7 syllables per second and 3.17 syllables per second—and two syllable lengths—8 syllables and 6 syllables—for each sentence length utterance (item). The embedded items, including the sandhi-variation features or absence of the features, were hypothesized to be determinants of any hierarchy. For example, items embedded with elision were designed to be more difficult items. Therefore, the *t*-tests were conducted to compare four conditions: (a) items with and without elision, (b) items with and without assimilation, (c) items with different utterance rates, and (d) items with different syllable lengths. Items with elision were found to be significantly more difficult than items without elision. The comparisons for b, c, and d were not significant, though the differences in scores for each comparison, seen in the CI (Figure 16), were in the hypothesized directions.

One theoretical explanation why sandhi-variation showed a tendency to be more difficult is based on coarticulation theory (Diaz-Campos, 2004). Coarticulation theory broadly stipulates that due to overlap of adjacent phonetic segments—actual physical segments that exist in time as opposed to abstract, phonological segments (Tatham & Morton, 2006, p. 73)—each phonetic segment is affected by or affects the adjacent or nearby segment, and, therefore, is not realized (i.e., uttered) identically (Kuhnert & Nolan, 1997; Salverda, Kleinschmidt, & Tanenhaus, 2014). The slight differences in aural realization (e.g., phonological effects such as a regional accent) can cause, in turn, variations in aural processing within general sandhi-variation types such as assimilation, because the utterances are unique to the paired segments within the targets. In other words, some targets in the current study might have been more difficult for the participants because of the degree of, or variation within, the effect, as opposed to simply the existence of an effect. Consider Item 7 (+A8+SEIT7), *The best estimate is secret*, where *best* assimilates with *estimate* as [bɛs ɛstəmət]. As a target, Item 7 has assimilation, and the confounds speed and length, yet Item 7 was relatively easy for the participants (Rasch difficulty estimate = -.51) compared to an item hypothesized to be easier such as Item 11 (-A8-SEIT11) (Rasch difficulty estimate = -.35), *I did not recognize the spot*, where *recognize* does not assimilate with *the* [rɛkəgnaɪz ði]. In the example, the filter of comprehension is unclear. Perhaps it is the [z] that filters comprehension as the sound is non-existent in the participants' L1, and sits next to the phonologically challenging [ð], which is also non-existent in their L1. The combination of [z] and [ð] might create phonemic confusion. The participants in the current study could comprehend one segment of one target item with, for example, assimilation, and then not comprehend another target item with assimilation because the two targets were unique, due to the variability in the specific speech signals. The physical mechanisms used for the production of each target item can easily create variations of, for example, stress or intonation, that might vary utterances in such a way that one utterance is more difficult to comprehend (Tatham & Morton, 2006). The effects of coarticulation together with the participants' level of experience and the specific target items could have created filter effects for comprehension on the EIT, regardless of the participants' knowledge of the vocabulary in the target utterances.

Table 47 shows a small sample of the more difficult items on the EIT. The table shows that elided phrases and the phoneme values of assimilation for the six most difficult items. The meaning of the items could also have caused difficulty for the participants. For example, item +E8+SEIT16, *Your joint effort to win paid off*, required the participants to comprehend an imagined scenario where two or more players worked together to win something. The item has combined expressions *joint effort* and *paid off*, both of which require collocational knowledge to be comprehensible. Furthermore, controlling for the conceptual content of the item, as the EIT does by including a second-stage, parsed repetition focused on the elided portion, the most difficult item (Rasch difficulty estimate = 1.65), +E8+SEIT16, *joint effort*, is heard as /dʒɔɪn ɛ.fərt/. Participants who are unfamiliar with the expression might interpret *joint* with an elided [*t*] to be *join*. In other words, the embedded filter effects of the experiment might or might not be triggered depending on other filters such as lexical and collocational knowledge in each target item and on each participant's negotiation of the filters. Thus, the variability of the speech stream, such as the use of elision, can negatively influence phonological processing.

Table 47. *Examples of More Difficult Items for EIT Instrument*

Item code	Rasch item difficulty estimate	Item [Phonemic Alphabet]
+E8+SEIT16	1.65	Your joint effort to win paid off. [dʒɔɪnɛ.fərt]
+A6+SEIT41	1.60	We breathe in air deeply. [brið ən]
+E8-SEIT43	1.45	Will the owners of the coat be found? [wi ði]
-A6-SEIT45	1.28	It is painted bright blue. [braɪt blu]
-A8-SEIT48	1.24	The actual labor was cheap. [wʌz ʧip]
+A6+SEIT2	1.20	He disappeared in snow. [ɪ snoʊ]

Note. All statistics were based on Rasch logits.

Research Question 1b: Input Rates Combined with Elision and Assimilation

Research question 1b asked how the two input rates affected the participants' comprehension when combined with the phonological features of elision and assimilation. The descriptive statistics indicated that there was a tendency for items with elision and the faster input rate to be more difficult than items without elision and the slower input rate. Twelve paired-sample *t*-tests were conducted for each combination of phonological effect (i.e., elision and assimilation) and utterance rate. None of the comparisons between input rates and the elided and non-elided targets were significant, low effect sizes < .40 were found for each combination, and the 95% CIs overlapped. The non-significant relationship between combinations of the utterance rates and phonological features of the items on the EIT indicates that the participants' responses were reconstructive in nature, so they did not rely on rote memory. In other words, the participants, when successfully comprehending targets, could process the target items' input rates as opposed to parroting the target items (Yann et al., 2015).

Likewise, all comparisons between input rates and the assimilated and non-assimilated targets were not statistically significant, which suggest that the participants could successfully process both target utterance rates. Also, small effect sizes < .40 were found for each combination (e.g., +A and –S versus +A and +S) and the 95% CIs overlapped.

For each target type, it is possible that the two utterance rates in the current study were not perceived differently because the maximum syllable count size for the target sentences was 8 syllables. According to Yann et al. (2015), 8 syllables is a medium size. The medium and small (i.e., 6 syllables) syllable counts of the items in the current study could have muted the effect of the utterance rate confound. However, the decision to have a medium-sized syllable count was taken after conducting several pilot tests because a medium-sized syllable count allowed for a range of correct and incorrect responses. Erlam (2006) also suggested that the length of the sentence targets should be comprehensible to the participants instead of a prescribed absolute that, presumably, does not take into account the participants' aural processing levels. In pilot testing conducted with high-proficiency, non-native English speakers for the current study, the relationship between the targets and the participants was that targets with more than 8 syllables were too difficult, as the participants could not repeat any of these pilot targets when combined with the embedded phonological effects. Rather than having targets that were beyond the participants' abilities, 8 syllables was chosen as the maximum length. However, this decision might have caused the lack of significance in the utterance rate variable.

Research Question 2: Language Ability Correlated with Elision and Assimilation

Research question 2 concerned the relationship between the participants' general English proficiency, listening proficiency, listening vocabulary level, and aptitude, and the phonological features of elision and assimilation. The question was first examined using a Pearson correlation with raw total scores of the TOEFL PBT, Rasch measures for the Listening Vocabulary Levels Test, and Rasch measures for the EIT. The results showed that all 10 coefficients correlated significantly at $p < .005$, which provides evidence for strong, positive relationships ($r > .57$) among measured variables. The results indicate that general English proficiency (TOEFL), and more specific listening and vocabulary proficiencies (TOEFL Listening and LVLT) support aural processing of the phonological features of elision and assimilation.

Likewise, a body of research shows a strong correlation between elicited imitation tests and general language proficiency (e.g., Burger & Chretien, 2001; Jensen & Vinther, 2003). The results indicate a controlled version of how second language learners process the English speech stream because participants' processing abilities include general proficiency, listening proficiency, and listening vocabulary proficiency.

Furthermore, the data showed that correlation coefficients were higher for elision and the general English proficiency scores provided by the TOEFL, TOEFL Listening, and LVLT than for assimilation and the same scores (Table 39). One possible reason for the difference is that elision and assimilation require different levels of proficiency for second language learners. An elided sound within an utterance affects the meaning of an utterance in a way that assimilation does not, for example, an assimilated sound within an utterance might be perceived as creating a non-word. For instance, *ten people* [tɛn pipəl] becomes *tem people* [tɛm pipəl] with assimilation, and *tem* is a non-word in English. In contrast, an elided sound within an utterance might create a new word. For example, *next spring* [nɛkst sprɪŋ] becomes *neck spring* [nɛk sprɪŋ] with elision, and *neck* is a word in English, which changes the meaning of the original word. However, because of the effect of these acoustic cues on the conveyance of meaning and given that the act of speaking is a cognitive process, native English speakers are cognitively aware that specific dropped sounds will or will not alter the meaning of an utterance (Tatham & Morton, 2006). Examples from the current study include Item 6 (+E6-SEIT6), *Send them to me by mail*, where *send them* is elided as [sɛndəm], creating ambiguity in the second part of the combination of words with [əm], and Item 16 (+E8+SEIT16), *Your joint effort to win paid off*, where *joint effort* is elided as [dʒɔɪn ɛ.fərt], creating a new initial word, *join*. Elision, then, requires more knowledge of how the English sound system can change in a variety of lexical environments. Assimilation, on the other hand, can be accomplished by just producing an utterance quickly (Roach, 2009), so the sound changes, relative to elision, are minor and no syllables are lost. Examples from the current study include Item 2 (+A6+SEIT2), *He disappeared in snow*, where *in snow* is assimilated as [ɪˈsnoʊ], and the change amounts only to a loss of the pause between the words; and, Item 12 (+A6+SEIT12), *Her daughter writes quite well*, where *quite well* is assimilated as [kwaɪwɛl], and again the change amounts to the loss of a pause between the two words. Thus, as has been argued by other researchers (e.g., Cutler, 2015; Jusczyk, 2000), first and second language listeners need to have acquired underlying language skills, such as the accurate processing of acoustic cues, in order to create aural maps of target utterances before they can comprehend them. In other words, listeners cannot fully comprehend a message if they do not know the parts that make up the message.

This research question was further examined by conducting a Pearson correlation of Rasch measures for four subsections on the Modern Language Aptitude Test–Elementary (MLAT–E) and the Rasch measures from the EIT. The results showed that three of the four MLAT–E sections—Hidden Words, Matching Words, and Finding Rhymes—had a strong ($r > .73$) positive relationship with the EIT scores for elision and assimilation. A fourth section, Number Learning, did not have a significant correlation with elision and assimilation. The MLAT–E results describe for the first time the strength of the relationship between facets of second language aptitude and the ability to comprehend the phonological features of elision and assimilation.

First, the Hidden Words section of the MLAT–E measures knowledge of English vocabulary and sound-symbol association ability. There was a strong positive correlation between the Hidden Words section and the EIT, $r(25) = .73$ ($p < .005$). Participants with strong abilities to perform on the EIT would also have an inventory of target phonemes to use for associations. In both the EIT and the Hidden Words section participants must recognize known vocabulary and make aural sound associations, which can account for the positive correlation between the two tests. For example, Hidden Words Item 1, *rivr* (answer: *large stream of water*) is a lexical item from the 2,000 high-frequency words of English according to the BNC/COCA list, which is the same list used to control for the frequency of lexical items on the EIT (i.e., also limited to the 2,000 high-frequency words of English according to the BNC/COCA list for all words used in each item). In other words, the correlation is partly a product of the participants' knowledge of the high-frequency lexical items on both tests, and the high-frequency items relate to strong aural sound associations. This finding is in line with the theory that the attention required by a second language listener for linguistic coding can be reallocated when the discourse is familiar (Kormos, 2013).

Second, the Matching Words section of the MLAT–E measures sensitivity to grammatical structures by way of associating the function of a word in one sentence with the task of recognizing a word in the target sentence with the same grammatical function. There was a positive correlation between the Matching Words section and the EIT, $r(25) = .75$ ($p < .005$). The participants had all had a minimum of 7 years of EFL and/or ESL English instruction, with grammar instruction assumed to be a part of that instruction, as it is a common part of junior high and high school English instruction in Japan. Moreover, grammar is woven into the structure of language, which includes the skill of listening. For example, Item 13 from the Matching Words section has the prompt, *Alice PLAYS the piano very well*, and the linked sentence, *Little babies drink lots of milk*. The participants must find the word in the linked sentence that performs the same grammatical operation as the word in all capitalized letters from the prompt. In this case, the prompt, *PLAYS* (capitalization from original) is a transitive verb, so the correct answer is *drink* because it is the transitive verb in the linked sentence. This type of grammatical matching task was within the participants' abilities. While the participants were not required to identify specific grammatical items in the EIT, grammar is assumed to be part of the identifying feature of the imitation. In other words, the participants had to process the targets in the EIT in order to repeat them, and therefore they needed to comprehend the lexical items and probably, also, the grammatical structures of the utterances. Moreover, all of the targets in both the EIT and the items in the Matching Words test are grammatically simple sentences; no compound or complex sentences were used. The Matching Words test items correlated strongly with the EIT items because of the simple sentence structures and the high frequency lexical items on both tests.

Third, the Finding Rhymes section of the MLAT–E measures the participants' abilities to hear speech sounds by asking them to find words that rhyme. There was a positive correlation between the Finding Rhymes section and the EIT, $r(25) = .77$ ($p < .005$). Although rhyming is not found in Japanese beyond the mora system of the language, for example, *ki* rhymes with *chi*, and, anecdotally, few of my students can produce with rhymes in English if asked, the participants were knowledgeable enough of phonology to find rhymes. Phonological ability is required also in performing on the EIT. In this section of the MLAT–E test items, the items in Finding Rhymes are all either within the 2,000 high-frequency words of English according to the BNC/COCA corpus. Definitions of the words used as prompts and answers are not required in this section, and the prompts and answers were all single syllable words. For example, the prompt *WARM* (capitalization from the original) has the answer choices, *arm*, *form*, *lamb*, and *term* (correct answer *form*), which are all high-frequency words of English according to the

BNC/COCA. The high-frequency test item words afforded the participants access to known aural sound associations that include the distractors, so the correct answers were accessible. The EIT also relied on high-frequency lexis and known aural sound associations to allow for comprehensible targets with elision and assimilation, syllable length (8 and 6 syllables), and utterance rate (4.7 and 3.17 syllables per second) confounds. The Finding Rhymes section is most similar to the EIT—it had the strongest correlation—in that the participants had to comprehend the aural sequence of phonemes within the target items on both tests.

Fourth, the Number Learning section of the MLAT–E was designed to measure the memory component of language aptitude. According to the MLAT–E manual (Carroll & Sapon, 2010), this section has a "large specific variance, which one might guess to be a special 'auditory alertness' factor which would play a role in auditory comprehension of a foreign language" (p. 2). There was a slightly negative correlation between the Finding Rhymes section and the EIT, $r(25) = -.04$ ($p > .005$). As stated above, there was also a "large specific variance" for the participants in the current study. They were either able to get all of the items correct, or missed them all. This outcome could have been related to whether or not they were paying attention during the instructions, which is when the numbers were defined. While the Finding Rhymes section might test one form of aptitude, auditory alertness was more tested in the EIT by the 48 items that varied each item with new words and unknown syntax as opposed to the fixed and learned numbers in the Number Learning section. Furthermore, unlike the items in the Finding Rhymes section, the items on the EIT were designed to not place stress on the participants' working memories. The 6- and 8-syllable items on the EIT were short and within the ability of the participants to process.

Research Question 3: Background Correlated with Elision and Assimilation

Research question 3 concerned an investigation of the relationship of the participants' English-learning backgrounds, and listening comprehension as measured by the TOEFL PBT and the Listening Vocabulary Levels Test and the phonological features of elision and assimilation. This research question was first examined by using a composite of *z*-scores calculated using Rasch person measures. The composite score of the TOEFL PBT, the Listening Vocabulary Levels Test, and the EIT scores were called *ability*, and each participant's ability score was then correlated with raw scores from the Pre-Listening Questionnaire data and Background and Length of Residency Interview data. The results of the correlation analyses showed that only one of the 21 coefficients—the Pre-Listening in English Questionnaire—correlated significantly at $r = .42$ with the participants' abilities to recognize assimilation. Perhaps this moderate correlation occurred because the Pre-Listening in English Questionnaire questions concerned the participants' feelings about listening in English and the sound system of English, both of which are required skills on the EIT. In contrast, the conceptual content of the items on the EIT were arguably basic enough so that experience living abroad was not required for processing the items on the test. Three more of the 21 coefficients also correlated: assimilation with elision at $r = .74$, English education with the Pre-Listening in English Questionnaire at $r = .53$, and English education with Media at $r = .47$. Elision and assimilation are both forms of sandhi-variation within the EIT, so they are similar in scope. The Pre-Listening in English questionnaire was also similar to the English education interview questions. However, English education and Media focused on different types of input (e.g., how much time spent studying versus time spent watching movies). Participants might have counted their non-classroom English experience as their interaction with media.

To the best of my knowledge, no previous researchers have investigated the relationship among residency or background knowledge with the ability to process sandhi-variation. The lack of a significant correlation in this investigation can perhaps be explained by the homogeneity of the participants, who were Japanese university students with similar backgrounds. Most of the participants had not been abroad and were limited to the 6-7 years of English instruction required by the Ministry of Education in Japan as they moved through junior high school and high school. All but 3 of the 25 participants had a narrow range of answers regarding their English language learning experiences and time abroad; most of their experiences prior to entering the university did not include classroom time spent working specifically on English listening skills or English speaking skills. Instead, they were working on grammar rules. Furthermore, few of the participants had experienced extended interactions with native English speakers prior to university. The limited range of variability within the pool of participants in terms of English speaking and listening experiences and experience abroad led to attenuated correlations.

That said, specific participant experiences varied, including examples such as *Aya* (all participant names are pseudonyms) who had very little experience, despite her 6 years of junior high and high school English instruction, with speaking or listening in English until she reached the university level. Nevertheless, *Aya* had a high score for elision on the EIT. Her high score for elision might be related to her TOEFL PBT Listening score, which was also high (Table 41). *Mak* had lived outside of Japan for 5 years, and he had the highest score for elision and assimilation on the EIT (Table 41). Overall, despite individual differences in exposure to English, the participant pool lacked broader experiences such as interacting with native English speakers, fluency-development situations (e.g., speaking practice in English), and residencies abroad.

Comparing the data from Tables 43 and 45, two obvious differences between the data were scores for the TOEFL PBT and the MLAT-E. The most successful participants' group had three participants who scored > 500 on the TOEFL PBT, while the least successful group had one participants who scored > 500. Furthermore, the most successful participants' group had four participants who scored > 90 on the MLAT-E, while the least successful group had none.

Comparing Tables 44 and 46, one potentially important difference was in the participants' scores for LOR. Three of the five participants in the most successful group answered that they had some LOR. In contrast, one of the six participants in the least successful group had some LOR. Perhaps experience abroad, even from a short residency, stimulated phonological awareness in the participants in support of their decoding abilities on the EIT.

A few researchers have investigated participants' backgrounds and background knowledge and listening comprehension (e.g., Anderson & Lynch, 1988; Fujio, 2007, 2010; Nishino, 1992; Oka, Yamauchi, & Fujio, 2005). While knowledge of semantic content has been shown to support second language listening comprehension when the content is directly related to the target listening text (i.e., bottom-up processing), global background knowledge (i.e., top-down processing) has not yet been shown to directly support second language listening *en masse* (Anderson & Lynch, 1988; Fujio, 2007). Direct links between second language processing and background, as operationalized by educational background and years studying English, and length of residency in countries where English is the primary language can easily be found at the individual level. Longer residency abroad supports better aural processing (e.g., sandhi-variation). Reasons for this include stimulus exposure (Au, Knightly, Jun, & Oh, 2002; Wright, Baese-Berk, Marrone, & Bradlow, 2015) inherent in the residency, even if the residency is casual, and, if the resident abroad uses the L2 (Wright & Sabin, 2007).

Implications for Models of Second Language Listening Comprehension

The results indicated that enhancing understanding of the phonological effects of sandhi-variation was important because it directly influenced how ESL participants coped with the speech stream in English. This section included five implications for models of second language listening research, including (a) insight into how learners deal with elision and assimilation, (b) clarification of previous studies related to second language listening, (c) the use of Rasch measurement as the basis for conducting significance tests, (d) instrument validation, and (e), the links between background knowledge and length of residency to elision and assimilation.

First, the participants' measures on the TOEFL PBT, the Listening Vocabulary Levels Test, the Modern Language Aptitude Test—Elementary, and the EIT indicated links between phonological features of elision and assimilation and second language aural processing. The links provided insight into how learners deal with sandhi-variation when processing aural input. The results provide evidence that (a) the phonological features of elision and assimilation affect the listening comprehension of non-native English learners, and (b) there is evidence of a hierarchy of difficulty of sandhi-variation that learners follow as they develop L2 listening proficiency, namely utterances without the phonological effects of elision and assimilation are more easily comprehended.

Second, the results clarify previous studies related to second language listening comprehension. For example, the results support the theory that processing constraints involved in parsing the speech stream at or below a second language learner's proficiency level can still negatively influence the learner's ability to detect known linguistic information (Field, 2008b; Rost, 2015; Skehan & Foster, 2001). In other words, even when learners have adequate schema, lexical knowledge, and morpho-syntactic knowledge, they might not understand an utterance because of problems with phonological processing caused by features such as elision and assimilation. For example, Item 6 (+E6-SEIT6) *Send them to me by mail*, had a slower utterance rate and was a six-syllable item, yet more than two-thirds of the participants were unable to understand the utterance well enough to repeat it, due, arguably, to the elided portion, that is, [sɛndəm]. This theory of processing constraints is supported by the current study because the results showed that the participants were challenged to comprehend the utterances even with full knowledge of the lexical items on the EIT, with items that had slower input rates of 3.17 syllables per second and that were 6 syllables in length. Also, all of the items on the EIT were at or below the participants' proficiency level, but the results still exhibited a hierarchical pattern that signified reduced comprehension dependent upon elision and assimilation of the target items. The results of the current study give a clearer understanding of the facets required by second language listeners to process the speech stream because the results show that utterance rate, utterance length, elision, and assimilation can affect decoding for second language listeners and filter acquisition. For second language acquisition research, having a better understanding of the filters to aural comprehension sets a baseline for further studies of aural input, especially regarding the types of aural input that participants can comprehend.

Third, using Rasch measurements as the basis for conducting *t*-tests and Pearson correlations moved the research of second language speech processing forward by placing the measurement of sandhi-variation on an interval scale, as well as making fit indices and dimensionality statistics available. The analyses provide precise statistical results for the relationship between second language learning participants and the semantic mode of listening. The current study also provides a basic model for how second language listening content used for comprehension should be measured in order to insure stability and replicability of item and person estimates.

Fourth, while both Ahn (1987) and Henrichsen (1984) used *t*-tests and Pearson correlations, neither researcher used Rasch measurement to validate their instruments; as a result, their data were not placed on an interval scale. Moreover, neither of the previous researchers secured participants' scores for language proficiency (TOEFL PBT), listening vocabulary level (LVLT), and aptitude (MLAT–E) prior to testing. The results of the current study give a more accurate presentation of learner processing of sandhi-variation, and more insight into the type of input that English language learners can process. Furthermore, Rasch measurement was used to identify a hierarchy of difficulty for processing elision and assimilation. Frequency of error-type was investigated, and the tendency (see Figure 16) was for items with sandhi-variation to be more difficult. The information from this study can help to specify types of sandhi-variation errors made during a test of second language listening comprehension and thereby help to identify barriers to L2 listening comprehension discussed by L2 researchers (Buck, 1988; Field, 2008a; Goh, 2000). Specific identification of errors was and still is limited for L2 speech perception and processing research, but the current study adds to the body of information.

Fifth, the results of the current study did not identify a significant link to background and length of residency; thus, it is difficult to establish a connection between such variables and the phonological features of elision and assimilation. Schrooth, Smith, and Kyles (2015) described background, including educational background, in terms of learner preferences toward different aspects of language (e.g., grammar). In other words, learners prefer some language learning activities over others based on their experiences with language learning. Perhaps the lack of a significant link between background and length of residency and the results on the EIT indicates that attitudes to language were not part of processing the phonological features of sandhi-variation. In other words, the items on the EIT were not salient to the general analysis of background and length of residency analyses included in this study. However, the conceptual content of the items on the EIT was basic enough that living abroad was not required for processing the test items. After looking again at the lexical items, there is no obvious reason to think that a person would need overseas experience to understand them.

Pedagogical Implications for the Second Language Classroom

In this section I describe three pedagogical implications from the current study for the second language classroom, which include the need for (a) practice and exposure to known aural vocabulary within the speech stream, (b) practice parsing of the L2 speech stream, and (c) exposure to problematic constructions and unique properties of the L2 speech stream.

The results of the current study showed that English language proficiency, listening comprehension, vocabulary knowledge, and aptitude were associated with the decoding of phonological effects of elision and assimilation. An effective way to facilitate automatized decoding is through combinations of productive practice of and exposure to large amounts of comprehensible L2 aural input (Wright, Baese-Berk, Marrone, & Bradlow, 2015). Learners need opportunities to develop aural fluency and acquire aural vocabulary, and they need to have as many opportunities as possible to listen to the target language, both simplified and authentic. They also need to be able to construct and strengthen a sound-map that works with the target L2 (Cutler, 2015). The process of developing or strengthening listening fluency takes time and opportunities to listen. Furthemore, a word's pronunciation, along with its context, becomes more ingrained in a learner's memory with experience encountering and using the word. Listening fluency takes a large amount of comprehensible aural input that has little to no unknown lexis that learners can become used to.

One practical way to provide more aural input, including exposure to aural vocabulary, to learners is to assign extensive listening texts that can be used outside of the classroom (Chang & Millet, 2014; Rost, 2016). In this way, learners can control the playback in terms of the number of times and repetition of more challenging parts in a target listening text. Students in my own classes often report that they listen to graded-reader book-length extensive listening texts more than five times. Each listening of the target audio texts gives them an opportunity to make phonological-semantic connections and to develop their sound-maps.

Moreover, learners need to develop strong parsing skills for the target language sound stream. This process can be aided by explicit tasks that focus' learner attention on word boundaries. Field (2008a) provides many examples of such exercises. For example, one exercise asks the listeners to count the number of words in a sentence.

Aural input: *Does your family live here still?* (6 words)

Students are asked to count the words that they have heard, thereby focusing their attention on parsing. The example sentence above includes the opportunity to work on parsing elided segments (i.e., live here = lɪvhir).

Another task, also suggested in Field (2008a) focuses the listeners' attention on sandhi-variation in another way:

Aural input (choices): a. *Does your family live here still?*

b. *Did your family live here?*

c. *Do families live here still?*

Teacher query: *Which sentence did you hear?*

Learners then focus their attention on function words rather than content words (e.g., the slight differences between *Does your, Did your* and *Do*) in order to work on assimilation, auxilliary tense, and recognizing syllable stress within the utterance.

Rost (2016) suggested several strategies to develop second language listening skill, including intensive listening, focusing, for example, on phonology, and extensive listening, focusing, for example, on listening continuously to large amounts of aural input. Both forms of instruction present situations for learners to understand, utilize, and respond to input. How learners interact with the input leads to achievement (Rost, 2015).

Other researchers, such as Celce-Murcia et al. (2010/2016), suggested exercises for second language learners that include elision and assimilation. These types of exercises introduced phonological features, such as elision, and drills to familiarize learners by, for instance, identifying “silent syllables” (Celce-Murcia et al., 2010/2016, p. 180), which could increase learners’ awareness that such phonological features exist in the L2. However, there was no empirical evidence available to support them as beneficial as support for second language listening acquisition. In my own classroom experience, I have found that such exercises were phonologically rigid. In other words, the changes represented by target examples were possible but also, inevitably, of low frequency because of the variation inherent in spoken language and in sandhi-variation.

Finally, by spending time focusing attention on problematic constructions and unique properties of the L2 (e.g., speech stream differences, in this case stress-timed and morae-based speaking cues), instructors can give learners opportunities to notice more challenging areas of the speech stream, and help learners understand how to negotiate them. This type of explicit instruction coupled with input frequency expedites the possibility of learner intake (Ellis, 2002; Wright et al., 2015). In other words, learners need a great deal of comprehensible input with instructional guidance to notice the unique aspects of the target language input, including sandhi-variation.

In this chapter, I discussed the outcomes of the current study. I first discussed the four research questions. Next, I discussed implications for models of second language listening comprehension. Finally, I discussed pedagogical implications for the second language classroom. In the following chapter, I present my conclusion to this study, which includes a summary of the findings, limitations of the study, suggestions for future research, and my final conclusions.

CHAPTER 7

CONCLUSION

In this chapter I summarize my specific findings in the study. I then discuss three specific limitations I encountered and how these may be remedied. Next, I outline three ways that the current study could be expanded or used as a springboard for future research. Finally, I discuss my reasons for conducting the study and my hope for responses to it.

Summary of the Findings

This study has produced four important results. First, an empirical item hierarchy for the phonological features of sandhi-variation, elision and assimilation, was investigated, along with the determinants of the hierarchy. The Rasch item-person map for items with and without elision and with and without assimilation showed that there was a general tendency for items with elision and items with assimilation to have higher difficulty estimates. The results were consistent with previous explorations that also indicated sandhi-variation acts as a filter for second language comprehension (Ahn, 1987; Henrichsen, 1984). However, the determinants of the empirical hierarchy were unclear. The results of four dependent *t*-tests for items with and without elision and assimilation, and with and without two other variables, utterance rate (4.7 syllables per second and 3.17 syllables per second) and syllable length (8 syllables per utterance and 6 syllables per utterance), showed that items with elision were significantly more difficult than items without, that other comparisons (e.g., items with different syllable lengths) were not significant. For example, items with elision, 6 syllables, and a slower utterance rate (3.17 syllables per second) were more difficult than items without elision, but with 6 syllables, and a slower utterance rate (Figure 20). On the other hand, items with elision, 6 syllables, and a faster utterance rate (4.7 syllables per second) were not more difficult than items without elision, but with 6 syllables, and a faster utterance rate (Figure 22). Overall, the tendency was for items with the phonological features of sandhi-variation, elision and assimilation, to be more difficult. Confidence intervals for each of the item comparisons showed that the differences in scores were in the hypothesized direction, even when they did not reach statistical significance.

Second, the results indicated that the two input rate variables combined with the phonological features of sandhi-variation, elision and assimilation, affected the non-native participants' listening comprehension. Assumptions were first made about the results of the descriptive statistics for the phonological features of elision and assimilation from the Elicited Imitation Test (EIT) and the two utterance rates (4.7 syllables per second and 3.17 syllables per second). Twelve paired-sample *t*-tests were then used to investigate the combinations of the phonological features of elision and assimilation with the two utterance rates, and the results indicated that when elision and assimilation were removed from the items, and the utterance rates were slower, the participants' comprehension increased.

Third, the strength of the relationship between two measures of the participants' language ability, proficiency and aptitude, and their comprehension of items with and without the phonological features of elision and assimilation, were investigated. English proficiency, as measured by the TOEFL PBT, the TOEFL PBT listening section, and the

LVLT, was correlated with comprehension on the EIT of sandhi-variation. The Pearson correlations (Table 39) indicated correlations higher than .57, which indicated a relatively strong positive relationship between language proficiency and aural vocabulary ability with the comprehension of the phonological features of elision and assimilation. Next, the strength of the relationship between the participants' language aptitude, as represented by the Modern Language Aptitude Test–Elementary (MLAT-E), sections 1-4, and their comprehension on the EIT was investigated. The results (Table 40) showed that correlation coefficients computed among the four scales from the MLAT-E (i.e., Hidden Words, Matching Words, Finding Rhymes, and Number Learning), and the one scale for the EIT scores, elision and assimilation, were positive for 6 of the 10 correlations, with correlation coefficients greater than .73. Only the Number Learning scale correlations were not significant, with one negative correlation found between the Number Learning scores and the EIT scores (i.e., $r = -.04$). The results confirmed a positive relationship between language aptitude as measured by the MLAT-E and the comprehension of the phonological features of elision and assimilation.

Fourth, the results indicated that there were no significant, positive correlations between *English language proficiency* scores and both the Pre-Listening Questionnaire, which measured the participants' feelings about second language listening, and the Background and Length of Residency Interview. The *English language proficiency* scores were formulated from three interval scales—TOEFL PBT, the LVLT, and the EIT—consisting of Rasch person measures, which were then made into a single interval score. Moreover, the correlations among the English language proficiency scores and Background and Length of Residency Interview responses, which measured conversations with native English speakers, length of residency abroad, English education, and media use, produced no significant correlations and two negative correlations. More research needs to be conducted to determine how learners' backgrounds are related to listening comprehension in order to better prescribe aural input in second language listening classrooms.

Limitations

This study has three main limitations. First, the number of participants was small ($N = 25$). Increasing the number of participants would improve the clarity of the Rasch item measures by reducing the standard errors and would increase the statistical power of any significance tests. These improvements might provide a clearer measure of the hierarchy of the phonological features of sandhi-variation, elision and assimilation. In addition, many of the participants did not have experience traveling to or living in a country where English is the native language; thus, it was not possible to identify a relationship between time spent abroad and their scores on the EIT. Increasing the sample size and including more participants with experience abroad might uncover a relationship between the participants' backgrounds and experience processing English and their aural processing abilities.

The second limitation is that pinpointing the filter to comprehension in the targets where errors occurred might not have been completely accurate. Incorrect answers on the EIT were considered incorrect due to the filter effects embedded in half of the targets due to sandhi variation; however, the incorrect answers could have been due to a variable not tested in this study. For example, errors could have been caused by words that were “spuriously activated” (Cutler, 2015; Sebastian-Galles, Echeverria, & Bosch, 2005), which are words that are similar to the target but that are incorrectly activated by the listener. Moreover, the level of specificity used in experiments such as the one conducted by Kohler and Niebuhr (2011) of place assimilation in sibilant sequences, was not used in the current study. In order to investigate more specific effects of sandhi-variation the targets need to be controlled and balanced for specific features such as regressive assimilation and progressive assimilation.

The third limitation is that the participants were not challenged enough by the maximum number of syllables (8 syllables) and the faster utterance rate (4.7 syllables per second) in the study (e.g., Figure 12 and Figure 13). The decision for the maximum number of syllables and faster utterance rate was taken from literature on elicited imitation (e.g., Yan et al., 2015) and from piloted versions of the EIT used in this study. However, the Rasch difficulty estimates revealed hierarchal tendencies instead of clear hierarchies for items with and without the phonological features of elision and assimilation. Such enhancements of length and speed might shed more light on the process of second language listening comprehension for second language learners because they might uncover defined aural filters to the speech stream.

Finally, section 4, Number Learning, of the MLAT-E worked poorly, as all participants were able to correctly answer more than half of the items (Figure 8). Number Learning was designed to measure the memory component of language aptitude, but the instrument was too easy to estimate the abilities of the participants precisely.

Suggestions for Future Research

Three suggestions can be made for future research. First, future researchers can replicate this study with a larger number of participants, which would help to improve the differential clarity of the Rasch item measures by reducing the standard errors. A larger number of participants would also increase the statistical power of the *t*-tests. One benefit of this change might be a more precise measure of the hierarchy of the phonological features of sandhi-variation, elision and assimilation. Such a hierarchy could inform instructional interventions that develop students' ability to detect these features. In other words, instructional interventions might be better tuned to developing learners' ability to process the speech stream and develop greater automaticity and listening fluency. In addition, alternative types of background information, such as the dominant L1 accent of the area where a student lives abroad, or a more in-depth record of the types of L2 input a participant received, such as the types of media a participant interacted with, could be gathered to provide a more precise background information. Expanding into alternative types of background information might increase the possibility of finding a correlation between types of input and listening proficiency. Such a correlation might inform input that is more suitably comprehensible to learners based on the difficulty of the filter effects to comprehension by linking historical intake that differs from learner to learner (e.g., how a learner's background visiting Spain might give the learner an advantage with a target L2 regardless of their L1's mora-timing) at a more precise level.

Second, the aural input could be varied in numerous ways. For instance, other types of sandhi-variation (e.g., epenthesis) could be investigated to expand the understanding of how sandhi-variation filters second language listening comprehension.

Another variation of aural input would be to spuriously activate (Cutler, 2015; Sebastian-Galles et al., 2005) words within test items, by using items on the EIT with similarities that might be incorrectly activated by the listener. For example, one item might contain the phrase *two lips*, while another might contain the word *tulips*, thereby creating the possibility for a filter for aural comprehension. Furthermore, the level of specificity used in creating the items on the EIT in future studies could test the target filter (e.g., assimilation) in more detail. For example, Kohler and Niebuhr (2011) studied place assimilation in sibilant sequences, and such phonological specificity of the filters might produce results that are useful for understanding second language listening comprehension better by informing instructors and learners about which sandhi-variation types or examples should be noticed first. Furthermore, the number of syllables in the target items could be increased (e.g., > 8 syllables), and the speed of the target utterances could be increased (e.g., > 4.7 syllables per second) in order to better understand how higher proficiency listeners process aural input. Other modifications might include investigating the effect that vocal quality has on aural processing. For instance, the effects of female and male voices could be investigated. The benefit of expanding the input to include characteristics other than those used in the current study is that more accurate levels of textual aural input could be made from understanding how these characteristics filter second language aural input.

Third, introducing multiple types of interventions in which sandhi-variation targets are taught would give insight into what constitutes effective classroom instruction. Multiple types of sandhi-variation could be taught over one or two semesters, to investigate the effectiveness of instruction on the comprehension of the phonological effects of the instructional intervention. A specific example of an instructional intervention is to test the effect of shadowing on the comprehension of sandhi-variation in aural texts. Shadowing, which is a cognitive activity in which learners track the speech that they hear and vocalize it as clearly as possible while simultaneously listening (Tamai, 1997), is commonly used as a classroom technique to improve learners' processing of aural texts. It can be investigated to determine the degree to which it affects learners' abilities to process sandhi-variation over time. Finally, accurately gauging the processing ability of second language listening learners would provide further research directions, provide instructors with insight into the needs of such learners, and might inform classroom content related to aural texts used as input in listening lessons.

Final Conclusions

I undertook the present study to add new findings to the field of second language aural processing, especially second language listening. While theories abound within the field about what second language listening is, and how it should be taught, there is still a dearth of research at fundamental levels on how learners process aural language. For example, according to Hayes-Harb (2005), L2 listening proficiency is, basically, the development of "target-language-like perception of non-native sounds" (p. 1). However, without more empirical evidence to support or oppose such theories, and without knowing what the phonological features of those non-native sounds are, researchers cannot move the field forward. This study was designed to improve our understanding of the ways that learners process aural input; the results contributed to existing research on sandhi-variation as a filter to second language listening comprehension.

I hope that this study contributes to the literature regarding second language listening comprehension, especially regarding aural input, second language listening strategies (e.g., focusing learners on adjustments, such as elision and assimilation, made to authentic spoken language), and the hierarchical tendency of learner-intake for phonological filters embedded in spoken English. The results of this study indicate that processing two types of sandhi-variation, elision and assimilation, are interrelated with sentential parsing and therefore related to comprehension. When learners can parse a sentence accurately, they have a better chance at fully comprehending the sentence. To the best of my knowledge, this study has illuminated new factors that contribute to university learners of English as a second language. These new factors are (a) significant links between second language aural processing and both second language proficiency, as measured by TOEFL Listening, and second language aptitude, as measured by the MLAT-E, (b) insight into the filter effects that cause difficulty based on items with elision and assimilation, (c) strong support of the idea that filters to second language listening comprehension should be taken into consideration when an aural text is chosen as aural input for second language learners, and (d) strong support of the idea that aural input needs to be comprehensible so that learners have opportunities to notice the unique phonological aspects of the target language.

My wish is that other researchers continue to investigate phonological features of English and to explore ways that these features filter comprehension for second language learners. By investigating the filter effects of the phonological features of English, researchers can continue to assist learners of English and instructors of English by better understanding input.

REFERENCES

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh, Scotland: Edinburgh University Press.

Ahn, S. (1987). *Sandhi-variation and affective factors as input filters to comprehension of spoken English among Korean learners*. Seoul, Korea: Hanshin.

Allen, H. W., & Herron, C. (2003). A mixed-methodology investigation of the linguistic and affective outcomes of summer study abroad. *Foreign Language Annals 36*, 370-385. doi:10.1111/j.1944-9720.2010.01058.x

Anderson, A. & Lynch, T. (1988). *Listening*. Oxford, England: Oxford University Press.

Archila-Suerte, P., Zevin, J., Bunta, F., & Hernandez, A. E. (2012). Age of acquisition and proficiency in a second language independently influence the perception of non-native speech. *Bilingualism: Language and Cognition*, *15*(01), 190-201. doi:10.1017/S1366728911000125

Au, T. K., Knightly, L. M., Jun, S. A., & Oh, J. S. (2002). Overhearing a language during childhood. *Psychological Science*, *13*(3), 238-243. doi:10.1111/1467-9280.00444

Baese-Berk, M., Bradlow, A., & Wright, B. (2013). Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America, 133*(3), 174-180. doi:10.1121/1.4789864

Baker, R., & Bradlow, A. (2009). Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech, 52*(4), 391-413. doi:10.1177/0023830909336575

Bandura, A. (2001). Social cognitive theory: An Agentic perspective. *Annual Review of Psychology, 52*(1), 1-26. doi:10.1146/annurev.psych.52.1.1

Bazeley, P. (2004). Issues in mixing qualitative and quantitative approaches to research. In R. Buber, J. Gardner, & L. Richards (Eds.), *Applying qualitative methods to marketing management research* (pp. 141-156). Basingstoke, England: Palgrave Macmillan.

Beattie, J., Valls-Ferrer, M., & Pérez-Vidal, C. (2014). Listening performance and onset level in formal instruction and study abroad. In C. Pérez-Vidal (Ed.), *Language Acquisition in Study Abroad and Formal Instruction Contexts* (pp. 195-216). Amsterdam, The Netherlands: Benjamins. doi: 10.1075/aals.13.03ch2

Berkling, K., Zissman, M. A., Vonwiller, J., & Cleirigh, C. (1998). Improving accent identification through knowledge of English syllable structure. *ICSLP*, *98*, 89-92.

Best, C. & Tyler, M. (2007). Nonnative and second language speech perception: Commonalities and complementarities. In O. Bohn & M. Munro (Eds.), *Language experience in second language speech learning–In honor of James Emil Flege* (pp. 13-34). Amsterdam, The Netherlands: Benjamins. doi:10.1075/lllt.17.07bes

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London, England: Longman.

Bley-Vroman, R., & Chaudron, C. (1994). Elicited imitation as a measure of second-language competence. In E. Tarone, S. Gass, & A. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 245–262). Hillsdale, NJ: Erlbaum.

Bond, T., & Fox, C. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Erlbaum.

Bowen, J. D. (1975). *Patterns of English pronunciation*. Rowley, MA: Newbury House.

Bradlow, A., Nygaard, L., & Pisoni, D. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics*, *61*(2), 206-219. doi:10.3758/bf03206883

Briscoe, T. (Ed.). (2002). *Linguistic evolution through language acquisition*. New York, NY: Cambridge University Press. doi:10.1017/CBO9780511486524

Broersma, M. (2005). Perception of familiar contrasts in unfamiliar positions. *The Journal of the Acoustical Society of America*, *117*(6), 3890-3901. doi:10.1121/1.1906060

Broersma, M., & Cutler, A. (2011). Competition dynamics of second language listening. *Quarterly Journal of Experimental Psychology, 64*, 74-95. doi:10.1080/17470218.2010.499174

Brooks, N. (1966). Language learning: The new approach. *Phi Delta Kappan*, 357-359.

Brouwer, S., & Bradlow, A. (2015). The temporal dynamics of spoken word recognition in adverse listening conditions. *Journal of Psycholinguistic Research, 45*(5), 1151-1160. doi:10.1007/s10936-015-9396-9

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (Ed.), *Second language acquisition and linguistic theory*. (pp. 4-63). Malden, MA: Blackwell.

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

Brown, J. D., & Kondo-Brown, K. (2006). Introducing connected speech. In J. D. Brown, & K. Kondo-Brown (Eds.), *Perspectives on teaching connected speech to second language speakers* (pp. 1-15). National Foreign Language Resource Center, University of Hawai'i at Mānoa. doi:10.1017/S0025100308003472

Brunfaut, T., & Révész, A. (2015). The role of task and listener characteristics in second language listening. *TESOL Quarterly, 49*(1), 141-168. doi:10.1002/tesq.168

Buck, G. (1988). Testing listening comprehension in Japanese university entrance examinations. *JALT Journal*, *10*(1), 15-42.

Bundgaard-Nielsen, R., Best, C., Kroos, C., & Tyler, M. (2012). Second language learners' vocabulary expansion is associated with improved second language vowel intelligibility. *Applied Psycholinguistics, 33*(3), 643-664. doi:10.1017/S0142716411000518

Bundgaard-Nielsen, R., Best, C., & Tyler, M. (2011). Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics*, *32*, 51-67. doi:10.1017/S0142716410000287.

Burger, S., & Chretien, M. (2001). The development of oral production in content-based second language courses at the University of Ottawa. *Canadian Modern Language Review*, *58*(1), 84-102.

Bybee, J., Chakraborti, P., Jung, D., & Scheibman, J. (1998). Prosody and segmental effect: Some paths of evolution for word stress. *Studies in Language, 22*, 267-314. doi:10.1075/sl.22.2.02byb

Carroll, J. & Diller, K. (1981). *Twenty-five years of research in foreign language aptitude, individual differences and universals in language learning aptitude.* New York, NY: Newbury House.

Carroll, J. B., & Sapon, S. (2010). *Modern Language Aptitude Test – Elementary: Manual, 2010 Edition*. Rockville, MD: Second Language Testing.

Carroll, S., Swain, M., & Roberge, Y. (1992). The role of feedback in adult second language acquisition: Error correction and morphological generalizations. *Applied Psycholinguistics*, *13*(2), 173-198. doi:10.1017/S0142716400005555

Cawley, G. C., & Green, A. D. P. (1991, November). The application of neural networks to cognitive phonetic modelling. In *Artificial Neural Networks, 1991* (pp. 280-284). *Second International Conference on Artificial Neural Networks*. IET.

Ceci, S., & Papierno, P. (2005). The rhetoric and reality of gap closing: When the "have-nots" gain but the "haves" gain even more. *American Psychologist, 60*(2), 149-160. doi:10.1037/0003-066x.60.2.149

Celce-Murcia, M., Brinton, D. M., Goodwin, J. M., Griner, B. (2010/2016). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge, England: Cambridge University Press.

Chan, E., Skehan, P., & Gong, G. (2011). Working memory, phonemic coding ability and foreign language aptitude: Potential for construction of specific language aptitude tests–the case of Cantonese. *Ilha do Desterro, A Journal of English Language, Literatures in English and Cultural Studies, 60*, 45-73. doi:10.5007/2175-8026.2011n60p045

Chang, A. C., & Millett, S. (2014). The effect of extensive listening on developing L2 listening fluency: Some hard evidence. *ELT Journal 68*(1), 31-40. doi:10.1093/elt/cct052

Chaudron, C., Nguyen, H., & Prior, M. (2006). *Manual for the Vietnamese elicited imitation test*. Honolulu, HI: University of Hawai'i.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159. doi:10.1037/0033-2909.112.1.155

Collins, B., & Mees, I. (2013). *Practical phonetics and phonology: A resource book for students*. New York, NY: Routledge.

Cook, V. (Ed.) (2002). *Portraits of the L2 user*. Bristol, England: Multilingual Matters.

Coxhead, A. (2000). A new Academic word list. *TESOL Quarterly, 34*, 213-238. doi:10.2307/3587951

Crystal, D. (1997). *A dictionary of linguistics and phonetics* (4th ed.). Oxford, England: Blackwell.

Cumming, A., Grant, L., Mulcahy-Ernt, P., & Powers, D. E. (2004). A teacher-verification study of speaking and writing prototype tasks for a new TOEFL. *Language Testing*, *21*(2), 107-145. doi:10.1191/0265532204lt278oa

Cutler, A. (2002). Listening to a second language through the ears of a first. *Interpreting*, *5*(1), 1-23. doi:10.1075/intp.5.1.02cut

Cutler, A. (2015). *Native listening: Language experience and the recognition of spoken words.* Cambridge, MA: MIT Press.

Cutler, A., & Broersma, M. (2005). Phonetic precision in listening. In W. J. Hardcastle, & J. M. Beck (Eds.), *A figure of speech: A festschrift for John Laver* (pp. 63-91). Mahwah, NJ: Erlbaum.

Cutler, A., & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, *33*, 824-844. doi:10.1006/jmla.1994.1039

Darcy, I., Peperkamp, S., & Dupoux, E. (2007). Bilinguals play by the rules. Perceptual compensation for assimilation in late L2-learners. *Laboratory Phonology*, *9*, 411-442.

Díaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition*, *26*(02), 249-273. doi:10.1017/S0272263104262052

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2015). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience, 19*(1), 158-164. doi:10.1038/nn.4186

d'Inverno, M., & Luck, M. (2004). *Understanding agent systems*. New York, NY: Springer. doi:10.1007/978-3-662-10702-7

Doughty, C. (2013). Assessing aptitude. In A. Kunnan (Ed.) *The companion to language assessment* (pp. 23-46). Hoboken, NJ: John Wiley & Sons. doi:10.1002/9781118411360

Edwards, J. G. H., & Zampini, M. L. (2008). Phonology and second language acquisition. *Studies in Second Language Acquisition*, *33*(1), 127-128. doi:10.1075/sibil.36

Ellis, L., & Hardcastle, W. J. (2002). Categorical and gradient properties of assimilation in alveolar to velar sequences: Evidence from EPG and EMA data. *Journal of Phonetics*, *30*(3), 373-396. doi:10.1006/jpho.2001.0162

Ellis, N. C. (2002). Frequency effects in language acquisition: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*, 143-188. doi:10.1017/S0272263102002024

Ellis, N. C. (2008). Usage-based and form-focused language acquisition: The associative learning of constructions, learned-attention, and the limited L2 endstate. In P. Robinson & N. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 372-405). London, England: Routledge. doi:10.1111/j.1540-4781.2009.00946.x

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics, 27*(3), 464-491. doi:10.1093/applin/aml001

Fabra, L. R. (2005). Predicting ease of acquisition of L2 speech sounds: A perceived dissimilarity test. *Vigo International Journal of Applied Linguistics, 2*, 75-92.

Fay, A., & Buchweitz, A. (2014). Listening comprehension and individual differences in working memory capacity in beginning L2 learners. *Letrônica*, *7*(1), 113. doi:10.15448/1984-4301.2014.1.16839

Field, A. (2013). *Discovering statistics using SPSS*. London, England: Sage.

Field, J. (2003). Promoting perception: Lexical segmentation in L2 listening. *ELT Journal*, *57*(4), 325-334. doi:10.1093/elt/57.4.325

Field, J. (2008a). *Listening in the language classroom.* Cambridge, England: Cambridge University Press.

Field, J. (2008b). Revising segmentation hypotheses in first and second language listening. *System, 36*, 35-51. doi:10.1016/j.system.2007.10.003

Flege, J., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition, 23*(4), 527-552.

Flowerdew, J., & Miller, L. (2005). *Second language listening: Theory and practice*. Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511667244

Fox, C. M., & Jones, J. A. (1998). Use of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology, 45*(1), 30-45. doi: 10.1037/0022-0167.45.1.30

Fox, R., & McGory, J. (2007). Second language acquisition of a regional dialect of American English by native Japanese speakers. In O. Bohn & M. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp.117-134). Amsterdam, The Netherlands: Benjamins. doi:10.1075/lllt.17

Frisch, S., Large, N., & Pisoni, D. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language, 42*(4), 481-496. doi:10.1006/jmla.1999.2692

Fujio, M. (2007). *Communication strategies for the negotiation, establishment, and confirmation of common ground: A longitudinal study of Japanese-British conversational interaction.* Unpublished doctoral thesis. Tokyo, Japan: The University of Tokyo.

Fujio, M. (2010). The role of strategic competence in listening comprehension. *Bulletin of the Foreign Language Teaching Association*, *9*(1), 1-21.

Gaillard, S., & Tremblay, A. (2016). Oral proficiency assessment in second language acquisition research: The elicited imitation task. *Language Learning, 66*, 419-447. doi: 10.1111/lang.12157

Gass, S. M., & Mackey, A. (2006). Input, interaction and output: An overview. *AILA Review*, *19*(1), 3-17. doi:10.1075/aila.19.03gas

Goh, C. C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, *28*(1), 55-75. doi:10.1016/S0346-251X(99)00060-3

Graham, C. R., McGhee, J., Millard, B. (2010). The role of lexical choice in elicited imitation item difficulty. In M. Prior, Y. Watanabe, & S. Lee (Eds.), *Selected proceedings of the 2008 second language research forum* (pp. 57-72). Somerville, MA: Cascadilla Proceedings Project.

Granena, G, & Long, M. (2013). Age of onset, length of residence, aptitude and ultimate L2 attainment in three linguistic domains. *Second Language Research , 29*, 311-43.

Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics*, *25*(5), 408-422. doi:10.1016/j.jneuroling.2009.08.006

Hammer, C. S., Jia, G., & Uchikoshi, Y. (2011). Language and literacy development of dual language learners growing up in the United States: A call for research. *Child Development Perspectives, 5*(1), 4-9. doi:10.1111/j.1750-8606.2010.00140.x

Hayes-Harb, R. (2005). Optimal L2 speech perception: Native speakers of English and Japanese consonant length contrasts. *Journal of Language & Linguistics, 4*(1), 1-29.

Henrichsen, L. E. (1984). Sandhi-variation: A filter of input for learners of ESL. *Language Learning*, *34(*3), 103-123. doi:10.1111/j.1467-1770.1984.tb00343.x

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393-402. doi:10.1038/nrn2113

Hieke, A. E. (1990). Toward listener strategies for decoding fluent speech. *International Review of Applied Linguistics in Language Teaching*, *28*(3), 221-233.

Husserl, E. (1991). *On the phenomenology of the consciousness of internal time (1893-1917)* (Vol. 4). New York, NY: Springer Science & Business Media. doi:10.1007/978-94-011-3718-8

Jackendoff, R. (2002). Précis of foundations of language: Brain, meaning, grammar, evolution. *Behavioral and Brain Sciences*, *26*(6), 651-665. doi:10.1017/S0140525X03000153

Jarvella, R. J. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior*, *10*(4), 409-416. doi:10.1016/S0022-5371(71)80040-3

Jensen, E. D., & Vinther, T. (2003). Exact repetition as input enhancement in second language acquisition. *Language Learning*, *53*(3), 373-428. doi:10.1111/1467-9922.00230

Jones, D. (1956). *The pronunciation of English*. Cambridge, England: Cambridge University Press.

Juffs, A., & Harrington, M. (1996). Garden path sentences and error data in second language sentence processing. *Language Learning*, *46*(2), 283-323. doi:10.1111/j.1467-1770.1996.tb01237.x

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching, 44*(2), 137-166. doi:10.1017/S0261444810000509

Jusczyk, P. W. (2000). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Kaisse, E. M. (1985). *Connected speech: The interaction of syntax and phonology*. Orlando, FL: Academic.

Keating, P. A. (1990). The window model of coarticulation: articulatory evidence. *Papers in Laboratory Phonology I*, *26*, 451-470. doi:10.1017/CBO9780511627736.026

Kennedy, S., & Blanchet, J. (2014). Language awareness and perception of connected speech in a second language, *Language Awareness, 23*(1-2), 92-106, doi:10.1080/09658416.2013.863904.

Kinginger, C. (2008). Language learning in study abroad: Case studies of Americans in France. *The Modern Language Journal, 92*, 1-124. doi:10.1111/j.1540-4781.2008.00821.x

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, *7*(312), 1-26.

Kohler, K. J., & Niebuhr, O. (2011). On the role of articulatory prosodies in German message decoding. *Phonetica*, *68*(1-2), 57-87. doi:10.1159/000328656

Kösem, A., Basirat, A., Azizi, L., & van Wassenhove, V. (2016). High-frequency neural activity predicts word parsing in ambiguous speech streams. *Journal of Neurophysiology, 116*(6), 2497-2512. doi:10.1152/jn.00074.2016

Kostin, I. (2003). Exploring item characteristics that are related to the difficulty of TOEFL dialogue items. *ETS Research Report Series*, *2004*(1), i-59. doi:10.1002/j.2333-8504.2004.tb01938.x

Krashen, S. D. (1981). The "fundamental pedagogical principle" in second language teaching. *Studia Linguistica*, *35*(1-2), 50-70. doi:10.1111/j.1467-9582.1981.tb00701.x

Kröger, B. J., Kannampuzha, J., & Neuschaefer-Rube, C. (2009). Towards a neurocomputational model of speech production and perception. *Speech Communication*, *51*(9), 793-809. doi:10.1016/j.specom.2008.08.002

Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*(15), 9096-9101. doi:10.1073/pnas.1532872100

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), 13-21. doi:10.1111/j.1467-7687.2006.00468.x

Kühnert, B., & Nolan, F. (1999). The origin of coarticulation. In W. J. Hardcastle, & N. Hewlett (Eds.), *Coarticulation: Theory, data and techniques* (pp. 1-30). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511486395.002

Ladefoged, P., & Johnson, K. (2014). *A course in phonetics*. Boston, MA: Cengage Learning.

Laver, J. (1994). *Principles of phonetics*. Cambridge, England: Cambridge University Press. doi:10.1017/CBO9781139166621

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition, 38*(4), 801-842. doi:10.1017/S027226311500042X

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431-461. doi:10.1037/h0020279

Liberman, A. M., & Mattingly, I. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1-36. doi:10.1016/0010-0277(85)90021-6

Lightbown, P. M., & Spada, N. (1990). Focus-on-form and corrective feedback in communicative language teaching. *Studies in Second Language Acquisition*, *12*(4), 429-448. doi:10.1017/S0272263100009517

Linacre, J. M. (2016a). Winsteps® (Version 3.92.0) [Computer Software]. Beaverton, OR: Winsteps.com. Retrieved January 1, 2016.

Linacre, J. M. (2016b). *Winsteps® Rasch measurement computer program User's Guide*. Beaverton, OR: Winsteps.com

Mackey, A., & Gass, S. (2005). *Second language research: Methodology and design*. Mahwah, NJ: Erlbaum.

Mai, L. H., Ngoc, L. T. B., & Thao, V. T. (2014). Enhancing listening performance through schema construction activities. *Journal of Language Teaching and Research*, *5*(5). doi:10.4304/jltr.5.5.1042-1051

Major, R. C. (2001). *Foreign accent: The ontogeny and phylogeny of second language phonology*. London, England: Erlbaum.

Mathews, J., O'Toole, J., and Chen, S. (2017). The impact of word recognition from speech (WRS) proficiency level on interaction, task success and word learning: Design implications for CALL to develop L2 WRS. *Computer Assisted Language Learning, 30*, 22-43. doi: 10.1080/09588221.2015.1129348

McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research,19*(6), 741-760. doi:10.1177/1362168814567889

Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics*, *65*(6), 956-969. doi:10.3758/bf03194826

Mixdorff, H., & Munro, M.J. (2013). Quantifying and evaluating the impact of prosodic differences of foreign-accented English. *Proceedings of Slate, 2013*. Grenoble, France.

Moyer, A. (2011). An investigation of experience in L2 phonology: Does quality matter more than quantity?. *Canadian Modern Language Review, 67*, 191-216. doi:10.3138/cmlr.67.2.191.

Munnich, E., Flynn, S., & Martoharjono, G. (1994). Elicited imitation and grammaticality judgment tasks: What they measure and how they relate to each other. In E. Tarone, S. Gass, & A. Cohen (Eds.) *Research methodology in second language acquisition* (pp. 227-245). Mahwah, NJ: Erlbaum.

Munro, M., & Bohn, O. (2007). The study of second language speech learning: A brief overview. In O. Bohn & M. Munro (Eds.), *Language experience in second* *language speech learning–In honor of James Emil Flege* (p. 3). Amsterdam, The Netherlands: Benjamins. doi:10.1075/lllt.17.07bes

Nishino, T. (1992). What influences success in listening comprehension? *Language Laboratory*, *29*, 37-52.

Nunan, D. (2002). Listening in language learning. In J. C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching* (pp. 238-241). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511667190.032

Oka, H., Yamauchi, Y., & Fujio, M. (2005). How do strategies work? Investigating practical English ability. *Bulletin of the Foreign Language Teaching Association*, *9*, 1-21.

Perkins, K., Brutten, S., & Angelis, P. (1986). Derivational complexity and item difficulty in a sentence repetition task. *Language Learning, 36*, 125-141. doi:10.1111/j.1467-1770.1986.tb00375.x

Pezzulo, G. (2008). Coordinating with the future: The anticipatory nature of representation. *Minds and Machines*, *18*, 179-225. doi:10.1007/s11023-008-9095-5

Pinker, S. (2007). *The language instinct (1994/2007)*. New York, NY: Harper Perennial Modern Classics.

Piske, T., MacKay, I. & Flege, J. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics, 29*(2), 191-215. doi:10.1006/jpho.2001.0134

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878-912. doi:10.1111/lang.12079

Porro, C. A., Baraldi, P., Pagnoni, G., Serafini, M., Facchin, P., Maieron, M., & Nichelli, P. (2002). Does anticipation of pain affect cortical nociceptive systems?. *The Journal of Neuroscience*, *22*(8), 3206-3214. doi:10.1016/S1053-8119(00)90940-6

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen, Denmark: Danmarks Paedagogiske Institut. doi:10.4135/9781412961288.n335

Rebuschat, P., & Mackey, A. (2013). Prompted production. In C. A. Chappelle (Ed.), *The encyclopedia of applied linguistics*, *5* (pp. 1-5). Oxford, England: Wiley Blackwell. doi:10.1002/9781405198431.wbeal0964

Renandya, W., & Farrell, T. (2010). ‘‘Teacher, the tape is too fast!’’ Extensive listening in ELT. *ELT Journal, 65*(1), 52-59. doi:10.1093/elt/ccq015

Révész, A., & Brunfaut, T. (2013). Text characteristics of task input and difficulty in second language listening comprehension. *Studies in Second Language Acquisition*, *35*(1), 31-65. doi:10.1017/S0272263112000678

Roach, P. (2009). *English phonetics and phonology: A practical course* (4th ed.). Cambridge, England: Cambridge University Press.

Rosen, R. (1985). *Anticipatory systems*. Oxford, England: Pergamon Press. doi:10.1016/B978-0-08-031158-6.50006-X

Rosen, R. (1987). On complex systems. *European Journal of Operational Research*, *30*(2), 129-134. doi:10.1016/0377-2217(87)90089-0

Rost, M. (2016). *Teaching and researching listening*. London, England: Longman. doi:10.4324/9781315833705

Rost, M., & Ross, S. (1991). Learner use of strategies in interaction: Typology and teachability. *Language Learning, 41*, 235-273. doi:10.1111/j.1467-1770.1991.tb00685.x

Rubin, J. (1994). A review of second language listening comprehension research. *Modern Language Journal*, *78*, 199-221. doi:10.1111/j.1540-4781.1994.tb02034.x

Saito, K. (in press). Advanced segmental and suprasegmental acquisition. In P. Malovrh & A. Benati (Eds.). *The handbook of advanced proficiency in second language acquisition.* Hoboken, NJ: Wiley Blackwell.

Saito, K. (2015). Experience effects on the development of late second language learners' oral proficiency. *Language Learning, 65*, 563-595. doi:10.1111/lang.12120

Saito, K., Trofimovich, P., & Isaacs, T. (2015). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, *35*, 217-240. doi:10.1017/S0142716414000502

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, *71*(1), 145-163. doi:10.1016/j.jml.2013.11.002

Schrooth, T., Smith, B., & Kyles, E. (2015). Language learning background and attitudes in ESL contexts: The case of verb acquisition. *The Journal of Language Teaching and Learning, 2*, 74-85.

Sebastian-Galles, N., Echeverria, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language, 52*, 240-255. doi:10.1016/j.jml.2004.11.001

Segalowitz, N., Gatbonton, E., & Trofimovich, P. (2009). Links between ethnolinguistic affiliation, self-related motivation and second language fluency: Are they mediated by psycholinguistic variables?. In Z. Dornyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 172-192). Bristol, England: Multilingual Matters.

Simon, E., & D'Hulster, T. (2012). The effect of experience on the acquisition of a non-native vowel contrast. *Language Sciences, 34*(3), 269-283. doi:10.1016/j.langsci.2011.10.002

Skehan, P. (2002). Theorizing and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69-93). Amsterdam, The Netherlands: Benjamins. doi:10.1075/lllt.2

Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183-205). Cambridge, England: Cambridge University. doi:10.1017/CBO9781139524780.009

Sobkowiak, W. (2001). *English phonetics for Poles: A resource book for learners and teachers*. (pp. 121-135). Poznań, Poland: Poznańskie.

Spada, N., Shiu, J., & Tomita, Y. (2015). Validating an elicited imitation task as a measure of implicit knowledge: Comparisons with other validation studies. *Language Learning, 65*, 723-751. doi:10.1111/lang.12129

Spinelli, E., & Gros-Balthazard, F. (2007). Phonotactic constraints help to overcome effects of schwa deletion in French. *Cognition, 104*(2), 397-406. doi:10.1016/j.cognition.2006.07.002

Suzuki, Y. & DeKeyser, R. (2015). Comparing elicited imitation and word monitoring as measures of implicit knowledge. *Language Learning, 65*, 860-895. doi:10.1111/lang.12138

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. Boston, MA: Pearson Education.

Tamai, K. (1997). Shadowing no koka to chokai process ni okeru ichizuke [The effectiveness of shadowing and its position in the listening process]. *Current English Studies*, *36*, 105-116.

Tatham, M., & Morton, K. (2006). *Speech production and perception*. New York, NY: Palgrave MacMillan. doi:10.1057/9780230513969

Tauroza, S., & Allison, D. (1990). Speech rates in British English. *Applied Linguistics, 11*, 90-105. doi:10.1093/applin/11.1.90

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Toates, F. (2005). Evolutionary psychology–Towards a more integrative model. *Biology and Philosophy*, *20*(2-3), 305-328. doi:10.1007/s10539-004-0756-3

Tracy-Ventura, N., McManus, K., Norris, J., & Ortega, L. (2014). "Repeat as much as you can": Elicited imitation as a measure of oral proficiency in L2 French. In P. Leclercq, H. Hilton, & A. Edmonds (Eds.), *Measuring L2 proficiency: Perspectives from SLA* (pp. 143-166). Bristol, England: Multilingual Matters.

Trochim, W. M. (2008). *Research methods knowledge base*. Cincinnati, OH: Atomic Dog.

Trofimovich, P. (2011). Language experience in L2 phonological learning: Effects of psycholinguistic and sociolinguistic variables. *International Review of Applied Linguistics, 49*, 135-156. doi:10.1515/iral.2011.007

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28*(01), 1-30. doi:10.1017/s0272263106060013

Trueswell, J., Tanenhaus, M., & Garnsey, S. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33*(3), 285-318. doi:10.1006/jmla.1994.1014

Tuinman, A., & Cutler, A. (2011). L1 knowledge and the perception of casual speech processes in L2. In M. Wrembel, M. Kul, & K. Dziubalska-Kolaczyk (Eds.), *Achievements and perspectives in SLA of speech: New sounds 2010. Volume I* (pp. 289-301). Frankfurt am Main, Germany: Peter Lang.

Underhill, N. (1987). *Testing spoken language: A handbook of oral testing Techniques.* Cambridge, England: Cambridge University Press.

Vandergrift, L., & Goh, C. (2012). *Teaching and learning second language listening: Metacognition in action*. New York, NY: Routledge.

Vitevitch, M., Luce, P., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech, 40*(1), 47-62. doi:10.1177/002383099704000103

Warner, N., Otake, T., & Arai, T. (2010). Intonational structure as a word-boundary cue in Tokyo Japanese. *Language and Speech*, *53*(1), 107-131. doi:10.1177/0023830909351235

Webb, S. A., & Sasao, Y. (2013). New directions in vocabulary testing. *RELC Journal*, *44*(3), 263-277. doi:10.1177/0033688213500582

Weber, A. (2001). Help or hindrance: How violation of different assimilation rules affects spoken-language processing. *Language and Speech*, *44*(1), 95-118. doi:10.1177/00238309010440010401

Wells, J. C. (1982). *Accents of English* (Vol. 1). Cambridge, England: Cambridge University Press. doi:10.1177/007542428301600112

Wojtaszek, A., & Arabski, J. (2011). *The acquisition of L2 phonology*. Toronto, Canada: Multilingual Matters. doi:10.1111/j.1540-4781.2013.01438.x

Wolvin, A. D. (2005). *Listening engagement: Intersecting theoretical perspectives*. Cambridge, England: Cambridge University Press. doi:10.1002/9781444314908.ch1

Wong, S., Mok, P., Chung, K., Leung, V., Bishop, D., & Chow, B. (2015). Perception of native English reduced forms in Chinese learners: Its role in listening comprehension and it phonological correlates. *TESOL Quarterly, 51*, 7-31.

Wright, B. A., Baese-Berk, M. M., Marrone, N., & Bradlow, A. R. (2015). Enhancing speech learning by combining task practice with periods of stimulus exposure without practice. *The Journal of the Acoustical Society of America*, *138*(2), 928-937. doi:10.1121/1.4927411

Wright, B. A., & Sabin, A. T. (2007). Perceptual learning: how much daily training is enough? *Experimental Brain Research*, *180*(4), 727-736. doi:10.1007/s00221-007-0898-z

Wu, S., & Ortega, L. (2013). Measuring global oral proficiency in SLA research: A new elicited imitation test of L2 Chinese. *Foreign Language Annals, 46*, 680-704. doi:10.1111/flan.12063

Yan, X., Maeda, Y., Lv, J., & Ginther, A. (2015). Elicited imitation as a measure of second language proficiency: A narrative review and meta-analysis. *Language Testing*, *33*(4), 497-528. doi:10.1177/0265532215594643

Yu, A. (2013). Individual differences in socio-cognitive processing and the actuation of sound change. In A. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (pp. 201-227). Oxford, England: Oxford University Press.

APPENDICES

APPENDIX A

THE LISTENING VOCABULARY LEVELS TEST (TRANSLATION [FROM JAPANESE])

This is a vocabulary test.
Each English word will be read together with an example sentence. Select the Japanese word from the choices (a~d) that has the closest meaning to the English word being read.

Each question will be read only once.

Example Problem

1.
a. 食(た)べた (ate)
b. 待(ま)った (waited)
c. 見(み)た (saw)
d. 寝(ね)た (slept)

(Recorded Material)

waited: I waited for a bus.

The correct answer is **b**.

If you do not know a word at all, please leave it blank.
However, if you think there is a chance that you may know the word, please try to answer.

Let's practice some problems.

Practice Problem 1
a. 強(つよ)い (strong)
b. 幸(しあわ)せな (happy)
c. 食(た)べすぎる (eats too much)
d. 親切(しんせつ) (kind)

Practice Problem 2
a. 〜について話(はな)します (talk about)
b. 〜を運(はこ)ぶ (carry)
c. ~に名前(まえ)を書(か)く (write your name on)
d. ~を振(ふ)る (shake)

Because this is a listening test, please do not speak until it is finished.

Let's begin.

1. time: They have a lot of <time>.
a. money
b. food
c. **time**
d. friends

2. stone: She sat on a <stone>.
a. **stone**
b. bench
c. rug
d. twig

3. poor: We are <poor>.
a. **poor**
b. feel happy
c. are very interested
d. tall

4. drive: She <drives> fast.
a. swims
b. learns
c. throw balls
d. **drives**

5. jump: She tried to <jump>.
a. float
b. **jump**
c. park
d. run

6. shoe: Where is your other <shoe>.
a. parent
b. wallet
c. pen
d. **shoes**

7. test: We have a <test> in the morning.
a. meeting
b. trip
c. **test**
d. plan

8. nothing: He said <nothing> to me.
a. very bad things
b. **nothing**
c. very good things
d. something

9. cross: Don't <cross>.
a. **cross**
b. push
c. eat too fast
d. wait

10. actual: The <actual> one is larger.
a. **actual**
b. old
c. round
d. other

11. any: Does she have <any> friends?
a. **any**
b. no
c. good
d. old

12. far: You have walked <far>!
a. for a long time
b. very fast
c. **far**
d. to your house

13. game: I like this <game>.
a. food
b. story
c. group of people
d. way of playing

14. cause: He <caused> the problem.
a. **caused**
b. fixed
c. explained
d. understood

15. many: I have <many>.
a. none
b. enough
c. a few
d. **many**

16. where: <Where> did you go?
a. at what time
b. for what reason
c. **where**
d. in what way

17. school: This is a big <school>.
a. bank
b. sea animal
c. **school**
d. house

18. grow: All the children <grew>.
a. drew pictures
b. spoke
c. **grew**
d. cried a lot

19. flower: He gave me a <flower>.
a. pajamas
b. small clock
c. **flower**
d. piece of bread / bread

20. handle: I can't <handle> it.
a. open
b. remember
c. **handle**
d. believe

21. camp: He is in the <camp>.
a. sea
b. **camp**
c. **hospital**
d. hotel

22. lake: People like the <lake>.
a. **lake**
b. very young child
c. leader
d. quiet place

23. past: It happened in the <past>.
a. **past**
b. shocking event
c. night
d. summer

24. round: It is <round>.
a. friendly
b. very big
c. very quick
d. **round**

1. maintain: Can they <maintain> it?
a. **maintain**
b. magnify
c. improve
d. obtain

2. period: It was a difficult <period>.
a. question
b. **period**
c. thing to do
d. book

3. standard: Her <standards> are very high.
a. heels
b. grades
c. prices
d. **standards**

4. basis: This was used as the <basis>.
a. answer
b. resting place
c. next step
d. **basis**

5. upset: I am <upset>.
a. strong
b. famous
c. rich
d. **upset**

6. drawer: The <drawer> was empty.
a. **drawer**
b. garage
c. refrigerator
d. cage

7. pub: They went to the <pub>.
a. **pub**
b. bank
c. shopping area
d. swimming pool

8. circle: Make a <circle>.
a. rough draft
b. blank space
c. **circle**
d. large hole

9. pro: He's a <pro>.
a. a spy
b. a stupid person
c. a writer
d. **a pro**

10. soldier: He is a <soldier>.
a. a businessman
b. a student
c. a carpenter
d. **a soldier**

11. result: They were waiting for the <results>.
a. the right time
b. questions
c. money
d. **results**

12. resist: They <resisted> it.
a. repaired
b. looked at it twice
c. thought hard about
d. **resisted**

13. lend: She often <lends> her books.
a. **lends**
b. scribbles
c. cleans
d. writes her name

14. refuse: She <refused>.
a. returned
b. considered
c. **refused**
d. stayed late

15. speech: I enjoyed the <speech>.
a. **speech**
b. sprint
c. music
d. food

16. pressure: They used too much <pressure>.
a. money
b. time
c. **pressure**
d. bad words

17. refer: She <referred> to him.
a. supported him
b. let him go first
c. **referred to**
d. answered him

18. army: They saw the <army>.
a. black and white animal
b. bookshelf
c. neighbor
d. **army**

19. knee: Take care of your <knee>.
a. small child
b. **knee**
c. money
d. possession

20. rope: He found a <rope>.
a. **rope**
b. tool for making holes
c. wallet
d. ladder

21. brand: This is a good <brand>.
a. dance party
b. first attempt
c. waiting room
d. **brand**

22. seal: They <sealed> it.
a. fixed it
b. seal it
c. looked at it carefully
d. opened it

23. warn: They were <warned>.
a. pushed away
b. invited in
c. **warned**
d. led into a war

24. reserve: They have large <reserves>.
a. **reserves**
b. oven
c. debt
d. employees

1. restore: It has been <restored>.
a. repeated
b. transferred
c. reduced in price
d. **restored**

2. compound: They made a new <compound>.
a. agreement
b. **compound**
c. company
d. prediction

3. latter: I agree with the <latter>.
a. pastor
b. reason
c. **latter** d.
answer

4. pave: It was <paved>.
a. blocked
b. shared between groups
c. given gold borders
d. **paved**

5. remedy: We found a good <remedy>.
a. **remedy**
b. restaurant
c. method of cooking
d. equation

6. bacterium: They didn't find a single <bacterium>.
a. **bacterium**
b. plant with red or orange flowers
c. camel
d. stolen item

7. behavior: Look at her <behavior>!
a. audience
b. **behavior**
c. large amount of money
d. island

8. fuel: Do you have any <fuel>?
a. **fuel**
b. painkillers
c. cloth
d. insulation

9. silk: It's made of <silk>.
a. **silk**
b. hard black wood
c. fur
d. a shiny metal

10. conceive: Who <conceived> the idea?
a. told to others
b. explained
c. **conceived**
d. criticized

11. legend: It is now a <legend>.
a. museum
b. habit
c. **legend**
d. periodic event

12. impose: This was <imposed>.
a. completely changed
b. in the middle of other things
c. made to look like something else
d. **impose**

13. solution: There is no <solution>.
a. time
b. support
c. problem
d. **solution**

14. celebrate: We have <celebrated> a lot recently.
a. discovered
b. inspected
c. worked hard
d. **celebrated**

15. independence: He has too much <independence>.
a. **independence**
b. solitude
c. power
d. pride

16. tunnel: We need a <tunnel> here.
a. **tunnel**
b. pole
c. hyphen
d. curtain

17. reward: He got a good <reward>.
a. praise
b. help around the house
c. **reward**
d. audience

18. review: The committee <reviewed> the plan.
a. **reviewed**
b. accepted
c. reproduced
d. discarded

19. mode: The <mode> of production has changed.
a. **mode**
b. speed
c. attitude
d. amount

20. personnel: I don't like the <personnel> there.
a. chairs
b. air quality
c. **personnel**
d. employer

21. competent: She was very <competent>.
a. efficient
b. angry
c. **competent**
d. easily hurt

22. devastate: The city was <devastated>.
a. decorated
b. isolated
c. **devastated**
d. polluted

23. constituent: This is an important <constituent>.
a. building
b. agreement
c. idea
d. **constituent**

24. weave: She knows how to <weave>.
a. **weave**
b. join pieces of metal together
c. persuade people
d. trick people

APPENDIX B

PRE-LISTENING IN ENGLISH QUESTIONNAIRE

Answer the following questions in Japanese.

日本語で以下の問いに答えよ。

1. How do you feel about English, in general?

一般的に、英語についてどう思いますか？

2. How do you feel when you listen to English?

あなたが英語を聞くとき、どのように感じていますか？

3. How do you feel about the English sound system (e.g., pronunciation)?

英語のサウンドシステム（例えば、発音, イントネーション等）についてどう思いますか？

4. What is your history of studying English? (How many years have you been studying English? Have you studies abroad in any way? Do you listen to English songs? Etc.)

これまでどのように英語を勉強してきましたか（何年間英語を勉強してきましたか？何らかの理由で英語で授業が行われる海外の学校で勉強したことがありますか？あなたは英語の歌を聴いていますか？など）

APPENDIX C

ELICITED IMITATION TEST

Elicited Imitation Test
口頭模倣テスト

This is an elicited imitation test.
これは口頭模倣（聞こえてきた語句や文を復唱し、次に完全な形に直してもう一度言葉に発する）テストです。

You will hear phrases or complete sentences.
初めに語句や文が話されます。

First, repeat each phrase or sentence exactly as you hear them.
まず、聞こえてきたフレーズや文を復唱してください。

Second, say the phrase or sentence a second time, pronouncing each word in its complete form.
次にもう一度その語句や文を言ってください。その際、あなたが聞いた語句や文が短縮系であれば完全な形に直して言ってください。

Your responses will be recorded.
あなたの回答は録音されます。

Examples:

You hear: I don’t like milk.
You repeat: I don’t like milk.
You then say: I *do not* like milk.

You hear: Watcha doing?
You repeat: Watcha doing?
You then say: *What are you* doing?

You hear: Are ya happy?
You repeat: Are ya happy?
You then say: Are *you* happy?

You hear: Wouldjou like a drink?
You repeat: Wouldjou like a drink?
You then say: Would you like a drink?

You hear: It’s a pinkcar.
You repeat: It’s a pinkcar.
You then say: It is a pink car.

[*Students then hear audio versions of the examples above. After the examples are heard, the interviewer checks with the participant to verify understanding of the test. Then the test begins with the following items.*]

Test Items:

1. 1-E@8–S Coffee is a waste of money.

   weɪst ʌv

2. 1+A@6+S He disappeared in snow.

   ɪn snoʊ

3. 1+A@8–S My mom used to like chocolate.

   laɪg ʧɔklət

4. 1-E@8+S I have never seen your family.

   fæməli

5. 1+E@6+S He hosts several games.

   hoʊs sɛvrəl

6. 1+E@6–S Send them to me by mail.

   sɛnd ɛm

7. 1+A@8+S The best estimate is secret.

   bɛs ɛstəmət

8. 1+E@6–S East side of the city.

   is saɪd

9. 1-E@6+S Tell him the drink you want.

   tɛl hɪm

10. 1+E@8–S Somehow, we woke up very late.

    sʌmhaʊ i

11. 1-A@8–S I did not recognize the spot.

    rɛkəgnaɪz ðə

12. 1+A@6+S Her daughter writes quite well.

    kwaɪt⁼ wɛl

13. 1+A@8–S Did your father used to work there?

    wɜrg ðɛr

14. 1+A@8+S There are ten cars in the car shop.

    tɛŋ kɑrz

15. 1-E@6–S The river runs <u>that way</u>.

ðæt weɪ

16. 1+E@8+S Your <u>joint effort</u> to win paid off.

dʒɔɪn ɛfərt

17. 1-A@8–S We <u>have to</u> buy a new car soon.

hæv tu

18. 1+E@6+S That video <u>shouldn't</u> be watched.

ʃʊdənt

19. 1-E@6+S He'll <u>be about</u> an hour.

bi əbaʊt

20. 1-A@8+S He's surprised <u>about the</u> pressure.

əbaʊt ðə

21. 1-E@6–S Concert <u>tickets are</u> cheap.

tɪkəts ɑr

22. 1-A@6+S <u>What do</u> you think is funny?

wʌt du

23. 1+A@6–S Can't you <u>meet me</u> tonight?

mit⁼ mi

24. 1+A@6–S What do you <u>think about</u>?

θɪŋg əbaʊt

25. 1+E@8+S My <u>employer ran</u> in the race.

ɛmplɔɪər æn

26. 1+A@8+S Satisfy your <u>dreams with</u> software.

drimz wɪð

27. 1-E@8–S Does your family <u>live here</u> still?

lɪv hir

28. 1-A@6–S I have a <u>good neighbor</u>.

gʊd neɪbər

29. 1-E@8–S She has <u>lived there</u> about a year.

laɪvd ðɛr

30. 1+E@6+S We are <u>grateful for</u> it.

greɪtfəɫ fɔr

31. 1-E@6–S It’s useful to clean it.
jusfəl tu
32. 1-A@8+S I don’t know what class I should take.
ʃʊd teɪk
33. 1+A@8–S She did not like his shirt very much.
nɑt˭ laɪk
34. 1-A@8+S Does its shape change all of the time?
ɪts ʃeɪp
35. 1-A@6–S What are we doing here?
duɪŋ hir
36. 1-A@6+S When did you go to the park?
dɪd ju
37. 1+E@8–S He will be about an hour late.
bi baʊt
38. 1+A@6–S I have to go to school.
hæf tu
39. 1+E@8+S It was attached to the panel.
ətæʧ tu
40. 1-A@6+S Our area is here.
aʊər ɛriə
41. 1+A@6+S We breathe in air deeply.
brið ɪn
42. 1-E@8+S Who is your regular doctor?
hu ɪz
43. 1+E@8–S Will the owners of the coat be found?
wɪł ði
44. 1-E@8+S The king got up to make a speech.
kɪŋ gɑt
45. 1-A@6–S It is painted bright blue.
braɪt blu
46. 1-E@6+S They will get them in time.
ðeɪ wɪl

47. 1+E@6–S <u>Best side</u> is on the left.

bɛs saɪd

48. 1-A@8–S The actual labor <u>was cheap</u>.

wʌz ʧip

APPENDIX D

BACKGROUND AND LENGTH OF RESIDENCY INTERVIEW QUESTIONS

Family Name
姓
First name
名
Student Number
学籍番号

Age
年齢

Gender	Female	Male
性別	女性	男性

What year are you at the school?
- first year university
- second year university
- third year university
- fourth year university

学年
- 大学１年生
- 大学２年生
- 大学３年生
- 大学４年生

Do you go to an English conversation school now? (This is a different school from the question above.)
現在、英会話学校に通っていますか？ (先の質問で尋ねた学校とは違います)

Have you attended a Juku to study English?
英語を学ぶために塾に通ったことがありますか？

Were you a member of an English club?
英語クラブに所属していましたか？

FOLLOW UP QUESTION
補足質問

If 'Yes', how many hours a week do/did you attend classes?
”はい”と答えた人は週に何時間授業を受けていますか、または受けていましたか？

English Education History
英語学習について

At what age did you first learn English in terms of speaking?
話す英語を最初に習ったのは何歳のときですか？

At what age did you first learn English in terms of reading?
読む英語を最初に習ったのは何歳のときですか？

At what age did you first learn English in terms of writing?
書く英語を最初に習ったのは何歳のときですか？

Have you ever had a formal listening or learned the phonetic alphabet in class?
学校の授業でリスニングや発音記号を学んだことがありますか？

How many years (cumulative) have you spent learning English?
トータルで何年英語を勉強していますか？

How good in general do you feel you are at learning language (e.g. relative to friends or people you know)? What is your strength (e.g., speaking, listening, etc.)?
あなたの語学能力を評価してください。または、あなたが語学を勉強する上でどのくらい優れていると思いますか？(e.g. 友達や他の人と比べて)また、あなたは特にどの能力に優れていますか？（例：会話、リスニング）

How many hours a week on average do you spend speaking in English?

- less than 1 hour
- 1 to 2 hours
- 2 to 6 hours
- 6 to 10 hours
- more than 10 hours

平均して週に何時間くらい英語で話をしますか？

- 1 時間以下
- 1 時間〜2 時間
- 2 時間〜6 時間
- 6 時間〜10 時間
- 10 時間以上

How many hours a week on average do you spend reading English?

- less than 1 hour
- 1 to 2 hours
- 2 to 6 hours
- 6 to 10 hours
- more than 10 hours

平均して週に何時間くらい英語を読みますか？

1 時間以下
1 時間〜2 時間
2 時間〜6 時間
6 時間〜10 時間
10 時間以上

How many hours a week on average do you spend writing English?

less than 1 hour 1 to 2 hours
2 to 6 hours
6 to 10 hours
more than 10 hours

平均して週に何時間くらい英語を書きますか？ 平均して週に何時間くらい英語を読みますか？

1 時間以下
1 時間〜2 時間
2 時間〜6 時間
6 時間〜10 時間
10 時間以上

How many hours a week on average do you spend listening to English?

less than 1 hour
1 to 2 hours
2 to 6 hours
6 to 10 hours
more than 10 hours

平均して週に何時間くらい英語を聞きますか？

1 時間以下
1 時間〜2 時間
2 時間〜6 時間
6 時間〜10 時間
10 時間以上

Conversations with native English speakers
英語を母国語とする人（英語のネイティヴスピーカー）との会話について

Do you currently have opportunities to speak to a native speaker of English outside of the classroom (informally)?

Yes
No

現在授業以外で英語のネイティヴスピーカーの人と話す機会はありますか？

はい
いいえ

If 'Yes' (for previous question), how often do you meet?

Almost daily
Once or twice a week
Once or twice per month
About once every two months

前問で”はい”と答えた人はどのくらいの頻度で英語のネイティヴスピーカーの人と会いますか？

ほぼ毎日
週に 1〜2 回
月に 1〜2 回
2 ヶ月に 1 回程度

How many of your friends are native speakers of English?
あなたの友達で英語のネイティヴスピーカーは何人いますか？

Do you have a job where you use English?

Yes
No

英語を使う仕事(アルバイト)をしていますか？

はい
いいえ

If you have a job where you use English, how many hours per week do you use English at work?

less than 1 hour
1 to 2 hours
2 to 4 hours
4 to 6 hours
6 to 10 hours
more than 10 hours

英語を使う仕事(アルバイト)をしている人は、仕事場で週に何時間くらい英語を使っていますか？

1 時間以下
1 時間〜2 時間
2 時間〜4 時間
4 時間〜6 時間
6 時間〜10 時間
10 時間以上

Length of Residency
海外居住の長さについて

Have you ever lived/stayed in a country where English is spoken as the first or as the common language? (This can include homestay.)

Yes
No

今までに英語が公用語または共通語として使われている国に住んだことがありますか？（ホームステイを含みます）

はい
いいえ

If 'Yes', where did you live/stay?

USA
Canada
Australia
England
New Zealand
Other:

"はい"と答えた人、それはどこの国ですか？

アメリカ
カナダ
オーストラリア
イギリス
ニュージーランド
Other:

How old were you when you entered the country of your residence?
何歳のときにその国に住み始めましたか？(ホームステイを含む)

How old were you when you left the country of your residence?
あなたが住んでいた国(上記の国)を離れたのは何歳のときですか？

What was the purpose of your residence?

Homestay
Family transfer
Formal education
Language school
Other:

住んだ目的は何ですか？

ホームステイ
家族での移住
公的教育
語学学校
Other:

Please define your residency environment by commenting on the following statements.
あなたの海外での居住環境について以下の質問に答え、できるだけ詳しく教えて下さい 。

How was your residency? Were you satisfied?
居住環境に満足しましたか。

Did you speak English most days during your residency?
滞在期間中、ほとんど毎日英語を話しましたか。

Did you immerse yourself in the new culture?
新しい文化に溶け込むことが出来ましたか。

FOLLOW UP QUESTIONS
補足質問

Did you have a lot of support (e.g., conversation partner, host family) to help with learning English?
英語を学ぶうえでの多くのサポートを受けましたか。(例:ボランティアの会話パートナーやホストファミリーなど)

Did your English improve during your residency?
滞在期間中に英語が上達しましたか。

English Language Media
英語で発信されるメディアについて

How often do you watch English language media (television, movies)?

- 1-3 hours per week
- 3-7 hours per week
- 7-12 hours per week
- More than 12 hours per week
- I don't watch English language media

あなたはどの頻度で英語のメディア(例えばテレビ番組や映画)を観ますか？

- 週に 1〜3 時間
- 週に 3〜7 時間
- 週に 7〜12 時間
- 週に 12 時間以上
- 英語のメディアは観ない

FOLLOW UP QUESTION 補足質問
If there is anything else that you feel is interesting or important about your English language background (e.g., you love to sing English songs at karaoke), please comment below.
あなたの英語の経歴に影響を与えた興味深い事または重要なことがあれば、下の空欄にコメントしてください。(例: カラオケで洋楽を歌うことが好き。)