

ABSTRACT

Title of Dissertation: THE RELATIVE SIGNIFICANCE OF SYNTACTIC KNOWLEDGE AND VOCABULARY KNOWLEDGE IN SECOND LANGUAGE LISTENING COMPREHENSION

Payman Vafaei, Doctor of Philosophy, 2016

Directed By: Dr. Steve Ross, Second Language Acquisition

The main purpose of the current study was to examine the role of vocabulary knowledge (VK) and syntactic knowledge (SK) in L2 listening comprehension, as well as their relative significance. Unlike previous studies, the current project employed assessment tasks to measure *aural* and *proceduralized* VK and SK. In terms of VK, to avoid under-representing the construct, measures of both *breadth* (VB) and *depth* (VD) were included.

Additionally, the current study examined the role of VK and SK by accounting for individual differences in two important cognitive factors in L2 listening: metacognitive knowledge (MK) and working memory (WM). Also, to explore the role of VK and SK more fully, the current study accounted for the negative impact of anxiety on WM and L2 listening.

The study was carried out in an English as a Foreign Language (EFL) context, and participants were 263 Iranian learners at a wide range of English proficiency from lower-intermediate to advanced. Participants took a battery of ten linguistic, cognitive and affective measures. Then, the collected data were subjected to several preliminary analyses, but structural equation modeling (SEM) was then used as the primary analysis method to answer the study research questions.

Results of the preliminary analyses revealed that MK and WM were significant predictors of L2 listening ability; thus, they were kept in the main SEM analyses. The significant role of WM was only observed when the negative effect of anxiety on WM was accounted for. Preliminary analyses also showed that VB and VD were not distinct measures of VK. However, the results also showed that if VB and VD were considered separate, VD was a better predictor of L2 listening success.

The main analyses of the current study revealed a significant role for both VK and SK in explaining success in L2 listening comprehension, which differs from findings from previous empirical studies. However, SEM analysis did not reveal a statistically significant difference in terms of the predictive power of the two linguistic factors. Descriptive results of the SEM analysis, along with results from regression analysis, indicated to a more significant role for VK.

THE RELATIVE SIGNIFICANCE OF SYNTACTIC KNOWLEDGE AND
VOCABULARY KNOWLEDGE IN SECOND LANGUAGE LISTENING
COMPREHENSION

by

Payman Vafae

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2016

Advisory Committee:

Professor Steve Ross, Chair
Professor Robert DeKeyser
Professor Cathy Doughty
Professor Michael Long
Dean's Representative: Professor Donald (D.J.) Bolger

©Copyright by

Payman Vafaei

2016

Acknowledgements

Many people contributed to the completion of this dissertation, and so it is a pleasure to thank all who helped me along the way. First, I would like to extend my most sincere gratitude to my advisors and mentors, Dr. Steve Ross, Dr. Mike Long, Dr. Robert DeKeyser, and Dr. Catherine Doughty, for their unwavering support and dependability throughout this process. Their guidance, mentorship, and friendship were invaluable for the completion, not only of this dissertation but also my doctoral studies at the University of Maryland. They have always been sterling examples of researchers dedicated to the pursuit of understanding and knowledge, and their tremendous dedication to their students is impressive.

I would also like to thank my former mentors at Columbia University, Dr. Jim Purpura and Dr. Kirby Grabowski. They welcomed me to the US and helped me to gain trust and confidence in myself. They also taught me the most fundamental principles of research in second language acquisition and assessment.

In addition, I would like to thank my peers, in particular, Nancy Boblett, Dan Mann, Ryan DeRentz, Yuna Seong, Saerhim Oh, Ian Blood, Reese Heintner, Eddie Getman, Yuichi Suzuki, Illina Kachisnke, Eric Pelzl, Stephen O'Connell, and Jon Malone. The laughs, levity and support I got from you restored my sanity more than a few times.

Thanks also to my colleagues, Dr. Nahal Akbari, Dr. Ali Abasi, Dr. Fatemeh Keshavarz, and Manijeh Galedari, at the Persian program at the University of Maryland.

Working with them was a great pleasure, and they always fully supported me throughout my years of work as a Persian teacher at Maryland.

Data collection for the current project was impossible without the help of my dear friends in Iran. My sincere thanks go to Bahman Shahri, Mehdi Yaghoubi, Farokh Ziyae, Behrooz Yaghmayean and Iman Alipour for their unconditional support.

Last but certainly not least, my deepest gratitude to my family. To my wife, Sarvenaz, for her understanding, patience, optimism, and constant encouragement. Sarvenaz, thank you for being in my life and giving meaning to it. My deepest gratitude also to my parents, brother and sisters, whose care, support and faith in me knows no end.

Table of Contents

Acknowledgements.....	ii
List of Tables	vi
List of Figures	vii
Chapter 1: Introduction	1
Chapter 2: Literature Review	5
2.1 Componential Models of L2 Listening Ability	5
2.2 Top-down versus Bottom-up Processes in L2 listening Comprehension	10
2.3 Process-oriented Models of L2 Listening Ability	13
2.4 Interim Summary.....	17
2.5 The Role of Vocabulary Knowledge in L2 Listening	18
2.6 The Role of Syntactic Knowledge in L2 Listening.....	34
2.7 The Role of Metacognitive Knowledge in L2 Listening	38
2.8 The Role of Working Memory in L2 Listening	41
2.9 Influence of Anxiety on L2 Listening Performance.....	48
2.10 Overall Summary	51
Chapter 3: The Current Study	54
3.1 Research Questions	55
3.1.1 Preliminary Questions:	55
3.1.2 Main Questions:.....	56
3.2 Expected Results	56
Chapter 4: Method	59
4.1 Participants	59
4.2 Instruments	60
4.2.1 Listening Comprehension Test.....	60
4.2.2 Vocabulary Knowledge Measures.....	61
4.2.3 Syntactic Knowledge Measures.....	66
4.2.4 Metacognitive Knowledge Measure.....	72
4.2.5 Working Memory Measures	73
4.2.6 Anxiety Questionnaires	76
4.2.7 Background Questionnaire	78
4.3 Procedure.....	78
4.4 Data Analysis	79
Chapter 5: Results	83
5.1 Item and Reliability Analysis	83
5.2 Descriptive Statistics.....	86
5.3 Correlation Analysis.....	88
5.4 Regression Analysis	90
5.5 Confirmatory Factor Analysis	92
5.6 Structural Equation Modeling	96
Chapter 6: Discussion of Results and Conclusions.....	109
6.1 Discussion of Results	109
6.1.1 Preliminary research question 1	110
6.1.2 Preliminary research question 2	112
6.1.3 Preliminary research question 3	113
6.1.4 Preliminary research question 4	114

6.1.5 Preliminary research question 5	118
6.1.6 Main research question 1	120
6.1.7 Main research question 2.....	124
6.1.8 Main research question 3.....	127
6.2 Conclusions	130
6.3 Limitations and Directions for Future Research	137
Appendix A: VBT Sample Items	140
Appendix B: VDT Sample Item	141
Appendix C: GJT Sample Items	142
Appendix D: SCT Sample Items.....	143
Appendix E: MALQ Questionnaire	144
Appendix F: Anxiety Questionnaires.....	146
Bibliography	148

List of Tables

Table 1. Syntactic Target Structures Used in the Present Study.....	69
Table 2. Sample Items from the Aural GJT	70
Table 3. Sample Items from the Aural SCT	71
Table 4. Timing and Order of Task Administration	79
Table 5. Summary of Measures Used in the Present Study	82
Table 6. Reliability Estimates	85
Table 7. Descriptive Statistics.....	87
Table 8. Correlations among Separate Measured Variables	89
Table 9. Correlations among Composite Scores of the Measured Variables	89
Table 10. Multiple Regression Results for all Independent Variables	90
Table 11. Multiple Regression Results without MK	91
Table 12. Regression Models Comparison	91
Table 13. Simple Regression Results for MK	92
Table 14. Summary of Fit Indices for CFA Model 1.....	93
Table 15. Correlation Coefficients among Latent Variables	94
Table 16. Summary of Fit Indices for CFA Models 2 and 3	94
Table 17. Summary of Fit Indices for CFA Models 4, 5 and 6	95
Table 18. Chi-square Difference Tests Comparing Models 4, 5 and 6.....	96
Table 19. Summary of Fit Indices for CFA Models 7 and 8	96
Table 20. Summary of Fit Indices for SEM Models 1 to 6b.....	98
Table 21. Summary of Fit Indices for SEM Models 7, 8 and 9	100
Table 22. Summary of Fit Indices for SEM Model 10	103
Table 23. Summary of Fit Indices for SEM Model 11	104
Table 24. Chi-square Difference Tests Comparing SEM Models 11 to 15	105
Table 25. Summary of Fit Indices for SEM Model 12	108

List of Figures

Figure 1. Sample sequence for Blockspan task	74
Figure 2. Example of the Shapebuilder task	75
Figure 3. SEM Models 1 to 5, and 6b.	99
Figure 4. SEM Model 8	101
Figure 5. SEM Model 9	102
Figure 6. SEM Model 11	104
Figure 7. SEM Model 16	107

Chapter 1: Introduction

More than 45 percent of communication time is spent listening (Feyten, 1991; Lee & Hatesohl, 1993), and the development of second language (L2) listening skill has a significant influence on the development of other L2 skills (Dunkel, 1991; Rost, 2013; Vandergrift, 2011; Vandergrift & Baker, 2015). These studies have highlighted how important listening skill is for everyday language use, as well as the process of L2 acquisition. However, L2 listening is “the least understood and least researched skill” (Bae & Bachman, 1998, P. 383; Vandergrift, 2007, p. 191; Vandergrift & Goh, 2012), and “little is known about the variables that contribute to the development of L2 listening ability” (Vandergrift & Baker, 2015). More fine-grained analyses of the variables contributing to L2 listening ability are essential both for pedagogy and assessment.

From a pedagogical perspective, teachers need an informed grasp of the variables and development of L2 listening ability in order to adequately serve their students. Traditionally, classroom listening pedagogy has followed a testing model in which, after performing listening tasks, answers are given, with little feedback on how to improve listening skills themselves (Harding, Alderson & Brunfaut, 2015; Field, 1998, 2008). This limitation in pedagogy may be related to the fact that many classroom listening practices are influenced by large-scale proficiency testing. With teachers’ primary focus on proficiency outcome tests, there is little opportunity for the development and improvement of underlying factors contributing to listening ability.

In addition, the under-researched nature of L2 listening ability has negative implications for the validity of its assessment. L2 listening assessment research has paid extensive attention to test-taking or external factors that influence learners’ performance

on L2 listening tests; however, not much attention has been given to the listener or internal factors that underlie this important skill. Numerous studies have investigated the influence of different characteristics and conditions of listening assessment on test takers' performance. These studies investigated the influence of passage, task and question type (e.g., Brindley & Slatyer, 2002; Gilmore, 2007; Rost, 2006; Winke & Gass, 2013; Yinghui, 2006), time limits (Siemer & Reisenzen, 1998), number of and control over hearings (e.g., Henning, 1990) and note-taking (e.g., Lin, 2006). Although this line of research on external factors is valuable in terms of identifying variables that lead to construct-irrelevant variance in test scores, the key validity question remains: what *underlies* L2 listening ability?

Performance on an L2 listening test is a function of the interaction between test and testing situation characteristics – *external* factors – and listener characteristics – *internal* factors (Vandergrift, 2007). Only after the identification of *internal* factors, can *external* factors influencing listening ability be determined, as well as the extent to which external factors induce construct-relevant or irrelevant behaviors.

Internal factors in L2 listening ability are a set of general cognitive and linguistic abilities that explain individual differences in L2 listening comprehension. The purpose of the present dissertation was to investigate the relative significance of two major linguistic factors in L2 listening ability: (1) vocabulary knowledge, and (2) syntactic knowledge. Along with phonological knowledge, these two types of linguistic knowledge are recognized by Bloomfield et al. (2010) as the most important linguistic components of L2 listening ability. It is also important to note that vocabulary knowledge and syntactic knowledge are two linguistic components that can be explicitly taught and

practiced in classroom settings, and that teachers can focus on the development of their students in these two areas in order to improve L2 listening skill.

Pre-existing vocabulary knowledge has been presented in models of L2 listening comprehension as a way of assigning meaning, at least in part, to aural language (Buck, 2001; Rost, 2013). Syntactic knowledge has likewise been recognized as essential for the perception and interpretation of aural language (Call, 1985; Mecartty, 2000; Richards, 1983). However, empirical studies of the contribution of these two linguistic factors have been rare (Mecartty, 2000), and existing studies suffer from several methodological limitations, making conclusions hard to interpret and/or generalize. The most notable limitation of these studies is that in measuring vocabulary and syntactic knowledge, the two constructs have either been under or misrepresented.

In terms of measuring vocabulary knowledge, only one dimension of the construct – size or breadth – has been extensively researched, while depth of vocabulary knowledge has almost been ignored. Additionally, vocabulary knowledge has been tested only in the untimed written modality, which does not provide direct evidence of *proceduralized spoken* vocabulary knowledge. Also, in measuring syntactic knowledge, tests have been untimed written grammar tests, which reveal little about *proceduralized spoken* syntactic knowledge. According to models of L2 listening (e.g., Buck, 2001), *proceduralized spoken* knowledge of vocabulary and syntax play a key role in L2 listening, and results of studies that included measures of *controlled written* knowledge of the two constructs are incomplete and even misleading.

In addition, previous studies on internal factors of L2 listening have ignored the role of cognitive and affective factors when investigating the role of linguistic factors.

For example, metacognitive knowledge and working memory are the two important cognitive factors known to affect L2 listening comprehension. Anxiety, as an affective factor, has also been shown to influence the performance of L2 listeners (Bloomfield et al., 2010). In order to investigate the unique variance explained by linguistic factors, the influence of important cognitive and affective factors should be accounted for.

In the current dissertation, the role of *proceduralized spoken* knowledge of vocabulary (both its *breadth* and *depth*) and of syntax in L2 listening was examined, while the influence of metacognitive knowledge, working memory, and anxiety was also accounted for. To date, no study has investigated the concurrent role of all of these linguistic, cognitive and affective factors. Various studies have investigated the role of these different factors separately, or the joint effect of only two of these components. For example, in the most recent and comprehensive empirical study of internal L2 listening factors (Vandergrift and Baker, 2015) vocabulary depth, syntactic knowledge, and anxiety were all ignored.

Chapter 2 is a review of literature, beginning with a theoretical account of the components of the L2 listening construct and the processes underlying it. The chapter then moves to an evaluation of several major empirical studies investigating the significance of vocabulary knowledge, syntactic knowledge, metacognitive knowledge, working memory, and anxiety in L2 listening. Methodological limitations in the results of previous studies will be noted, as well as the manner in which they motivate the current dissertation.

Chapter 2: Literature Review

2.1 Componential Models of L2 Listening Ability

Theoretical constructs of L2 ability and its sub-constructs, including L2 listening ability, determine what tests should measure and what test scores mean. For this reason, construct definition is considered the first step in the process of test development, and is subsequently used for test score interpretation (Messick, 1994). There are two basic ways to define a construct. In the first, often referred to as the “competence-based” construct definition, underlying ability is defined in terms of its components. These are argued to be different types of knowledge, skills and abilities test-takers should be able to demonstrate on a test. In the second way, known as the “task-based” construct definition, tasks to be performed, rather than abilities or components, are identified (Chapelle, 1998). Both approaches to construct definition have advantages and disadvantages. However, for the purpose of the current study, the competence-based construct definition of L2 listening ability is used. A competence-based construct definition leads to componential models of learners’ ability, in which the components and their relationships can be delineated.

In componential models of L2 listening ability, it is assumed that consistencies in listening performance are attributable to the characteristics of the test-taker’s underlying competence, which reveals itself in a variety of settings and tasks (Buck, 2001). Therefore, in defining the construct of L2 listening ability, its components and processes must be outlined. Unsurprisingly, defining the construct of L2 listening ability has been challenging. As Buck (2001) noted, “[w]e still do not fully understand what the important sub-skills or components of listening ability are” (p. 97). As “listening comprehension is

a massively parallel interactive process taking advantage of information from a large number of sources, both linguistic and nonlinguistic,” it may be impossible to separate out its individual components (Buck, 1990, p.5; see also Rost, 2013).

Nevertheless, several attempts have been made to identify sub-components of L2 listening ability. For example, Richards (1983) proposed a list of “micro-skills” that underlie the broader skill of listening. Richards classified these micro-skills as sub-skills for both conversational and academic listening. Some examples of micro-skills for conversational listening are as follows:

- Ability to retain chunks of language of different lengths for short periods
- Ability to discriminate between the distinctive sounds of the target language
- Ability to detect meanings expressed in different grammatical form/sentence types
- Ability to process speech at different rates

(Richards, 1983, pp. 228–229)

However, the “micro-skills” approach posed by Richards (1983) has several limitations. This approach does not explicitly suggest how micro-skills might be operationalized, and the approach itself is atomistic in nature – there is no unifying principle to the list – so it lacks theoretical grounding (Harding et al., 2015).

To account for the limitations of Richards’ (1983) approach, especially its atheoretical nature, two major componential models of L2 listening ability have been proposed by Buck (2001) and Rost (2013). These models include both linguistic and non-linguistic components of L2 listening ability, and attempt to synthesize different sub-skills and sub-components of L2 listening into a single construct.

Following the principles of cognitive assessment (e.g., Tatsuoaka, 2009) and results from empirical studies using rule-space methodology (e.g., Buck & Tatsuoaka, 1998; Buck et al., 1997), Rost (2013) distinguished top-level and bottom-level abilities of L2 listening. Top-level abilities were defined as attributes generalizable to all language skills, while bottom-level abilities are attributes specific to listening skills. Using the results of a survey conducted to identify bottom-level abilities (viz., Rost, 2005), Rost (2013) proposed that a model of L2 listening ability should include the following components: phonological knowledge, syntactic knowledge, semantic knowledge, pragmatic knowledge and general knowledge.

Although Rost (2013) used the term “knowledge” to refer to the components of L2 listening ability, he defined the ability to *use* these types of knowledge as the building blocks of L2 listening ability rather than passive and declarative knowledge. To Rost, the use of syntactic, semantic and pragmatic knowledge *quickly* comprises the bottom-level (skill-specific) ability of L2 listening comprehension. Among non-linguistic components, Rost only included general knowledge, defined as “knowledge of the world and knowledge of how to utilize one’s knowledge in testing situations” (p. 212).

One major limitation of Rost’s (2013) model is that it does not account for general cognitive abilities in L2 listening. The model focuses on the ability to use different sources of knowledge quickly; however, it does not specify what it means to deploy that knowledge in the actual task of listening. This limitation of Rost’s model had been addressed in an earlier model of general language ability proposed by Bachman (1990) and Bachman and Palmer (1996).

According to Bachman and Palmer (1996), there is a division between knowledge of language – *linguistic* competence – and the ability to use this knowledge, defined as *strategic* competence. Linguistic competence or language knowledge is “a domain of information in memory that is available for use by the cognitive and metacognitive strategies in creating and interpreting discourse in language use” (p. 67). Language knowledge is thereby not synonymous with the ability to use language. Use of language knowledge requires, “a set of cognitive and metacognitive strategies that are higher order executive processes and provide a cognitive function in language use” (p. 70). Bachman and Palmer referred to these strategies as *strategic* competence.

Similar to general language ability, models of L2 listening ability should also include a component of general cognitive ability – strategic competence – in order to provide a more complete picture of L2 listening ability. In a modified form of Bachman’s (1990) model of general language ability, Buck (2001) proposed a model of L2 listening ability which included the cognitive as well as the linguistic component.

In Buck’s (2001) model, linguistic competence is defined as the knowledge of language that a listener brings to the listening situation. It is divided into four types: (1) grammatical, (2) discourse, (3) pragmatic, and (4) sociolinguistic. Grammatical¹ knowledge was defined as the ability to understand “short utterances on a literal semantic level, which includes knowledge of phonology, stress, intonation, spoken vocabulary and spoken syntax” (p. 104). Buck emphasized that *controlled* and *declarative* knowledge of these components of grammatical knowledge is not sufficient for strong L2 listening

¹ According to Buck’s (2002) model, syntactic knowledge is one of the components of grammatical knowledge. However, in L2 assessment literature, the two terms, grammatical knowledge and syntactic knowledge, are used interchangeably. To avoid confusion in the present dissertation, only the term syntactic knowledge is used in reference to listening ability.

skills. Instead, learners should focus on using grammatical knowledge automatically in real time. Thus, for effective L2 listening comprehension, declarative grammatical knowledge must be proceduralized or even automatized (Goh, 2005; Johnson, 1996). Discourse knowledge, pragmatic knowledge, and sociolinguistic knowledge refer, respectively, to the ability to understand longer utterances, the function or the illocutionary force of utterances, and the meaning of utterances in the context of situation.

Bachman and Palmer (1996) argued that strategic competence consists of cognitive and metacognitive strategies² that fulfill cognitive management functions. Cognitive strategies refer to the mental activities of processing and storing input in WM or long term memory. Metacognitive strategies are conscious or unconscious mental activities that perform an executive function in the management of cognitive strategies (Buck, 2001, p. 104).

Although Buck's (2001) model provides insight into the potential components of L2 listening ability, it is not without limitations. As Buck explains, the model does not specify the relative importance of the components, and does not provide criteria with which to operationalize the components. More importantly, the model does not explain the relationship between linguistic and strategic competence. This last limitation is because the distinction between linguistic and cognitive components is unclear, making it difficult to separate knowledge *of* the language from the general cognitive ability to *apply* that knowledge (Buck, 2001, p. 103; McNamara, 1996).

² The terms "metacognitive strategies" and "metacognitive knowledge" are used interchangeably in L2 listening literature. In the current dissertation, both terms refer to the same construct.

One way of examining the relationship among components of L2 listening ability and their relative significance is by comparing two important types of processes involved in L2 listening. These processes have usually been categorized as bottom-up and top-down processes, with a combination of the two leading to successful comprehension. Essentially, the two types of processes refer to the order in which different types of L2 listening components are applied during the task of L2 listening comprehension. The relative significance of the components of L2 listening ability varies, depending on the type of processing most dominant at a given point in an L2 listening task (Buck, 2001).

2.2 Top-down versus Bottom-up Processes in L2 listening Comprehension

The terms ‘top-down’ and ‘bottom-up’ processing are typically used in the context of L2 listening research to mark a distinction between the source of information learners rely on to comprehend aural input. In bottom-up processing, comprehension is achieved by relying on the information derived directly from perceptual sources. In top-down processing, the major source of comprehension is general knowledge and familiarity with context. However, it should be noted that these two terms refer to the direction of processing, rather than levels of processing (Vandergrift, 2011).

This means that in bottom-up processing, smaller units of information are progressively reshaped into larger ones. Conversely, in top-down processing, larger units exert an influence over processing and comprehension of smaller units (Field, 2004). In bottom-up processing, the stream of sound is segmented and decoded into meaningful units. When bottom-up processing is involved, listeners construct meaning by gradually combining smaller units at the phoneme level to create larger units at the discourse level.

On the other hand, in top-down processing, prior knowledge and knowledge of context is used to achieve comprehension (Vandergrift, 2011).

Generally, listening comprehension is primarily a top-down process, because larger units of information are always at work in the processing of even the smallest units of information. For example, processing a string of phonemes is influenced by the knowledge that a particular word exists. However, the importance of processing smaller units should not be minimized, as larger units can assist comprehension only if smaller units are perceived and processed effectively and accurately (Buck, 2001).

In the strict sense of the terms, top-down and bottom-up processing refer to the influence of larger and smaller units on mutual processing. However, in the context of empirical studies, the two terms are operationalized in a narrower sense. Top-down processing refers to activating and relying on world knowledge, general knowledge or schematic knowledge to understand input (Vandergrift, 1996; Young, 1998). Bottom-up processing, on the other hand, refers to relying heavily on perceptual data, and comprehension of input in terms of its smallest meaningful units, such as focusing attention at the word level (Flowerdew, 1994).

The relationship between bottom-up and top-down processing is complex, and listening comprehension is achieved as the result of an interaction between various information sources. Consequently, these two types of processes rarely operate independently (Buck, 2001; Field, 2004; Vandergrift, 2007). However, at a given point in L2 listening, one type of processing predominates, depending on listening task characteristics and requirements. For instance, verifying a specific detail – such as a date

or a number – requires more bottom-up processing than obtaining the gist of a message, which requires more top-down processing (Vandergrift, 2011).

Regardless of listening task characteristics, L2 proficiency level and listening ability determine which of the two processes plays a more significant role in L2 listening. One common view is that listeners at lower levels of listening ability mostly rely on bottom-up processing and focus attention at the word level, occupying much of their working memory. This can lead to a failure to combine words into larger meaningful units (Field, 2004; Lynch, 1998; Rubin, 1994). For low proficiency learners, it is therefore extremely important to develop and proceduralize linguistic knowledge (phonological, lexical and syntactic). Learners who process input efficiently can free up working memory resources to generate these larger, meaningful units. Relying on bottom-up processing at lower levels of ability also points to the role of metacognitive knowledge in top-down processing. Metacognitive knowledge allows learners to compensate for inaccurate and non-effective processing of the linguistic units due to limited vocabulary or weak syntactic knowledge (Goh, 2005).

On the other hand, studies also suggest that learners at higher levels of L2 listening ability rely on schematic or background knowledge for making inferences, predictions and elaborations. Skilled listeners frequently rely on background knowledge and contextual information to infer the meanings of unknown words to create meaningful interpretations. Better listeners are able to notice cues and activate top-down processing by making greater use of metacognitive knowledge (Chamot, 2005; Vandergrift, 2003a). The importance of metacognitive knowledge for success in L2 listening comprehension, especially by skilled listeners, has been well documented (e.g., Vandergrift, Goh,

Mareschal, Tafaghodatari, 2006). However, it is difficult to say that activation of background knowledge or reliance on metacognitive knowledge is *adequate* for more able listeners to comprehend aural input successfully (Goh, 1998, 2005; Vandergrift, 1996; Young, 1997). As Bonk (2000) explained, success in top-down processing depends on information obtained through bottom-up processing. If listeners fail to recognize certain words in aural input, they cannot employ top-down processing to access relevant contextual information. Therefore, word segmentation and recognition, as well as assigning meaning to recognized words, during bottom-up processing forms the basis for aural comprehension at all L2 listening ability levels (Rost, 2013).

As a result of these studies, it is now easier to understand what distinguishes more and less able L2 listeners. Better listeners often have more proceduralized linguistic knowledge. This allows more effective and rapid processing of information from perceptual sources, preserving enough available working memory capacity for chunking processed information into larger units. More skilled listeners are also more successful at top-down processing, as they make better use of background knowledge and metacognitive knowledge. In short, successful comprehension by L2 listeners at higher levels of ability is the result of effective processing of both types.

2.3 Process-oriented Models of L2 Listening Ability

In order to gain insight into the interplay between bottom-up and top-down processes, as well as linguistic and cognitive components in L2 listening, it is appropriate at this point to assess the stages of listening comprehension as outlined in “process-oriented” models of listening comprehension. Anderson (1995, 2009), whose initial work

on first language listening ability was later adopted by L2 scholars, broke down the process of listening comprehension into three interconnected stages: perception, parsing and utilization. These stages are interactive and do not demonstrate linear, one-way relationships. During the perception stage, primarily involving bottom-up processing, the sound stream is divided into phonemes, pauses and acoustic emphases and held in working memory. This is the first stage of the word-segmentation process, during which listeners may fail to recognize words and lose concentration, due to working memory overload (Goh, 2000). A phonetic representation of what is retained from the perception stage is passed on for further processing in the parsing stage (Vandergrift, 2011).

During the parsing stage, both top-down and bottom-up processing are involved. Listeners continue segmenting information retained in working memory from the perception stage, and begin to activate linguistic and contextual knowledge from long-term memory (Rost, 2005). By activating previous knowledge, listeners use cues, such as word onset, with other prosodic and phonotactic information to identify potential word candidates (Al-Jasser, 2008; Harley, 2000). Listeners then create propositions consisting of meaning-based representations of possible words in working memory. At the parsing stage, greater amounts of language knowledge – particularly of vocabulary – facilitate activation of word candidates. This allows listeners to hold meaning in larger units, or chunks, which frees up space in working memory. The outcome of the parsing stage of comprehension is “parsed speech,” i.e., identified words held in working memory for the next stage of processing, utilization (Vandergrift, 2011, pp.427-428).

During the utilization stage, top-down processing of previous stages results in new information being compared with existing information in long-term memory for

interpretation. In this stage, top-down processing is assumed to be dominant because listeners access non-linguistic information to interpret retained information from parsing. Listening comprehension thus becomes a problem-solving activity, during which listeners reconcile linguistic input with non-linguistic knowledge to interpret the speaker's meaning (Goh, 2000; Vandergrift, 2011).

Success in all three stages of processing, as well as bottom-up and top-down processing, depends on the degree to which listeners can coordinate these processes efficiently. L2 learners typically have limited language knowledge, and are unable to automatically process everything they hear. For this reason, controlled processing is more common among L2 learners. Controlled processing requires more time than automatic, so given limited working memory capacity and the speed of incoming input, comprehension is inhibited.

Two studies illustrate this point. Drawing on Anderson's (1995) model, Goh (2000) and Sun (2002) examined listening problems at perception, parsing, and utilization stages and found that recognizing vocabulary and segmenting speech were listeners' two major problems at the perception stage. Difficulty in vocabulary recognition and speech segmentation is caused by limited proceduralized vocabulary and syntactic knowledge. To overcome comprehension problems caused by limited linguistic knowledge and to avoid comprehension breakdowns, L2 listeners possibly use compensatory strategies, such as metacognitive strategies (Vandergrift, 2011).

More specific to the context of L2 listening, and similarly to Anderson (1995, 2009), Field (2013) divided the process of L2 listening comprehension into five stages: input decoding, lexical search, parsing, meaning construction, and discourse

representation. In Field's model, input decoding, lexical search and parsing are considered low-level or bottom-up processes, while meaning construction and discourse representation are high-level, or top-down, processes.

At lower processing levels, input for processing is derived directly from aural input, and different stages of comprehension are supported by various *linguistic* knowledge sources (Field, 2013, p. 97). Thus, processing at lower levels is considered to be bottom-up. According to Field, input decoding is supported by phonological knowledge, lexical search by lexical knowledge, and parsing by syntactic knowledge. At the parsing stage, incoming aural input has been transformed from a phonological string into a proposition, which becomes the basis of literal understanding by the listener.

In Field's model, at higher levels of processing, the literal understanding gained at the parsing stage is related to the "circumstances in which it was produced" (Field, 2013, p. 100). This results in understanding at the "meaning construction" phase. These "circumstances" include both the context of the utterance and understanding the speaker's intentions. During the final stage, labeled "discourse representation", listeners integrate what they have understood from the utterance into an understanding of the entire listening event. Meaning construction and discourse representation stages are supported by pragmatic knowledge, as well as world knowledge listeners bring to a listening task. Therefore, these two stages are considered top-down.

In line with Anderson (1995, 2009), Field (2013) also explained that L2 listening comprehension is not a linear progression from lower to higher processing levels. Instead, different levels operate concurrently, with breakdowns at one level compensated for by

processed information from another. Simultaneous breakdowns at more than one level can lead to complete misunderstanding.

2.4 Interim Summary

Successful L2 listening comprehension can be achieved if sufficient proceduralized linguistic knowledge is available. Such knowledge leads to fluency both in top-down and bottom-up processing at multiple stages of comprehension (i.e., perception, parsing and utilization). The amount of available proceduralized linguistic knowledge can also influence the amount of available working memory capacity, storage and processing mechanism, which is instrumental at all stages of comprehension. L2 listeners with more proceduralized linguistic knowledge, particularly vocabulary and syntactic knowledge (Mecarty, 2000; Thompson, 1995), have more available working memory resources, thereby saving space for retaining information and achieving successful comprehension (Vandergrift, 2004).

The use of metacognitive knowledge can also impact results of L2 listening comprehension, especially when sufficient linguistic knowledge (vocabulary and syntactic knowledge) and cognitive resources (i.e., working memory) are not available (Field, 2001; Vandergrift, 2004, 2007, 2011). In addition, affective factors such as anxiety can significantly influence L2 listening comprehension performance. Listening comprehension is a complex cognitive task, and “substantial evidence suggests that anxiety is associated with poor performance in complex cognitive tasks” (Dutke & Stober, 2001, p. 2). Obviously, anxiety is not an underlying component of L2 listening

ability; however, in empirical studies on the linguistic and cognitive factors of L2 listening ability, the role of this affective factor should be accounted for.

Empirical studies on the role of linguistic and non-linguistic factors of L2 listening ability are intended to reveal which factors explain the largest amount of individual differences in this ability. However, these studies are limited both quantitatively and qualitatively. The following sections will summarize results of empirical studies on the role of vocabulary knowledge, syntactic knowledge, metacognitive knowledge, working memory, and anxiety in L2 listening ability.

2.5 The Role of Vocabulary Knowledge in L2 Listening

It has long been argued that vocabulary knowledge is an important component of global L2 proficiency (Read, 2000), and it plays a central role in understanding aural input. Aural word *recognition* skills are crucial in effective use of bottom-up processing, which allows listeners enough available working memory capacity for further processing of the input (Segalowitz & Segalowitz, 1993; Tyler, 2001; Vandergrift, 2004). Knowledge of word *meanings* has also been recognized as essential for deriving meaning at least in part from aural input (Mecartty, 2000; Rost, 1990).

In more recent decades, researchers have argued that vocabulary knowledge is a multidimensional construct (Qian & Schedle, 2004). There is a consensus among vocabulary researchers that the construct of vocabulary knowledge should comprise at least three dimensions of *breadth*, *depth* and *fluency* (Daller, Milton, & Treffers-Daller, 2007; Nation, 2001; Read, 2000; Qian, 1999; Wesche & Paribakht, 1996).

Vocabulary breadth, or size of the lexicon, is typically defined as the number of the words for which a learner knows the meaning (Nation, 2001; Staehr, 2009; Qian, 2002), and depth of vocabulary knowledge, or quality of word knowledge, refers to how well a learner knows a word (Qian & Schedle, 2004, p.29). Fluency of vocabulary knowledge has been defined as the extent to which L2 learners can recognize, process, or access the form and meaning of a word for language *use*. Numerous empirical studies have revealed that measures of different aspects of vocabulary knowledge are correlated with measures of general L2 proficiency in statistically significant ways (e.g., Beglar, 2010; Crossley, Salsbury, McNamara, 2011; Gyllstad, 2005, 2007; Meara, 1996; Nizonkiza, 2011; Zareva et al., 2005), and distinguish learners of different L2 proficiency levels (e.g., Milton & Meara, 2003; Milton & Alexiou, 2009; Milton, Wade & Hopkins, 2010; Nation, 2001, 2006).

Breadth has been considered the most basic aspect of vocabulary knowledge, and learners with larger vocabulary sizes are more proficient language users (Meara, 1996a). However, as breadth of vocabulary knowledge can only mean having a *superficial* knowledge of meaning, depth of vocabulary knowledge seems to be playing the primary role in distinguishing success of L2 learners in complex linguistic and cognitive tasks such as L2 reading and listening comprehension (Qian & Schedl, 2004). This more substantial role of depth of vocabulary knowledge can be better understood based on the proposed continuum of vocabulary knowledge by Henriksen (1999).

Henriksen (1999) proposed a partial-to-precise continuum of vocabulary knowledge. On this continuum, breadth of vocabulary knowledge is located toward the partial-knowledge end, and depth of vocabulary knowledge falls toward the precise-

knowledge end. However, as Qian (2002) argues, vocabulary knowledge grows in an incremental fashion. This means that words acquired at earlier stages of L2 learning are more likely to have much more depth than words recently acquired. Qian also argues that the breadth of vocabulary knowledge forms the basis for the depth of vocabulary knowledge to develop because “the more words a learner knows, the more likely it is that he or she will have a greater depth of knowledge for those words” (p. 517). According to Schmitt (2010), L2 learners surely know a word’s basic meaning before they have a full knowledge of the word in terms of its associations with other words and develop, for example, collocational competence. These accounts of the relationship between breadth and depth of vocabulary knowledge suggest that rather than being separate dimensions of vocabulary knowledge, they are different levels of vocabulary knowledge on a continuum from partial to precise, with breadth as the prerequisite for depth.

However, it should be noted that breadth and depth of vocabulary knowledge interact closely with one another in fundamental processes of L2 vocabulary use. The importance of these important dimensions or levels of vocabulary knowledge varies according to the specific purposes of language use (Qian & Schedl, 2004). For example, in complex linguistic tasks such as reading or listening comprehension, in which precise meaning of propositions should be understood clearly, it seems that depth of vocabulary knowledge plays a more fundamental role. Partial or superficial recognition and knowing the meaning of words is not sufficient to fulfil the linguistic needs of communicative L2 use. In meaningful and communicative L2 contexts, “depth of vocabulary knowledge occupies a primary and central place in the multidimensional domain of vocabulary knowledge” (Qian & Schedl, 2004, p. 30). Although recognizing words and knowing the

form-meaning links for the recognized words are the most basic and essential aspects of vocabulary knowledge, much more must be known about words if they are to be effectively used in online L2 production and comprehension (Schmitt, 2010).

According to Henriksen (1999), depth of vocabulary knowledge mainly involves the knowledge of the relationships between words through, for example, knowledge of antonyms, synonyms and collocations. Qian (1999; 2002) operationalized the construct of depth of vocabulary knowledge in terms of knowledge of synonyms and collocations, and investigated the role breadth and depth of vocabulary knowledge play in L2 reading comprehension ability. The study found an important role for both breadth and depth of vocabulary knowledge in reading. Unfortunately, the nature of the data analysis methods in Qian's studies could not reveal the relative significance of these two dimensions of vocabulary knowledge.

In terms of the relationship between vocabulary knowledge and L2 listening ability, a number of studies have suggested that vocabulary knowledge plays a central role in successful listening comprehension among listeners at different levels of ability (Andringa et al., 2012; Bonk, 2000; Hasan, 2000; Kelly, 1991; Liao, 2007; Mecarty, 2000; Milton, 2010, 2013; Milton & Hopkins 2006; Milton, Wade & Hopkins, 2010; Staehr, 2007, 2008, 2009; Vandergrift & Baker, forthcoming). Kelly (1991) examined listening comprehension problems among advanced EFL learners by analyzing errors from BBC news radio transcriptions. Kelly categorized the errors into three types: perceptual, lexical, and syntactic. The study concluded by arguing that lack of vocabulary knowledge is the primary obstacle to successful listening comprehension for EFL learners. The results of Kelly's study revealed that lexical errors accounted for 65.5% of

all errors impairing comprehension. However, Kelly's methodology was error analysis, focusing on data from a very small corpus of only 148 listening errors. 100 of these errors were committed by a single participant.

Hasan (2000) conducted a questionnaire study to identify factors contributing to L2 listening comprehension difficulty. His participants were Arabic EFL learners at the intermediate level. He asked his participants to state what they thought had impaired their listening comprehension, and found that the majority reported experiencing difficulty in predicting a missing word or phrase. Hasan concluded that this difficulty could be attributable both to learner reliance on word-by-word processing approaches and to limited vocabulary knowledge. As in Kelly's (1991) study, Hasan (2000) did not include direct measures either of vocabulary knowledge or listening comprehension. An important final note on Hasan's design is that results from self-report studies such as these should be interpreted cautiously.

Bonk (2002) carried out a more rigorous study of the relationship between L2 vocabulary knowledge and listening comprehension, measuring vocabulary knowledge as familiarity with words in a passage through dictation. Listening comprehension was then measured through recall protocols. A modest association of .45 was found between vocabulary knowledge and listening comprehension. To explain such a relatively low association, Bonk analyzed the percentage of words participants knew, comparing it with comprehension scores. Some participants were able to achieve a high level of comprehension even while recognizing less than 75% of the words in the listening passages. Oddly, some participants were familiar with 90% of the words in the passages, but did not achieve high comprehension scores. Bonk's (2002) results underscored the

complex nature of listening comprehension, as well as its relationship with vocabulary knowledge, but these results should be interpreted with caution for multiple reasons.

In Bonk's (2000) study, both listening comprehension and vocabulary knowledge measures were questionable. Bonk used dictation as the measure of word recognition, and recall protocols as a measure of gist comprehension. Results could have been different had he employed standardized measures of vocabulary knowledge and listening comprehension ability. Low reliability in non-standardized measures, which Bonk reported, can often lead to attenuation in the magnitude of correlation coefficients. In addition, results of studies such as these are limited, in the sense that the effect of other important effective variables in L2 listening comprehension is not accounted for. Thus, it remains unclear why some learners with knowledge of more words were outperformed by learners with more limited vocabulary knowledge.

To avoid the limitations of a lack of standardized measures, a study by Stæhr (2007) employed the Vocabulary Levels Test (VLT) (Nation, 1983, 1990) as a measure of breadth of vocabulary knowledge. A standardized listening measure was also used – the Cambridge Certificate of Proficiency in English – to investigate the relationship between vocabulary knowledge and L2 listening ability. Stæhr (2007) reported a substantial correlation (.70) between receptive vocabulary size and listening comprehension among advanced EFL learners (N= 115). The higher correlation discovered in Stæhr's (2007) study signifies the importance of standardized measures in these research contexts.

In a follow-up study, Stæhr (2008) examined the relationship between receptive vocabulary knowledge and listening comprehension among EFL learners (N=88) in lower

secondary education in Denmark. He also investigated the role of familiarity with the 2,000 most frequent word families in achieving passing grades in listening comprehension, as well as L2 reading and writing skills. Stæhr used an updated version of the VLT (Schmitt, Schmitt and Clapham, 2001) as the measure of vocabulary knowledge size, and a multiple-choice test of listening comprehension that Denmark uses as an exit test for its school system. Stæhr's (2008) analysis used binary logistic regression, and results showed that 39% of the variance in the listening scores – the ability of participants to score above the mean – was accounted for by variance in vocabulary size test scores. For any single factor, this amount of explained variance is considered substantial. These results again confirmed the centrality of receptive vocabulary size for L2 listening comprehension.

One major limitation of Stæhr's (2007, 2008) studies is that they included only one dimension of vocabulary knowledge, size. However, as explained previously and suggested by Milton (2013), different dimensions of vocabulary knowledge might need to be measured separately, and separate and combined effects of the two can give a fuller picture of the nature of the relationship between vocabulary knowledge and listening comprehension. For example, in the context of L2 reading comprehension, Qian's (1999; 2002) previously mentioned studies suggested that vocabulary depth, as well as breadth, contribute significantly to predictions of L2 performance in academic reading.

Relatively few studies have been undertaken regarding the role of multiple dimensions of vocabulary knowledge in L2 listening comprehension. Mecarty (2000) examined the relationship between vocabulary knowledge and syntactic knowledge in L2 listening ability among lower-intermediate learners of Spanish. Although the dimension

of vocabulary knowledge included in the study was not clearly outlined, the description of the vocabulary test indicated that it primarily tapped *depth* of vocabulary knowledge. Mecartty measured vocabulary knowledge through a word-association and a word-antonym task. Knowledge of association and antonyms are considered sub-constructs of depth of vocabulary knowledge (e.g., Read, 2007). Mecartty found that, unlike syntactic knowledge, vocabulary knowledge was a significant predictor of L2 listening ability. These results strongly support claims that depth of vocabulary knowledge, along with breadth of vocabulary knowledge, are both robust building blocks of L2 listening ability. However, since there was no vocabulary size measure in Mecartty's (2000) study, it is not clear whether depth remains a significant predictor of L2 listening when breadth of vocabulary knowledge is taken into account. Therefore, studies including measures of both aspects would be more informative in exploring the unique variance each can explain in L2 listening.

A recent and methodologically rigorous study conducted assessing both types of vocabulary knowledge was Stæhr (2009). Following up on previous studies, Stæhr included separate measures of vocabulary breadth and vocabulary depth, and investigated the relationship between the two in L2 listening ability among 115 advanced Danish EFL learners. The second version of the VLD (Schmitt et al., 2001) was employed as a measure of breadth, and the Word Associates Test (WAT) (Read, 1993, 1998) as a measure of depth. For the listening comprehension measure, the Cambridge Certificate of Proficiency in English (CPE), a standardized measure of listening, was utilized. The CPE listening test is designed to measure listening comprehension at a very advanced level

(CEFR level C2), involving various target-language situations. It tests listening sub-skills such as gist, details, stated and non-stated opinions, and inferences.

Stæhr (2009) conducted a Pearson product-moment correlation analysis to examine the strength of association between scores on the listening measure and the two vocabulary tests. The results revealed that VLD and WAT scores correlated with listening test scores at .7 and .65 levels, respectively. Both of these measures had a strong association with the listening test scores, and results of a regression analysis showed that vocabulary size and depth together accounted for 51% of the variance in listening scores.

To examine the relative contributions of breadth and depth of vocabulary knowledge to variance in listening test scores, Stæhr (2009) conducted a hierarchical regression analysis by entering VLD and WAT scores into the regression model in two separate steps. Because Stæhr argued for vocabulary size as the basis of lexical knowledge, he first entered VLD scores into the regression model, followed by WAT scores. The results from the hierarchical procedure showed that VLD scores alone accounted for 49% of the variance in listening comprehension scores, with only 2% of additional variance explained by WAT scores.

Stæhr (2009) concluded that both breadth and depth of vocabulary knowledge are associated with L2 listening ability. Each of these two vocabulary measures could explain a significant portion of the variance in the test scores (R^2 for VLD and WAT were .49 and .43, respectively). However, as revealed by the hierarchical regression, in contrast to vocabulary size, depth of vocabulary knowledge contributed very little to successful listening comprehension. Stæhr (2009) interpreted these results as evidence that vocabulary size is the basic component of vocabulary knowledge in L2 listening ability,

and depth of vocabulary knowledge does not play a separate role. However, these results and conclusions should be interpreted with caution for a number of reasons.

First, in Stæhr's (2009) study, the two measures of vocabulary knowledge were highly correlated ($r = .80$), resulting in possible multicollinearity between the two types of knowledge. This led to a very limited amount of unique variance explained by each of the two measures. If the researcher had entered the WAT scores into the regression model first, the results and conclusions about vocabulary size and depth and their relationship with L2 listening ability could easily have been the exact opposite.

In addition to the close relationship between different dimensions of vocabulary knowledge, the case of multicollinearity in Stæhr's (2009) study can be explained through examination of the measures used in his study. Stæhr used the VLD as the measure of vocabulary size. In the VLD, participants are asked to match words with definitions. In many items, these definitions are synonyms or words related to the target vocabulary items. There is therefore an overlap between size and depth of knowledge in the VLD measure. If the aim is to examine the unique contribution of the two dimensions, more distinct measures of them are needed.

In addition, both of the vocabulary measures used in Stæhr (2009) were in a written format. As Milton (2013) explains, correlations between vocabulary measures and listening ability are typically weaker than correlations with written skills. This may be due to the fact that oral language is less sophisticated lexically than written language, but smaller correlations may also be an artifact of study methodology. In the majority of L2 listening studies, vocabulary measures were delivered through the medium of writing. This could be a potential limitation in results and conclusions, given the fact that a word

recognized in its written form will not necessarily be recognized in aural input. Milton et al. (2010) argued further that word knowledge may be primarily stored in aural form in the mental lexicon. This may particularly be true if vocabulary learning takes place predominantly through oral input, as Ellis (1994, p. 24) suggests. Therefore, with regard to L2 listening, vocabulary tests may tap into vocabulary knowledge in the wrong form for the relationship to be made fully clear. To gain a clearer picture of the relationship between vocabulary knowledge and listening ability, vocabulary tests should be administered aurally.

Milton et al., (2010) investigated the relationship between vocabulary size scores and International English Language Testing System (IELTS) sub-skill scores with 29 intermediate and advanced level EFL learners. Researchers examined whether vocabulary knowledge can be tested in different modalities, written and aural, and whether the measurement of vocabulary size in two separate modalities can better explain performance in all four skills. To measure written vocabulary size, they used X-Lex (Meara & Milton, 2003), and to measure aural vocabulary size, A-Lex (Milton & Hopkins, 2006).

Milton et al. (2010) hypothesized that scores from the written test would correlate best with reading and writing test scores, and scores from the aural test with speaking test scores. They argued that a combination of the two tests would best explain scores from the listening test, as the IELTS listening sub-test involves both written and aural input. The results of correlation and regression analyses revealed that, in terms of reading and listening, scores from the written vocabulary size test (X-Lex) had relatively similar significant correlations (Reading: .54 , Listening: .52). However, the aural vocabulary

size test (A-Lex) correlated significantly with listening only. Results from Milton et al. (2010) suggest that there is no significant difference in the strength of the relationship between written vocabulary tests versus reading or listening tests. However, aural vocabulary tests are more directly related to listening test scores.

Vandergrift and Baker (2015) sought to examine the role of several components of L2 listening. In their wide-ranging study, they included measures for the following variables: first language (L1) listening ability, L1 vocabulary knowledge, L2 vocabulary knowledge, auditory discrimination ability, metacognitive knowledge, and working memory. To overcome limitations of previous studies that included written measures of L2 vocabulary knowledge only, they used an aural vocabulary test. Participants in Vandergrift and Baker were 157 seventh-grade students in year one of a French Canadian immersion program. Given the age of participants, Vandergrift and Baker used the Peabody Picture Vocabulary Test (PPVT) (Dunn & Dunn, 2007), which was developed for young learners of French.

The PPVT includes a test of meaning recognition, and assesses breadth of vocabulary knowledge. Rather than choosing from written alternatives, participants indicate comprehension of a spoken stimulus word by pointing to the correct image of that word from a panel of four picture choices. Vandergrift and Baker (2015) conducted a correlation analysis, finding that L2 vocabulary knowledge had the strongest association with L2 listening among all other variables. The correlation between the two measures was .51, indicating that L2 vocabulary knowledge accounted for 26% of the variance in the listening scores. These results again highlight the important role of vocabulary size in L2 listening, especially when tested aurally.

Neither Milton et al. (2010) nor Vandergrift & Baker (2015) included an aural measure of vocabulary depth in their studies. Additionally, both involved correlational analyses only, by which magnitude of associations are often attenuated due to measurement error. Studies including aural measures of both breadth and depth are needed to further illuminate the relative significance of each to listening skills. Moreover, more sophisticated statistical methods with less measurement error can better investigate the amount of unique variance each type of knowledge contributes to L2 listening.

An important question arises when the relationship between vocabulary knowledge and listening ability is discussed – how much vocabulary is needed for successful L2 listening comprehension? The question has been addressed in studies which investigated *lexical coverage* in written and spoken discourse, and its relationship to comprehension. Lexical or text coverage has been defined as the percentage of familiar words in written or spoken discourse that enables successful comprehension (Nation & Waring, 1997; Adolphs & Schmitt, 2003). Lexical coverage is an informative measure, allowing scholars, testers, teachers and learners to estimate vocabulary size necessary for successful comprehension of written and spoken texts (van Zeeland & Schmitt, 2012).

Most research on lexical coverage and its relationship with L2 comprehension has been carried out in reading. Robust findings include the percentage of vocabulary necessary for successful reading comprehension, along with the number of word families required to reach this percentage (van Zeeland & Schmitt, 2012). Results of several studies have shown that 98 percent lexical coverage is essential for successful L2 reading comprehension (e.g., Hu & Nation, 2000). Nation (2006) calculated a vocabulary size of

8,000–9,000 word families necessary to reach the 98 percent lexical coverage level for L2 English reading.

Unfortunately, few studies have focused on L2 listening comprehension and the issue of lexical coverage. Lack of direct research on listening has motivated speculations about lexical coverage in listening based on findings about reading. For example, Nation (2006) used Hu and Nation's (2000) 98 percent figure for reading to calculate vocabulary size requirements for listening. Using the Wellington Corpus of Spoken English (WCSE) and BNC word frequency lists to analyze the vocabulary of spoken English, Nation concluded that a vocabulary size of 6,000–7,000 word families is required to reach the 98 percent lexical coverage in spoken texts. However, because reading and listening texts are both qualitatively and quantitatively different, direct research was needed on the relationship between lexical coverage and required vocabulary size for L2 listening. Studies by Bonk (2000), Stæhr (2009), and van Zeeland & Schmitt (2012) have tried to address this issue.

Bonk (2000) showed that learners with the lexical coverage under 90 percent showed poor comprehension and those with scores above 95 percent were successful in L2 listening comprehension. However, as mentioned earlier, both vocabulary size and listening comprehension measures in Bonk's study were questionable. Stæhr (2009) used the scores from the VLT and CEP tests to indirectly calculate the percentage of lexical coverage and number of word families required to meet the percentage. Using the 98 percent estimation of Nation (2006), Stæhr found that participants at the level of 5,000 word families (according to the VLT) achieved a higher level of listening comprehension than participants with knowledge of fewer word families. However, the indirectness of

Stæhr's method of assessing lexical coverage should result in great caution when interpreting his estimates.

Van Zeeland & Schmitt (2012) *directly* investigated lexical coverage with regard to L2 listening comprehension. They manipulated the vocabulary coverage of four informal spoken narrative passages and measured listening comprehension for factual information through multiple-choice items. Results showed that learners could adequately comprehend spoken texts with only 90 percent coverage, although there was considerable variation among listening test scores at this level of vocabulary knowledge. 95 percent lexical coverage for learners enabled them to demonstrate relatively good comprehension, but with much less variation. Based on a 95 percent coverage figure, Van Zeeland & Schmitt concluded that learners would need to know between 2,000 and 3,000 word families for adequate listening comprehension, which is substantially smaller than Nation's (2006) calculation of 6,000–7,000 word families based on a 98 percent figure.

One major limitation of van Zeeland & Schmitt's (2012) study was that their estimations were based on comprehension scores only from short, informal spoken stories. As they admitted, short story narratives are most comprehensible for listeners, and other genres would require a larger percentage of lexical coverage and vocabulary size for successful comprehension. In this sense, it seems that Nation's (2006) estimate of needed vocabulary size (6,000–7,000 word families) based on 98 percentage of lexical coverage is more dependable for L2 listening research, especially at advanced levels of proficiency.

In summary, although there has been substantial evidence for the importance of vocabulary knowledge as one of the components of L2 listening comprehension, this

topic merits further attention. The empirical studies in this area, while contributing helpful insights into listening comprehension, have been limited in several ways. In the majority of studies, the relationship of only one dimension of vocabulary knowledge, size or breadth, has been examined. In a few studies that included depth of vocabulary knowledge, the measures used to tap breadth and depth were not distinct enough to tease them apart. In the majority of studies, vocabulary tests have been in the written format. This potentially obscures the relationship between vocabulary knowledge and aural skills. Milton et al. (2010) and Vandergrift & Baker (2015) used aural vocabulary tests, but their measures were limited to vocabulary size, with no measure of depth.

Moreover, all the studies in this area have used statistical analyses such as zero-order correlation or regression. Results of such studies are only at the measurement level, but claims have often gone beyond the data to conclusions at the construct level. More sophisticated data analysis methods, such as structural equation modeling, are required if construct-level conclusions about the relationships between vocabulary knowledge dimensions and L2 listening ability are to be made.

Finally, in order to gain a fuller picture of the role of different components of L2 listening (e.g., vocabulary knowledge), empirical studies should include as many components as possible. Only by inclusion of other important components in a single study can the unique contributions of a single component, such as vocabulary knowledge, be demonstrated. Very few studies in the area of L2 listening have accounted for the role of several concurrent components. The following section provides a summary and evaluation of studies that have examined the role of syntactic knowledge and vocabulary

knowledge concurrently, providing insight into the relative importance of these linguistic components of L2 listening ability.

2.6 The Role of Syntactic Knowledge in L2 Listening

Compared to lexical and phonological knowledge, the role of syntactic knowledge in L2 listening is the least examined (Goh, 2005). One reason for this underrepresentation in the literature could be theoretical hypotheses about its minimal role. Field (2008) argued that in terms of processing and identification of function and content words, L2 listeners appear to be more successful in identifying content words. Their relative success at identifying content words is not surprising, as content words carry meaning. Due to working memory limitations, L2 listeners need to attend to aural input selectively (Vandergrift, 2011). Thus, L2 listeners focus on content words and ignore function words, in order to optimize L2 listening comprehension. If content words fall within the domain of vocabulary knowledge, while function words belong to morphosyntactic knowledge, Field's and Vandergrift's arguments theoretically justify a weaker role for syntactic knowledge in L2 listening. Along the same lines, Widdowson (2003, p. 39) argued that listeners tend to "edit grammar out" to catch the proposition of what they hear. Therefore, the importance of grammar seems to be particularly elusive to define in comprehending speech (Ranta, 2009).

Despite these arguments, it seems intuitively and theoretically plausible to assume that syntactic knowledge plays a major role in L2 listening (Thompson, 1995). The role of syntactic knowledge in L2 listening has been explored both in componential (Buck, 2001; Rost, 2013) and process-oriented (Anderson, 2009; Field, 2013) models of L2

listening. Contrary to the claims of Field (2008b), Call (1985) and Richards (1983) asserted that syntactic knowledge plays a crucial role in segmenting and interpreting streams of speech. In addition, Rost (1990) claimed that listeners' syntactic knowledge influences their ability to predict what they will hear next while listening and processing spoken utterances. Sun (2002) also claimed that L2 listeners with limited syntactic knowledge may have difficulty segmenting streams of speech when many words are linked together.

Findings from empirical studies on the impact of morphological and syntactic input modifications on comprehension likewise attest to the significant role of syntactic knowledge in L2 listening. Long (1985) showed that nonnative speakers performed significantly better with a modified lecture that was less syntactically complex than an unmodified version. Similarly, Cervantes and Gainer (1992) showed that syntactic simplification improves L2 listening comprehension. They found that Japanese EFL learners performed better on a recall cloze test when listening to less syntactically complex passages. Chiang and Dunkel (1992) also reported that high-intermediate learners benefited from syntactic modifications in comprehending lectures. Teng (2001) examined the impact of syntactic modification on L2 listening comprehension, finding that learners comprehend passages with syntactic modification better than unmodified ones.

However, to date, very few empirical studies have directly investigated the role of syntactic knowledge in L2 listening comprehension. The extent to which syntactic knowledge contributes to L2 listening and serves to predict L2 listening ability remains an open question (Mecartty, 2000; Liao, 2007). In the comprehensive study by

Vandergrift & Baker (2015), the researchers examined the role of several related variables: L1 listening ability, L1 vocabulary knowledge, L2 vocabulary knowledge, auditory discrimination ability, metacognitive knowledge, and working memory. However, the role of syntactic knowledge was omitted.

In the most frequently cited study of the role of syntactic knowledge in L2 listening, Mecartty (2000) compared the roles of vocabulary and syntactic knowledge in L2 reading and listening. The participants were 154 learners of Spanish enrolled in a four-semester Spanish program in the USA, and were categorized as high beginners in terms of Spanish proficiency. Mecartty divided the participants into two groups, each of which took either a reading or listening comprehension test. In addition to the comprehension tests, measures of vocabulary and syntactic knowledge were utilized.

Mecartty (2000) found that scores on the vocabulary and syntactic knowledge measures correlated statistically significantly with scores on the listening comprehension test. However, when the researcher conducted a hierarchical regression analysis, he found that only scores from the vocabulary knowledge measure significantly predicted performance on the listening comprehension test. Mecartty's (2000) study is valuable, insofar as it is among the few studies reporting the role of syntactic knowledge in L2 listening. However, its results are not free from important limitations.

The major limitation of Mecartty's (2000) study is the way syntactic knowledge was measured, using a sentence-completion multiple-choice task ($K=12$) and a grammaticality judgment task ($K=12$). Mecartty combined scores from the two tasks to create a composite score for correlation and regression analyses. The reliability of the combined test -- still short ($K=24$) -- was relatively low, reported at .66. The researcher

did not specify the target structures used in constructing these measures. Both of the above measures were delivered in written form and were untimed. However, similarly to vocabulary knowledge, componential models of L2 listening (e.g., Buck, 2001) consider knowledge of *spoken* syntax as one of the building blocks of listening ability. Both the componential and the process-oriented accounts of L2 listening (e.g., Goh, 2005) also consider *proceduralized* syntactic knowledge to play an important role in L2 listening. In addition to issues of low reliability, the untimed written nature of the syntactic measures used in Mecartty (2000) makes it unsurprising that no significant findings were found with respect to predictive power. Importantly, Mecartty reported that the measures of vocabulary and syntactic knowledge overlapped in his study, which made it difficult to isolate and produce clear-cut results for the two knowledge sources.

On the role of syntactic knowledge in L2 listening, the only other available study, Liao (2007), had similar limitations. Although Liao used more complex statistical analyses (e.g., SEM and discriminant analysis) and found greater predictive power for syntactic knowledge, measures in that study were also in written form and untimed. Liao (2007) investigated the construct validity of grammar and vocabulary sections and the listening section of the Examination for the Certificate of Competency in English (ECCE), an English proficiency test developed at the University of Michigan. It similarly explored the relationship between vocabulary knowledge, syntactic knowledge and L2 listening ability. SEM results showed that both vocabulary knowledge and syntactic knowledge serve as strong predictors of L2 listening ability. However, SEM results also suggested that the vocabulary knowledge measure produced larger loadings on the listening factor than the syntactic knowledge predictor. In addition, a discriminant

analysis showed that using vocabulary and syntactic knowledge as predictors produced a moderately high percentage of the test takers classified correctly into groups of masters and non-masters in L2 listening. However, these results also showed that vocabulary knowledge contributed more to the discriminant function than syntactic knowledge.

Results from Liao (2007) may well be truly representative of the relative significance of vocabulary and syntactic knowledge in predicting L2 listening comprehension. Results were in line with several hypotheses (e.g., Field, 2008; Vandergrift, 2011), which suggest a more important role for vocabulary than syntactic knowledge. However, the untimed written format of the measure of syntactic knowledge may have obscured the picture.

In summary, very limited empirical evidence is available of the role of syntactic knowledge in L2 listening ability. Results from the only two extant studies (i.e., Mecarty, 2000; Liao, 2007) are limited because of the measurement issues discussed above. More rigorous empirical studies of this topic are needed, including use of more stringent measures and more fine-grained analyses. In order to have a fuller picture of the relative significance of the linguistic factors of L2 listening, the role of non-linguistic factors should be accounted for.

2.7 The Role of Metacognitive Knowledge in L2 Listening

Research on L2 listening ability initially examined the use of strategies by learners during aural comprehension tasks (Rubin, 1994). This led to extensive research on the role of metacognition or metacognitive knowledge in L2 listening and performance. Metacognitive knowledge has been defined as a listener's awareness of the

cognitive processes involved in learning and comprehension. It involves both self-reflection and self-direction, and entails knowledge and awareness in reflection during a learning and comprehension task. Metacognitive knowledge is considered to be the capacity to oversee, regulate and direct attention, thereby avoiding unsuccessful comprehension strategies (Bloomfield et al., 2010; Goh, 2002; Vandergrift, 2003a; Vandergrift & Goh, 2012).

It has been suggested that metacognitive knowledge is among a cluster of variables contributing to variance in L2 listening (Vandergrift, 2006, 2007). Vandergrift, et al. (2006) showed that self-reported metacognitive knowledge explained about 13% of the variance in L2 listening performance in a study of university-level language learners. This finding was confirmed by results from the comprehensive study by Vandergrift & Baker (2015). They accounted for the effect of several L2 listening variables and found a significant effect for metacognitive knowledge.

The relationship of metacognitive knowledge to L2 listening has also been examined by comparing the use of listening strategies among skilled and less skilled listeners. Empirical studies (e.g., Goh, 2000; Rost, 2013; Vandergrift, 2003a), often employing think-aloud measures, have suggested that skilled listeners employ more metacognitive strategies than less skilled listeners. Findings have indicated that skilled groups use approximately twice as many metacognitive strategies (e.g., comprehension monitoring) than less-skilled groups. However, studies also suggested that differences between the two groups of listeners are not limited to *quantity* of metacognitive strategies used (Vandergrift & Baker, 2015).

Skilled listeners appear to employ a repertoire of strategies in regulating listening processes, combining these strategies into a continuous metacognitive cycle (Vandergrift, 2003a; Vandergrift & Baker, 2015). Through a qualitative data analysis obtained from think-aloud protocols, Graham and Macaro (2008) argued that successful L2 listening is attributable to the combined effect of ‘clustering’ a set of different metacognitive strategies. Their results suggested that metacognitive knowledge is a multi-dimensional construct, which should be measured through instruments tapping various sub-constructs.

Vandergrift et al., (2006) considered problem-solving strategies, planning and evaluation, avoiding mental translation, self-knowledge, and directed attention or concentration as five distinct sub-constructs of metacognitive knowledge for L2 listening. Vandergrift et al., (2006) argued that *problem-solving strategies* are strategies that listeners use to make inferences and monitor them. One example of problem-solving is the ability to infer meanings of unknown words based on other clues in a listening passage. *Planning and evaluation strategies* are used by learners to prepare for and evaluate listening performance. *Avoiding mental translation* was defined as the ability *not* to translate a passage in a listening task. *Self-knowledge* is a set of strategies that involve understanding the difficulty of the L2 listening task and being aware of one’s confidence and anxiety level during the task. Finally, *directed attention or concentration* is a set of strategies that listeners use to stay on a task.

After defining the five sub-constructs of metacognitive knowledge for L2 listening, Vandergrift et al., (2006) developed and validated a questionnaire (the Metacognitive Awareness Listening Questionnaire or MALQ) assessing the metacognitive awareness of L2 listeners. They used the University of Ottawa’s Placement

Test (Wesche & Paribakht, 1996) as the listening comprehension measure, and exploratory and confirmatory factor analyses to validate the MALQ. Their results revealed five distinct factors predictive of listening comprehension scores. However, self-knowledge was found to be the sub-construct explaining the most variance. Vandergrift and Baker (2015) also used the MALQ in their study, finding that self-knowledge was the only significant metacognitive knowledge factor in predicting listening scores.

To summarize, results from both qualitative and quantitative studies have revealed a significant role for metacognitive knowledge in L2 listening. In studies examining the relative significance of vocabulary and syntactic knowledge, metacognitive knowledge is one of the main non-linguistic factors that must be accounted for. Including measures of such knowledge will help to clarify the role of linguistic factors in L2 listening.

2.8 The Role of Working Memory in L2 Listening

Another cognitive factor theoretically and empirically recognized to have an important role in L2 listening is working memory. Theories of comprehension postulate that the ability to temporarily store and process information in memory underlies comprehension (e.g., Ericsson & Kintsch, 1995; Townsend & Bever, 2001). Working memory is the cognitive system with oversight of processing, storage, and retrieval of information in memory (e.g., Baddeley, 2001; Baddeley & Hitch, 1974). Early theories of memory system(s) focused only on the *storage* aspect of memory --- short-term memory -- whereas the conceptualization of working memory entails both *storage* and *processing*

of information, which gives working memory a far greater role in complex cognitive activities like language comprehension (Baddeley, 2003).

There have been various theoretical accounts with regard to a definition and structure of WM. Juffs & Harrington (2011) noted that the most widely used and influential in SLA research has been the multi-component model proposed by Baddeley (2000)³. According to this model, working memory is a system composed of four components: two modality-specific input/storage sub-systems (slave systems), the central executive, and the episodic buffer. The two input/storage sub-systems are the phonological loop and the visuo-spatial sketchpad, in which both verbal information and visual/spatial information are processed. The central executive is a capacity-limited attentional control system coordinating the two input /storage sub-systems. The central executive is “the most important but least understood component of working memory” (Baddeley, 2003, p. 835). It plays a central role in focusing attention, dividing it between two important targets or stimulus streams, and switching between tasks. The episodic buffer holds integrated episodes and acts as a buffer store between working memory components, perception and long-term memory (Baddeley, 2012).

More recent working memory models exist (e.g., Conway & Engle 1994; Cowan, 2008; Engle, 2002). Unlike Baddeley’s model, these lack modality-specific input/storage sub-systems. Instead, they emphasize functions and processes of working memory over its structure (Kane, Conway, Hambrick, and Engle, 2007). These more “process-oriented rather than structural models” (Linck et al., 2014, p. 862) emphasize the critical role of attentional control and executive functions. More contemporary models of working

³ This model was originally proposed by Baddeley & Hitch in 1968. In 2000, Baddeley updated the model by adding the episodic buffer.

memory emphasize the role of the central executive in exerting cognitive control over the contents of working memory. According to Miyake and Friedman (2012), working memory executive functions consist of updating, inhibition, and task-switching. These components are responsible for retrieving, holding, and processing temporary information while inhibiting interference of irrelevant stimuli. Engle's (2002) controlled attention view of working memory also considered executive functions to play a key role in efficiently managing available attentional resources. In addition, Engle argued that central executive functions are the primary determiner of individual differences in working memory. This brings an important point into focus in working memory measurement. If central executive functions are the best determiners of variation in individual working memory capacity, in individual differences research, priority should be given to measuring these functions rather than the storage aspect of working memory.

In terms of measurement, working memory tasks measuring the ability to store and rehearse information are called "simple" span tasks. Forward digit span, word span, and non-word span tasks require recall of a string of unrelated letters, words, digits, or visual objects after a brief period of presentation. These tasks only measure the *storage* aspect of working memory. In other words, simple working memory tasks are measures of short term memory rather than working memory. Using Baddeley's (2012) terminology, simple working memory tasks measure the capacity of the two input/storage systems: the phonological loop and the visuo-spatial sketchpad. However, as stated before, Engle (2002) asserted that executive functions are better determiners of individual differences in working memory capacity. Working memory tasks measuring such functions are called "complex" working memory measures (Linck et al., 2014, p. 863).

Complex working memory tasks measure an individual's ability to store information while faced with additional processing tasks. These tasks require active processing of input while simultaneously remembering a string of letters, words, digits, or objects (Linck et al., 2014, p. 863). In line with Engle's (2002) theoretical accounts, meta-analyses of the role of working memory in both L1 (Daneman & Merikle, 1996) and L2 comprehension (Linck et al., 2014) have suggested that complex working memory tasks are better predictors of individual differences than simple working memory tasks. In other words, complex working memory tasks could explain more of the variation between L1 and L2 comprehension tasks than simple working memory tasks.

An extensive and growing body of research exists on the relationship between working memory and L2 reading comprehension (for an overview, see Andringa et al., 2012) and certain aspects of SLA (for an overview, see DeKeyser & Koeth, 2010; Kormos & Sáfár, 2008). However, little research has examined the relationship between working memory and L2 listening comprehension (Vandergrift & Baker, 2015). At this point, it is appropriate to summarize and evaluate them.

McDonald (2006) carried out two experiments, including both non-native (from a variety of L1s) and native speakers of English, seeking to examine the role of working memory in L1 and L2 sentence processing. In this study, participants were asked to complete an aural grammaticality judgment task (GJT), and the working memory measure was presented in the L1 of the native speakers (English). The working memory task was termed a "size judgment task," a complex task involving storage and processing aspects of WM.

In the first experiment, a significant correlation was found between working memory scores and the aural GJT scores in non-native participants. Unsurprisingly, L2 learners were poorer in terms of their accuracy than native speakers, as measured by the GJT. In the second experiment, the working memory of native speakers was measured with a more stressful task which required them to maintain 7-digit numbers in memory. The results revealed that when native speakers had their working memory stressed, they performed similarly to non-native participants. At face value, the results of the two experiments confirmed the significant role of working memory in aural input processing. However, the outcome measure used in McDonald (2006), an aural GJT, was not an authentic listening task. In GJTs, attention is heavily drawn to form, and little comprehension of the meaning of a sentence is required. As such, the spoken syntactic knowledge of participants was measured, rather than their aural comprehension. This limited McDonald's (2006) claims for the role of working memory in L2 listening comprehension.

Brunfeut and Révész (2011) examined the role of working memory and listening anxiety in listening task difficulty. They measured working memory with a visual forward digit span test and a visual backward digit span test. Both measures were simple tasks, tapping working memory storage only. The outcome measure was a listening task, which required participants to listen to a short passage and complete 30 multiple-choice items. Results indicated that neither of the working memory measures predicted listening scores. However, non-significant results such as these in studies examining the predictive power of working memory in L2 listening could be due to the types of measures used. Since both working memory measures were simple, they would be much weaker

predictors of L2 comprehension than complex tasks. Additionally, the listening task in this study was very simple. It was short and involved multiple-choice questions. When cognitive demands of a task are low, it could easily lead to null findings for the role of working memory. In other words, tasks which consume greater cognitive resources would better reveal differences between individuals with different levels of working memory capacity. If tasks are easy, with limited cognitive demand, working memory capacity variation is much less likely to be found (Linck et al., 2014).

Null findings for the role of working memory in Brunfeut and Révész (2011) may have resulted from the simple working memory tasks used in the study. Length of listening passage and type of task (e.g., multiple-choice items) certainly may have a significant influence on the role of working memory. However, results from Wayland et al. (2013) suggest that if complex measures of working memory are employed, the role of working memory in L2 listening comprehension is significant, regardless of the length or density of passages.

Wayland et al. (2013) conducted three related experiments to explore the effects of passage length, information density, and working memory on L2 listening comprehension. They compared the effect of these variables on two question types – multiple-choice and recall. Two visual-spatial *complex* working memory tasks were utilized, the Blockspan and Shapebuilder tasks. Regarding passage length and information density, Wayland et al. found that neither variable impacted performance on *multiple-choice* items. However, results revealed that both the length and information density of a passage affected performance on *recall* items. In terms of the role of working memory, the overall conclusion from this series of experiments was that individuals with

higher working memory capacity understood passages more accurately as measured by both multiple-choice and recall items. Interestingly, the relationship between working memory capacity and performance on multiple-choice comprehension items varied little based on passage length and information density. These results were also reported for recall items, and these findings suggest that effects of length and information density cannot be entirely attributed to increased working memory load (Wayland et al., 2013, p. 42).

To this point, theoretical and empirical evidence indicates that complex working memory tasks are superior measures of individual differences. These results are even more revealing about the role of working memory in complex tasks such as L2 listening comprehension. When simple working memory tasks are used, non-significant results for the role of working memory should be expected. This may explain non-significant findings in Andringa et al. (2012) and Vandergrift & Baker (2015), who reported no role for working memory in L2 listening comprehension.

Andringa et al. (2012) used four digit span tasks (forward and backward visual and forward and backward auditory) and one non-word recognition task as measures of verbal working memory, in order to examine the role of working memory in L2 listening comprehension (among other variables). All working memory measures were simple tasks, tapping working memory storage only. Andringa et al. found very low correlations between working memory and L2 listening ability. Vandergrift & Baker (2015) used a Backward Digit Recall (BDR) task and a Nonword List Recall (NLR) task to measure working memory. They argued that BDR taps both central executive functions and phonological loop capacity, while NLR taps the phonological loop only. Curiously,

despite this argument, the researchers combined the scores of these two tasks to create a composite working memory score. Like Andringa et al., Vandergrift and Baker did not find significant correlations between WM and L2 listening scores.

As Vandergrift & Baker (2015) used a composite working memory score, it is difficult to claim that the null relationship they found for working memory and L2 listening was due to the working memory task type (simple vs. complex). However, results from previous empirical and meta-analysis studies suggest that complex working memory measures can be used more confidently to investigate the role of working memory in complex tasks, such as L2 listening comprehension. All the cited studies used different working memory measures, with analyses based on the measured variables. Moreover, several (e.g., Vandergrift & Baker, 2015) did not report the reliability of the working memory measures, making the reported non-significant relationships difficult to interpret. It is unclear whether the null relationships were due to low reliability of scores – leading to attenuated correlations – or no relationship between the two constructs. In all previous studies (Andringa et al., 2012, notwithstanding), analyses were at measured variable levels, but claims about the role of working memory at the construct level. It appears that the role of working memory in L2 listening comprehension remains unclear, and results of previous studies cannot be interpreted confidently.

2.9 Influence of Anxiety on L2 Listening Performance

In addition to measurement issues, substantial evidence from cognitive psychology has shown that affective factors such as anxiety interfere with executive function and storage aspects of working memory. This means that anxiety can mask

individual differences in working memory capacity, and it is not surprising that empirical studies with no control over anxiety effects do not report a strong role of working memory in L2 listening comprehension ability. Anxiety negatively influences the acquisition, retention and use of language by L2 learners (MacIntyre, 1995; Spielmann & Radnofsky, 2001). There are three kinds of anxiety: trait, state and situation specific. Trait anxiety refers to general personality traits, state anxiety refers to a temporary emotional state, and situation specific anxiety refers to forms of anxiety within a given situation. According to MacIntyre and Gardner (1991), foreign language anxiety (FLA) is a kind of situation specific anxiety, and has been defined as “the feeling of tension and apprehension specifically associated with second language context, including speaking, listening and learning” (MacIntyre & Gardner, 1994, p. 284).

Recently, it has been recognized that FLA has different facets, and while some L2 learners experience FLA in general, others only report feeling anxious while performing in a specific skill such as listening. According to Scarcella and Oxford (1992), listening FLA usually occurs when learners feel they are faced with a difficult and unfamiliar L2 listening task. Listening FLA increases when listeners are under the false impression that in order to complete a listening task, usually in a testing situation, they must understand every single word they hear (Vogely, 1998).

There have been several studies (e.g., Elkhafafi, 2005; Lund, 1991; Vogely, 1998) on the effect of FLA on L2 listening performance, and each reported that FLA impedes L2 listening comprehension. For example, Elkhafafi (2005) investigated the effects of general and listening FLA on Arabic learners’ listening comprehension performance and found significant negative correlations between results from the anxiety

questionnaires and the listening comprehension measure in the study. Although the negative impact of FLA on L2 listening performance seems obvious, an important question regards the exact way in which anxiety exerts its negative influence on cognitive tasks such as listening in the L2. Research findings in cognitive psychology become relevant in this regard.

Anxiety is important within the field of cognitive psychology because it is often associated with adverse effects on the performance of complex cognitive tasks (Eysenck & Calvo, 1992). According to Eysenck, Derakshan, Santos & Calvo (2007), anxiety exerts its negative effect on performance of complex cognitive tasks by preempting the processing and temporary storage capacity of working memory. Eysenck and Calvo (1992) argued that the main effects of anxiety are on central executive or attentional control mechanisms in working memory.

As Eysenck et al. (2007) explained, anxiety interferes with the two main working memory attentional control mechanisms: inhibition and shifting. Miyake et al. (2000) defined *inhibition* as “one’s ability to deliberately inhibit dominant, automatic, or prepotent responses when necessary” (p. 57), and *shifting* as “shifting back and forth between multiple tasks, operations, or mental sets” (p. 55). Eysenck et al. (2007) explained that anxiety redirects the allocation of inhibition and shifting mechanisms to threat-related stimuli, whether internal (e.g., worrisome thoughts) or external (e.g., threatening task-irrelevant distractors). Therefore, performance on complex cognitive tasks demanding working memory inhibition and/or shifting functions become challenging when an individual experiences anxiety. This is because inhibition and/or

shifting functions of working memory are dedicated to dealing with anxiety, a task-irrelevant stimulus (Eysenck et al., 2007).

However, negative effects of anxiety are not limited to working memory attentional control mechanisms. Anxiety also exerts detrimental effects on storage aspects of working memory, and according to Rapee (1993), these negative effects are expected to be more detrimental to storage of aural input in the phonological loop rather than visual input in the visuo-spatial sketchpad. This is because, as Rapee argues, anxiety typically involves inner verbal activity rather than imagery representations, leading to depletion of phonological working memory resources.

Because L2 listening comprehension is a complex linguistic and cognitive task, it seems plausible to assume the same relationship between anxiety, working memory and L2 listening performance. For this reason, anxiety has been included as a variable in the current study.

2.10 Overall Summary

Compared to other language skills, L2 listening has received the least amount of attention by SLA, L2 pedagogy and assessment researchers. Internal factors or components of L2 listening lack empirical foundations. The studies that do exist in this area have been limited in terms of both quantity and quality.

Among internal factors, vocabulary knowledge has been most extensively researched. However, several unanswered questions remain. Extant studies either have not fully measured vocabulary knowledge, or have not measured it in the aural modality. The majority of previous studies have reduced the construct of vocabulary knowledge to

vocabulary size, making the relative significance of depth of vocabulary knowledge remains more or less unknown. The majority of previous studies measured vocabulary knowledge in the written modality, perhaps blurring the role of vocabulary knowledge in L2 listening. Additionally, statistical analyses employed to date lend little confidence in interpretation of results at the construct level. Finally, few studies have examined the role of vocabulary knowledge when the effect of other factors is accounted for. One of the important linguistic factors is syntactic knowledge, which has been nearly ignored in the research to this point.

To the best of our knowledge, only two empirical studies have examined the role of syntactic knowledge in L2 listening. In both studies, declarative and written syntactic knowledge was measured, rather than spoken or proceduralized syntactic knowledge. It is therefore not surprising that syntactic knowledge was found to play a minimal role in L2 listening ability. Limitations regarding statistical analyses in studies of vocabulary knowledge also hold true for research on the role of syntactic knowledge. In the context of L2 reading research, when Shiotsu & Weir (2007) adopted a more rigorous approach in their measurement design and statistical analyses, they reported evidence “for the relative superiority of syntactic knowledge over vocabulary knowledge in predicting performance on a text reading comprehension test” (p. 99). Even in L2 reading research, minimal research into the significant role of syntactic knowledge existed until recently. Only more methodologically rigorous studies with more advanced data analysis methods (e.g., SEM) could reveal the true role of syntactic knowledge in L2 reading. Conclusions by Shiotsu and Weir about the relationship between syntactic knowledge and reading should also be empirically tested in L2 listening.

Two important cognitive factors must be included in any empirical study examining the role of linguistic factors: metacognitive knowledge and working memory. Metacognitive knowledge has been studied extensively, resulting in clear evidence for its central role in L2 listening. However, results of empirical studies on the role of working memory in L2 listening are still inconsistent, with no clear consensus as to its role. The major limitation in many studies investigating the role of working memory in L2 listening is measurement. Either simple measures of working memory have been used exclusively, or only a single complex working memory task. In addition, the effect of working memory may have been masked by the influence of anxiety. For this reason, the impact of anxiety on working memory should be accounted for, if a clearer picture of role of working memory in L2 listening is to be found.

Chapter 3: The Current Study

The purpose of the current study was to investigate the relative significance of two linguistic components of L2 listening: vocabulary knowledge (VK) and syntactic knowledge (SK). The current study was exploratory in nature and adopted a differential or individual differences approach (Cronbach, 1957). In this approach, which implies a correlational research design, variation between individuals in their L2 listening ability and other relevant factors was investigated.

Unlike previous studies, measures of both *breadth* (VB) and *depth* (VD) of VK were included. Previous studies under-represented the construct of VK just by focusing on VB. However, VK is a complex and multidimensional construct, and for this reason, both dimensions of VK were included in the current study. In addition, by including both VB and VD, the distinction between the two dimensions of VK, which is an interesting empirical question, could have been examined.

Additionally, both dimensions of VK were measured in the *spoken* modality, and measures were used in a way to tap *proceduralized* (or fluent) VK. In addition to the investigation of the role of overall VD, the relationship between VB and VD and their relative significance in explaining individual difference in L2 listening comprehension ability was examined.

In terms of SK, *proceduralized spoken* SK was measured. Its relationship with L2 listening was examined, along with its relative significance compared to vocabulary knowledge.

In order to clarify the relative significance of the linguistic components, the effects of two cognitive components in L2 listening was examined – metacognitive knowledge (MK) and working memory (WM).

To obtain a clearer picture of the role of WM, the effect of listening foreign language anxiety was also accounted for. It is important to explore the influence of anxiety on WM, and how this may change the relationship between WM and L2 listening. No previous SLA study, to the best of our knowledge, has examined the influence of anxiety on WM and how this influence affects the relationship between WM and L2 listening ability.

3.1 Research Questions

The current study sought to answer the following research questions. The preliminary research questions were asked to determine whether all the variables in the current study should be included into the main analyses to answer the main research questions.

3.1.1 Preliminary Questions:

1. Are VB, VD, SK, MK, WM and anxiety associated with L2 listening ability?
2. Does MK make a significant contribution to L2 listening ability?
3. Does WM make a significant contribution to L2 listening? Does anxiety influence the effect of WM?
4. Do VB and VD factors make a significant contribution to L2 listening? If so, which one is the better predictor?

5. Should VB and VD be considered distinct predictors of L2 listening in the current study, or should they be combined as a single predictor factor?

3.1.2 Main Questions:

1. Does VK make a significant contribution to L2 listening?
2. Does SK make a significant contribution to L2 listening?
3. What is the relative significance of VK and SK in explaining success in L2 listening ability?

3.2 Expected Results

Given the exploratory nature of the current study and lack of empirical evidence from previous studies, no formal hypotheses were formulated. However, theoretical explanations were combined with limited results of previous empirical studies to make predictions. Linguistic, cognitive and affective factors that explain or influence individual differences in L2 listening ability were considered as predicted results were made.

For the first preliminary research question, it was expected that VB, VD, SK, MK, WM and anxiety factors would all be associated with L2 listening ability. For the second preliminary research question, it was expected that MK would have a significant role in explaining individual differences in L2 listening ability (as previous studies suggest). However, because other variables were included, it was predicted that the role of MK was not as significant as suggested by previous studies including MK as the central predictor variable.

For the third preliminary research question, given theoretical and empirical evidence on the role of WM, it was expected that WM would play a significant role in individual differences in L2 listening ability. However, it was also expected that the significance of this role would become clearer when the influence of anxiety on WM was accounted for.

For the fourth preliminary question, previous studies have reported VB to play the most important role in explaining individual differences in listening ability. However, in the current study it was predicted that VD would have a more significant role. According to theoretical and empirical studies on L2 reading ability, VD plays a more important role in explaining individual differences among L2 learners (Qian & Schedl, 2004). Measures of VD assess how precisely L2 learners know words, in contrast to superficial knowledge of word meanings (or VB). As such, measures of VD may also reveal individual differences in L2 VK more effectively. In addition, it was expected that rather than being separate constructs, VB and VD would be different aspects of the single construct VK.

For the first *main* research question, given theoretical and empirical evidence on the role of VK in L2 listening, a significant role for VK was expected. Previous studies have not found a significant role for SK in individual differences in L2 listening; however, for the second main research question in the present study, it was expected that SK would play a significant role. In both componential and process-oriented models of L2 listening, SK has been assigned a fundamental role. Previous empirical studies have failed to reveal the significant role of SK mostly due to measurement and analysis limitations.

For the third main research question regarding the relative significance of VK and SK, the lack of available empirical research made it difficult to make any predictions. In addition, theoretical models of L2 listening do not explain the relative significance of the two constructs. However, in the current study, proceduralized *spoken* knowledge of syntax was measured. Inferring from evidence from L2 reading research (e.g., Shiotsu & Weir, 2007), it was expected that SK would play a role at least as important as vocabulary knowledge.

Chapter 4: Method

4.1 Participants

Because the research involved human participants, approval for the study was obtained through the University of Maryland Institutional Review Board (Project number: 757226-1). The current study was carried out in an English as a Foreign Language (EFL) context in Iran. Because the study employed complex methods of data analysis, such as SEM and Rasch, a relatively large sample size was required. For SEM, Kass & Tinsley (1979) recommended five to ten participants per indicator variable, up to a maximum of 300 participants. As outlined below, the current study included twenty one indicator variables; therefore, a minimum of 210 participants was required. Originally, 300 participants were recruited; however, data from 263 participants were used for the data analysis. Data from 37 participants were deleted because these participants did not complete several tasks of the study.

These 263 participants were all EFL learners in Iran, enrolled in communicative English classes at several language schools. 162 were female, and 101 were male. The average age of the participants was 27.8 years old, with 19 and 65 as the minimum and maximum age, respectively. The participants' minimum level of education was matriculating as an undergraduate student at a university. None of the participants reported to have had the experience of living, working or studying in an English speaking country.

Although no direct measure of EFL proficiency was used for recruitment purposes, the participants' proficiency level was determined based on placement tests, number of hours of instruction they had received, and textbooks used in classes. To allow

for enough variation in the data, participants enrolled at lower-intermediate to advanced levels of EFL proficiency were recruited. In the current study, the listening sub-section of IELTS was used as the measure of listening ability. The results of the IELTS test revealed that the participants' listening scores fell within the range of 2.5 to 9 (0 and 9 are minimum and maximum scores, respectively), with a median score of 5. These results indicate that there was enough variation in the participants' level, at least in listening ability.

4.2 Instruments

A battery of ten linguistic, cognitive and affective measures was used in the current study. The battery included the following measures: an L2 listening comprehension test, a test of VB, a test of VD, an aural GJT (a SK measure), an aural sentence comprehension task (a SK measure), the Metacognitive Awareness Listening Questionnaire (a MK measure), Blockspan (a WM measure), Shapebuilder (a WM measure), and the two anxiety questionnaires.

4.2.1 Listening Comprehension Test

The listening sub-section of an institutional⁴ version of the IELTS test was used as a measure of EFL listening comprehension ability. The test was comprised of four passages administered in sequential order of difficulty from easiest to most difficult. Passages included recorded conversations, monologues and lectures by a range of English

⁴ The IELTS test used in the current study was a retired official test, which had been validated by the Cambridge University.

native speakers (British, American and Australian). There were ten questions for each passage, with a total of 40 items in the test. Response formats were diverse, including multiple-choice, sentence completion, table completion, sentence matching, fill-in-the-blanks and diagram/picture labeling. Questions tapped into a variety of abilities: understanding main ideas and factual information; understanding the opinions and attitudes of speakers; understanding the purpose of an utterance; and following the development of ideas. Participants listened to each passage once and marked or wrote their answers directly on the question booklet. At the end of the test, participants were given ten minutes to transfer their answers to an answer sheet. The test took a total of about 45 minutes.

4.2.2 Vocabulary Knowledge Measures

In the current study, breadth of vocabulary knowledge was operationalized as the ability to recognize words and match them with their definitions, and depth of vocabulary knowledge was operationalized as the ability to recognize words and match them with their synonyms and collocations. By considering the way these two constructs were operationalized, the following two tests were employed to measure breadth and depth of vocabulary knowledge in the current study.

4.2.2.1 Vocabulary Breadth Test (VBT)

To the best of our knowledge, only three standardized aural measures of VB are known to exist. The first is Fountain and Nation's (2000) lexically graded dictation test. Although the test has displayed high internal reliability ($\alpha > .95$), it is unclear to what

extent the test measures aural VK, because it has not been thoroughly validated.

Additionally, the dictation test format potentially includes confounding variables, such as spelling and orthographic knowledge. For these reasons, this test was not used for the current project.

The second extant test is A-Lex (Milton & Hopkins, 2006), designed to measure aural VB. The test is administered digitally in a yes/no format. Test-takers listen to a list of words and pseudowords, one at a time, and decide whether they heard an actual word or not. Ease of administration is the biggest advantage of A-Lex; however, the test has two weaknesses. First, the target words can be heard an unlimited number of times, an approach that is not representative of the on-line processing required in most authentic listening contexts (Buck, 2001). Second, the degree to which examinees know the meaning of target words is unclear, even if they correctly recognize the real words (Read, 2000). For these reasons, this test was also not used in the current study.

The third test is an *aural* version of the Vocabulary Size Test (VST). The original written VST was developed and validated by Nation and Beglar (2007) and Beglar (2010). Following the original format and development of the VST, McLean, Kramer & Beglar (2015) created a 150-item aural version of the VST and called it the Listening Vocabulary Levels Test (LVL). The LVL uses the same multiple-choice format as the VST, but item stems are presented aurally.

In the written VST, participants see a word and a very limited context in which the word is used. The context is intentionally kept very limited and non-defining, so learners cannot guess the meaning of unknown words from the context. The purpose of the limited context is just to show the most frequent environments for the target word. For

example, the part of speech chosen for the item in the limited context is a reflection of the highest frequency environment for the word (Beglar, 2010).

Answers for test items are chosen based on accuracy of the description or definition of the target word. Here is an example item:

- Miniature:** It is a miniature.
- A. a very small thing of its kind
 - B. an instrument for looking at very small objects
 - C. a very small living creature
 - D. a small line to join letters in handwriting

In order to create the LVLT, McLean et al. (2015) recorded an English native speaker reading the target words and their relevant limited-context sentences. LVLT test-takers hear the words and their relevant limited-context sentences; however, the four options are given in the written modality, with participants asked to mark answers on an answer sheet. Because McLean et al. (2015) developed this test to be used for Japanese EFL learners, they developed the test in Japanese. For this reason the LVLT could not be used in the current study.

Since there was no existing test of VB deemed usable for the current study, a new aural version of the VST was developed. The test used the first eighty items of the written 140-item version of the VST (Beglar, 2010; Nation & Beglar, 2007), followed the development procedure of McLean et al. (2015), and was called the Vocabulary Breadth Test (VBT).⁵ This test consisted of ten items taken from the first to the fourteenth 1000-word families of English. The words included in the VST were based on fourteen 1000-word British National Corpus word lists developed by Nation (2006).⁶ As explained above, the test is in multiple-choice format, and the target word in each item is placed in

⁵ The original written version of the VST is available at:
<http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Vocabulary-Size-Test-14000.pdf>.

⁶ Available at <http://www.victoria.ac.nz/lals/about/staff/paul-nation>.

a short, non-defining context. All four options for each item are suitable in the provided non-defining sentence, and the sentence was chosen in a way to reflect the most frequent context for the target word. To the extent possible, the words in the definitions (choices) were of higher frequency than the word in the stem of the item, although for the higher frequency stem items, this was not always possible.

We adopted the first eighty items of the VST (covering words from the first to the seventh 1000-word families) to develop the new aural version because Nation (2006) asserted that for successful comprehension of English spoken texts, a vocabulary size of 6,000–7,000 word families is required. To ensure full coverage of the test, we included items for the eighth 1000-word family also. Unlike the LVLT and as in the original VST, the multiple-choice answers in the VBT were in English.

A female native speaker of American English was recruited to create the audio portions of the new aural version of the VST. The speaker read the word in each item and its relevant non-defining sentence. Recording took place in a sound-proof recording booth using high quality recording equipment. Then, the recordings were edited using the digital editing program Audacity.

To complete the VBT in the current study, participants listened to the words and non-defining sentences once, and read and marked the multiple-choice answers. There was a ten-second pause between each item, a length of time determined based on a pilot study. The results of the pilot revealed that ten seconds was sufficient time to process aural input, read the multiple-choice questions, and mark answers for each VBT item. Note that this time is twice as long as that reported in McLean et al. (2015), because, unlike the LVLT, the answers in the VBT were not in the native language of the test

takers. However, to promote full understanding, the instructions in the current study were written in the first language of participants (Persian). To reduce the effect of guessing, participants were asked to refrain from random guessing. The total time of the test, including the time for reading the instructions and asking questions, was twenty minutes. See Appendix A for VBT sample items.

4.2.2.2 Vocabulary Depth Test (VDT)

There is no available standardized measure of *spoken* vocabulary knowledge *depth*. Therefore, the Word Association Test (WAT), developed and validated by Read (1993, 1998), was adapted to create an aural test of depth of vocabulary knowledge (VDT). The WAT is a vocabulary measure tapping knowledge of synonym/polysemy and collocation, two important aspects of VD. The test contains 40 items, with each item consisting of one adjective stimulus word and two boxes, each containing four words. Among the four words in the left box, one to three can be synonymous either to one aspect or the full meaning of the stimulus word. Among the four words in the right box, one to three words collocate with the stimulus word. Each item always has four correct choices. However, choices are not evenly spread. There are three possible situations:

- The left and right boxes both contain two correct answers.
- The left box contains one correct choice, while the right box contains three.
- The left box contains three correct answers, while the right box contains one correct choice.

Here is an example item:

To develop the VDT, a female native speaker of American English was recorded reading each stimulus word in the WAT. At the time of test administration in the current study, participants heard each stimulus word once, then saw answer choices on an answer sheet and marked their responses. The time limit for each item was twenty seconds, which had been determined to be a sufficient length of time for processing aural information, reading answer choices, and marking answers. To promote complete understanding, the instructions in the current study were written in Persian. To reduce the effect of guessing, participants were asked to refrain from random guessing. The total time of the test, including the time for reading the instructions and asking questions, was fifteen minutes. See Appendix B for VDT sample items.

4.2.3 Syntactic Knowledge Measures

In L2 reading research, Alderson (1993) argued that for investigating the relationship between SK and reading ability, tests measuring these two different constructs should be as separate as possible. As Urquhart and Weir (1998) recommended, the difference between reading and SK tests can be maximized if SK tests include items with decontextualized sentences or phrases. Along the same line of reasoning, and to measure proceduralized spoken SK in the current study, two timed aural sentence-level SK tasks were developed.

sudden

<input type="checkbox"/> beautiful	<input checked="" type="checkbox"/> quick	<input checked="" type="checkbox"/> surprising	<input type="checkbox"/> thirsty	<input checked="" type="checkbox"/> change	<input type="checkbox"/> doctor	<input checked="" type="checkbox"/> noise	<input type="checkbox"/> school
------------------------------------	---	--	----------------------------------	--	---------------------------------	---	---------------------------------

To operationalize *proceduralized SK*, two criteria proposed by Ellis (2005) were adopted. The first criterion was *available time* for accessing L2 knowledge. Proceduralization of L2 knowledge is a gradual process and the result of large amount of exposure to and practice of L2 knowledge. As L2 knowledge becomes more proceduralized, L2 learners rely less on controlled processing of L2 input and become faster and more fluent in accessing L2 knowledge to comprehend L2 input. For this reason, *timed* linguistic measures are considered measures of proceduralized L2 knowledge. In timed measures, L2 learners have limited time to access their controlled knowledge of language, and they can complete the task successfully only if they possess an acceptable level of proceduralized knowledge of target structures.

The other criterion was *focusing on meaning*, rather than form. Fluent language use entails focusing attention on meaning, and L2 learners with more proceduralized L2 knowledge can focus more on meaning than form in both L2 production and comprehension (Ellis, 2005). For this reason, linguistic measures that draw learners' attention to meaning while measuring their knowledge of forms are considered better measures of proceduralized L2 knowledge. By considering these two criteria, two SK measures (a timed grammaticality judgement task and a timed sentence comprehension task) were developed for the current study. Information on the target structures used in constructing the two tasks and the details of each task is provided below.

4.2.3.1 Target Structures

Two previous studies (i.e., Mecarty, 2000; Liao, 2007) that examined the role of SK in L2 listening focused on morphosyntactic target structures, such as articles, verb endings, nouns, adjectives and adverbs. Testing these target structures may reveal

participants' overall level of morphosyntactic knowledge; however, there may not be a direct relationship between knowledge of these structures and L2 listening comprehension. Inaccurate or incomplete knowledge of the target structures may or may not lead to misinterpretation of aural input, depending on whether an extended context is available.

Therefore, unlike in previous research, the current study focused on syntactic structures that potentially play a more direct role in comprehending the meaning of aural input. Inaccurate processing of these target structures would potentially lead to misinterpretation of the message. The idea for choosing these target structures emerged from two lines of research: (a) cognitive psychology studies regarding factors contributing to aural sentence comprehension (e.g., Ferreira, 2003; Reali & Christiansen, 2007; Robertson & Joanisse, 2010); and (b) corpus studies comparing and contrasting the frequency of occurrence of important syntactic structures in written and spoken corpora (Biber et al., 1998; Roland et al., 2007).

Sentence comprehension studies in cognitive psychology have primarily focused on comparing and contrasting comprehension of canonical and non-canonical sentences. These studies have examined the difficulty of comprehending active and passive sentences, subject and object relative clauses (RC) (Robertson & Joanisse, 2010), conditional clauses (Haigh & Stewart, 2011) and causative sentences (Dittmar et al., 2008). The reason for targeting these structures is their presumed importance in successful aural comprehension.

Corpus studies (both written and spoken) comparing and contrasting variation of key syntactic structures have typically focused on three elements: (1) frequency of

occurrence of different types of relative and conditional clauses; (2) causative sentences; and (3) active and passive sentences. Following these two lines of research, five target structures, with their canonical and non-canonical versions, were chosen for the current study. Table 1 contains the target structures and relevant examples.

Table 1. Syntactic Target Structures Used in the Present Study

Target	Canonical	Noncanonical
Active/passive	The man is pointing at the boy.	The boy is tapped by the girl.
Subject/object RCs	This is the boy that points at the man.	This is the boy that was pointed at by the man.
Passive object RCs		I read the book which was written by a Russian writer.
Hypothetical sentences (types two and three)	If I had money, I would travel.	If they were not tall, they could not play well.
Causative/non-causative	His friend helped him to move.	He got his friend to help him move.

In order to measure automatized spoken knowledge of these syntactic structures, the following two tasks were developed.

4.2.3.2 Aural Grammaticality Judgment Task (GJT)

Previous SLA studies (e.g., Granena, 2013) have demonstrated that aural, timed (played once) GJTs measure proceduralized L2 syntactic and morphosyntactic knowledge. For this reason, an aural timed GJT was developed. In order to avoid confounding SK with VK, the frequency of words used in GJT sentences were checked with corpus data, with only words from the first 1000-word families used. There were four correct and four incorrect sentences for each of the nine target structures. Therefore, the aural GJT had a total of 72 items. Table 2 contains sample items for each of the target structures.

Table 2. Sample Items from the Aural GJT

Target	Canonical	Noncanonical
Active/passive	G: The man is pointing at the boy. UG: The man pointing at the boy.	G: The boy is tapped by the girl. UG: The boy is tapping by the girl.
Subject/object RCs	G: This is the boy that points at the man. UG: This is the boy that he points at the man.	G: This is the boy that was pointed at by the man. UG: This is the boy that he pointed at by the man.
Passive object RCs		G: I read the book which was written by a Russian writer. UG: I read the book which wrote a Russian writer.
Hypothetical sentences (types two and three)	G: If I had money, I would travel. UG: If I had money, I will travel.	G: If I didn't have money, I wouldn't travel. UG: If I didn't have money, I cannot travel.
Causative/non-causative	G: His friend helped him to move. UG: His friend helped him to moving.	G: He got his friend to help him move. UG: He got his friend to help move.

A native speaker of American English was recorded reading the items at a normal speed. At time of test administration, the recording for each item was played once, with participants instructed to mark answers as correct or incorrect on a paper answer sheet. The time limit for each item was eight seconds, a length of time determined by a pilot study. The results of the pilot revealed that eight seconds is sufficient time to process aural input, read the choices, and mark the answer for each item. To promote complete understanding, the instructions in the current study were written in the first language of participants (Persian). To reduce the effect of guessing, participants were asked to refrain from random guessing. The total time of the test, including the time for reading the instructions and asking questions, was 10 minutes. See Appendix C for GJT sample items.

4.2.3.3 Aural Sentence Comprehension Task (SCT)

As another measure of aural proceduralized SK, an aural SCT was developed along the lines of other sentence comprehension studies, using the target structures selected for the current study. In order to avoid confounding syntactic and vocabulary knowledge, the frequency of words used in SCT sentences was compared with corpus data, with words from only the most common 1,000 word families chosen. In this task, participants listened to a sentence once and answered a short yes/no comprehension question. The questions were designed to tap accuracy of comprehension based on the target structures. There were six items for each of the nine structures, for a total of 54 items. Table 3 contains sample items for each of the target structures.

Table 3. Sample Items from the Aural SCT

Target	Canonical	Noncanonical
Active/passive	The doctor pinches the girl. Does the girl pinch?	The man is pointed at by the boy. Did the man point?
Subject/object RCs	This is the girl that pinches the doctor. Does the girl pinch?	This is the man that is pointed at by the boy. Did the boy point?
Passive object RCs		The gift, which was sent by Mary, arrived to Tom's office. Did Tom send the gift?
Hypothetical sentences (types two and three)	If he were tall, he could play well. Can he play well?	If she were not rich, she could not travel. Can she travel?
Causative/non-causative	I always help John with his work. Does John always help him?	John always gets me to help with his work. Does John always help him?

A native speaker of American English was recorded reading the items at a normal speed. At the time of administration, the recording for each item was played once, with participants instructed to read the comprehension questions and mark yes or no on an answer sheet. The time limit for each item was eleven seconds, a length of time determined by a pilot study. The results of the pilot revealed that eleven seconds is

sufficient time to process the aural input, read the comprehension question, and mark the answer for each item. To promote complete understanding, the instructions in the current study were written in the first language of participants (Persian). To reduce the effect of guessing, participants were asked to refrain from random guessing. The total time of the test, including the time for reading the instructions and asking questions, was 10 minutes. See Appendix D for SCT sample items.

4.2.4 Metacognitive Knowledge Measure

MK of L2 listening was measured using the Metacognitive Awareness Listening Questionnaire (MALQ), developed and validated by Vandergrift et al. (2006). The MALQ is a self-report instrument consisting of 21 randomly ordered items tapping MK related to L2 listening comprehension. Items measure perceived use of strategies and processes underlying five factors related to the regulation of L2 listening comprehension: Problem-solving (PS) ($K=6$), Planning and Evaluation (PE) ($K=5$), Mental Translation (MT) ($K=3$), Person Knowledge (PK) ($K=3$) and Directed Attention (DA) ($K=4$). Participants respond to the MALQ using a Likert scale from 1 to 6, with six signifying full agreement with the statement in the item.

In order to avoid confounding L2 proficiency with MK in the current study, the items in the MALQ were translated into Persian. To ensure that the English and Persian versions were parallel forms and measured the same construct, the translation was done by a professional translator, and then his work was checked and revised by two other native speakers of Persian. These two Persian native speakers had an advanced level of English proficiency and a solid familiarity with developing questionnaires for L2

research. They were both PhD students of applied linguistics in two major US universities.

There was no time limit for completing the questionnaire, but, on average, it took about ten minutes for participants to complete this task. See Appendix E for the complete set of English and Persian MALQ items.

4.2.5 Working Memory Measures

As explained earlier, if the role of WM in complex cognitive tasks like L2 listening is to be examined effectively, complex measures tapping both processing and storage aspects of WM need to be employed. For this reason, two complex WM tasks – Blockspan (Atkins, 2011, cited in Atkins et al., 2013) and Shapebuilder (Atkins et al., 2013) – were used in the current study. Both are spatial tasks, requiring participants to simultaneously store and process spatial information. There were five reasons for choosing Blockspan and Shapebuilder: (1) both tasks were used in several previous L2 listening studies (Clark et al., 2013; Nielson, 2014; Wayland et al., 2013), all of which reported significant correlations with L2 listening scores; (2) WM is a domain-general construct – in order to avoid confounding WM with a verbal construct such as L2 proficiency, it was better to use *non-verbal* WM measures; (3) both tasks have been subjected to validation studies and scores of them correlate highly with one another, as well as those on other complex WM tasks (Atkins et al., 2009); (4) both have yielded reliable results in previous studies; and (5) both are short and easy to administer.

4.2.5.1 The Blockspan Task (WM 1)

The following description has been adapted from Clark et al. (2013). In the Blockspan task, participants are shown a 4 x 4 series of squares and asked to remember the serial order in which a sequence of yellow blocks appeared on the grid (see Figure 1 for an example). Each block within a sequence flashes for one second in one cell on the 4 x 4 grid. Trials are segmented into sets by the appearance of a black square mask that covers the entire grid for one second. After viewing a series of locations flash in a given trial, participants are asked to recall the locations of flashing squares in the order they were presented by clicking on the squares in the same sequence.

Participants complete 16 trials within each trial block. For the first trial, there is one set with two stimuli. For the next trial, one set is presented with three stimuli, the next with four, followed by five.

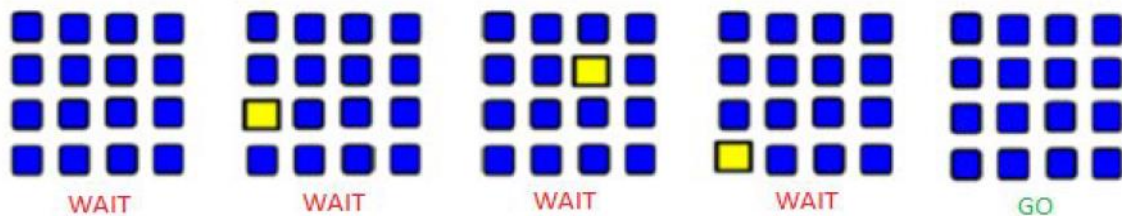


Figure 1. Sample sequence for Blockspan task

Following initial presentations, trials are made more difficult by including *two* sets of two, three, four, or five stimuli. Then, *three* sets of stimuli are presented. Finally, for the final four trials, participants view *four* sets of two, three, four, or five stimuli. The dependent variable for this task is a total score. Participants receive 10 points for the first item correctly recalled, 20 for the second consecutive item correctly recalled, 30 for the third, and so on. Each additional item in a series correctly recalled is worth 10 more

points than the previous item. If an item in the series was forgotten, scoring begins again at 10 for the next item in the sequence correctly recalled. This task takes fifteen minutes.

4.2.5.2 The Shapebuilder Task (WM 2)

The Shapebuilder description has also been adapted from Clark et al. (2013). Shapebuilder is similar to Blockspan, but involves the tracking of additional information. Participants are asked to remember the order and spatial position in which a series of colored shapes are presented. Participants see a 4 x 4 grid (see Figure 2 below), with a sequence of two to four colored shapes appearing sequentially in one of 16 possible locations. Participants are asked to remember the location, shape, and color of each item, as well as the order in which items appeared. After the final item of a trial is presented, participants are asked to recreate the sequence by clicking on the correct colored shape and dragging it to the appropriate location.

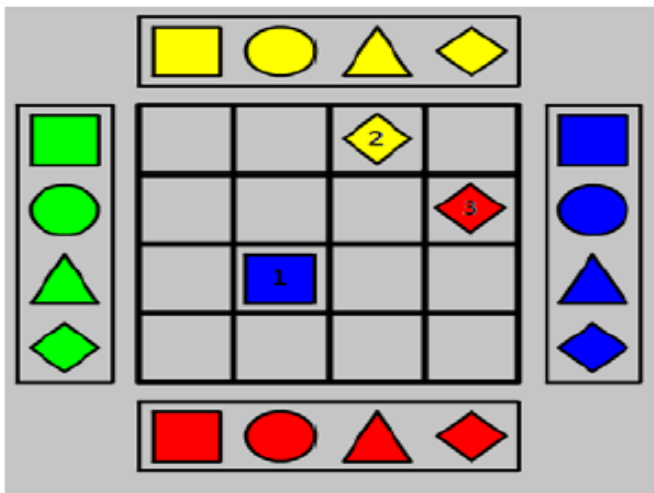


Figure 2. Example of the Shapebuilder task

This task increases in difficulty in two ways. First, trial length begins at two, then moves to three, then four. Within each set of trials of a given length, they become more difficult by including more diverse stimuli of different colors/shapes. At the easiest level, items are all the same shape or color, and at the most difficult level, items are all different colors and shapes. Participants see points awarded for each item immediately after releasing the mouse button.

Points are awarded as follows: participants receive 15 points for the first item correctly recalled, and an additional 15 points for every consecutive item correctly recalled in the sequence. Shapebuilder also awards points for partially recalled items; partial credit is awarded when the correct location is guessed, such that the participant earns 5 points for the correctly recalled color but not shape, and 10 points for the correctly recalled shape but not color. Every time an item is missed, scoring begins again at 15 for the next correctly recalled item. The task takes fifteen minutes.

Both Blockspan and Shapebuilder WM tasks were administered on personal computers connected to high speed internet on an individual basis. Participants received the instructions in their native language (Persian), and completed the tasks within the time limit determined by the computer.

4.2.6 Anxiety Questionnaires (AX 1 and AX 2)

For measuring L2 listening FLA, the current study employed two questionnaires. These measures were modified versions of the two questionnaires used in two previous studies on the role of FLA on L2 listening performance. The first one (AX 1) was a nine-item questionnaire adopted from Mills, Pajares and Herron (2006). However, as the

original study was conducted in the context of learning French as a foreign language, the word “French” was replaced with “English” in the item sentences.

The other questionnaire (AX 2) had eighteen items and was adopted from Elkhafaifi (2005). As the original study was conducted in the context of learning Arabic as a foreign language, the word “Arabic” was replaced with “English” in the item sentences. Also, the original questionnaire had twenty items, but two of the items were about Arabic “culture”, which were irrelevant to the purpose of the current study. Therefore, these two items were deleted.

In order to avoid confounding L2 proficiency with the ability to complete the anxiety questionnaires, in the current study both of the questionnaires were translated into Persian. To ensure that the English and Persian versions were parallel forms and measured the same construct, the translation was done by a professional translator, and then his work was checked and revised by two other native speakers of Persian. These two Persian native speakers had an advanced level of English proficiency and a solid familiarity with developing questionnaires for L2 research. They were both PhD students of applied linguistics in two major US universities.

There was no time limit for completion, but on average it took about ten minutes for participants to complete both tasks.

The primary reason for using two anxiety questionnaires was that in the SEM analysis, there should be at least two measured variables for each latent variable. See Appendix F for the English and Persian versions of the AX 1 and AX 2.

4.2.7 Background Questionnaire

A background questionnaire was administered before the experimental tasks. It asked participants about their age, gender, occupation, education/major, starting age for learning English, amount of classroom instruction, and length of residence in an English-speaking country (if applicable).

4.3 Procedure

Directors, administrators and teachers at two main language schools in Iran assisted with data collection for the current project. Invitation letters for participation in the study were sent to 657 EFL learners, all of whom were enrolled in lower-intermediate to advanced level EFL classes. The following website was used for participant registration: <https://evand.ir/>. 478 participants registered to take part in the study; however, due to time, space and budget limitations, only the first 300 were recruited.

Administration of the measures was conducted during one group and one individual session, with both sessions on the same day. A room equipped with high quality audio devices, used for official IELTS test administration, was used for group-level administration of the measures. Administrations of group measures were limited to ten participants at a time. Group-level administration took about 130 minutes, and included all measures except for the WM tasks. There were two ten-minute breaks during the group administration of the measures. A lab equipped with ten computers was used for the individual-level administration of the WM tasks. This session lasted for about 30 minutes for each participant. Individual sessions were arranged after a twenty-minute break following the group-level administrations. Table 4 below summarizes the timing and order of administration of the measures.

Table 4. Timing and Order of Task Administration

Group Session		
Order	Task	Time (in minutes)
1	Consent Form & Background Questionnaire	10
2	IELTS listening test	45
3	MK Questionnaire	10
4	AX Questionnaires	10
Break 1		10
5	VBT	20
6	VDT	15
Break 2		10
7	Aural SCT	10
8	Aural GJT	10
Break 3		20
Individual Session		
9	Blockspan (WM 1)	15
10	Shapebuilder (WM 2)	15
Total Time		200

4.4 Data Analysis

All linguistic measures were scored dichotomously, with data recorded as zero and one. For the IELTS listening measure ($K= 40$), scores were recorded in two ways: the whole test (including all 40 items), and the four separate sub-sections (coded IELTS 1-4) with ten items in each section. For the VBT, scores were recorded both for the whole test ($K= 80$), and for its four sub-sections labeled VBT 1 to VBT 4. Each sub-section included scores from paired 1000-word family bands ($K= 20$). For the VDT, scores were recorded for the whole test ($K= 157$), as well as for its two sub-sections VDT 1 and VDT2. VDT 1 ($K= 70$) included scores from synonym/polysemy items, and VDT 2 ($K= 87$) included scores from collocation items. In addition, scores from VBT and VDT measures were combined to create a composite score for the construct of VK ($K= 237$).

For SK measures, scores from the GJT ($K= 72$) and SCT ($K= 54$) were first recorded separately. Then, scores from both measures were combined to create a composite score for the construct of SK ($K= 126$).

Responses to the MK questionnaire were recorded as numbers ranging from one to six for all items ($K= 21$), as well as sub-sections. Subsections for the MK measure were as follows: Planning and Evaluation (PE) ($K=5$), Directed Attention (DA) ($K=4$), Person Knowledge (PK) ($K= 3$), Mental Translation (MT) ($K=3$), and Problem-solving (PS) ($K=6$). The MK questionnaire had several negatively worded items, so responses were inverted in order to have positive correlations with the listening ability measures.

Responses to the two anxiety questionnaires, AX 1 ($K= 9$) and AX 2 ($K= 18$), were recorded as numbers in a range of zero to seven and one to five, respectively. The scores from the two questionnaires were also combined to create a composite score for the construct of anxiety ($K= 27$). Responses to items on the anxiety questionnaires were inverted in order to have positive correlations with the WM and listening ability measures.

Total scores from the two WM measures, Shapebuilder (WM 1) and Blockspan (WM 2), were automatically calculated by the computer. Scores from these two measures were also combined to create a composite score for the construct of WM.

Before conducting the main analyses on the data set to answer the research questions, a set of preliminary analyses were conducted. First, classic test theory and Rasch item and reliability analyses were conducted. According to the results of the item analysis, misfitting items were deleted, and reliability of the measures then estimated.

Rasch person ability logits for all measures except WM tasks were generated for the subsequent analyses. Descriptive statistics were also computed, and assumptions of univariate and multivariate normality checked. In addition, outliers in the data were detected and addressed.

To compare the results of the current study with several previous studies, a set of zero-order correlation and multiple regression analyses were conducted. Afterwards, to examine associations among latent factors in the current study, a set of confirmatory factor analyses (CFA) were conducted. CFA was also used to examine the distinctness of VB and VD factors. Finally, a set of SEM analyses were conducted to answer the main research questions of the current study. Table 5 provides an overview of the measures used in the current study. SPSS (version 20), WINSTEPS (version 3.92.0) and LISREL (version 9.1) were used for these analyses.

Table 5. Summary of Measures Used in the Present Study

Task	K
L2 Listening Ability Measures (Lis)	
IELTS 1	10
IELTS 2	10
IELTS 3	10
IELTS 4	10
IELTS Total	40
Vocabulary Breadth Test (VBT)	
VBT 1	20
VBT 2	20
VBT 3	20
VBT 4	20
VBT Total	80
Vocabulary Depth Test (VDT)	
VDT 1 (Synonym/ polysemy)	70
VDT 2 (Collocation)	87
VDT Total	157
Vocabulary Knowledge (VK) Measure	
VK (VBT + VDT)	237
Syntactic Knowledge (SK) Measures	
SCT	72
GJT	54
SK Total	126
Metacognitive Knowledge (MK) Measures	
MK_PE	5
MK_DA	4
MK_PK	3
MK_MT	3
MK_PS	6
MK Total	21
Anxiety (AX) Measures	
AX 1	9
AX 2	18
AX Total	27
Working Memory (WM) Measures	
WM 1 (Shapebuilder)	
WM 2 (Blockspan)	

Chapter 5: Results

5.1 Item and Reliability Analysis

First, classic test theory and Rasch item and reliability analyses were conducted. The classic test theory item analysis revealed that none of the items in the measures had a zero or negative item discrimination index (item-total correlation). However, Rasch item analysis showed that the infit mean-square of several items in the battery was outside the acceptable range of .75 to 1.3 (Bond & Fox, 2013). Therefore, before reliability analyses, misfitting items were deleted. The number of deleted items from each measure was as follows: one from IELTS 4, VBT 2, GJT, SCT, MK_DA, MK_PE and MK_PK; two from VBT 4 and VDT 2; three from AX1 and AX2, and five from VBT 3.

After deleting the misfitting items, another round of Rasch item analysis revealed that no more items had an infit mean-square beyond the acceptable range of .75 to 1.3. Additionally, the infit mean-square of Rasch person ability logits was evaluated to detect misfitting persons (cases). By considering the acceptable range of .75 to 1.3, several misfitting cases were found in the data set. However, for all measures, the number of misfitting cases was less than five percent of all participants. The maximum number of misfitting cases was found in VBT4 with twelve misfitting cases. As the number of misfitting cases for none of the measures exceeded five percent of the total number of cases, and because a few misfitting cases in a data set should not be a point of concern (i.e., they have negligible impact on anything else) (Wright, Linacre, Gustafson, & Martin-Lof, 1994), it was decided to keep data from all participants for the subsequent analyses.

Then, the reliability of all measures (except WM) was estimated using Cronbach's Alpha (α) and Rasch. For the WM measures, following Atkins et al. (2014), Spearman-Brown split-half reliability estimates were computed. Scores from items in even-numbered trials were correlated with the ones in the odd-numbered trials for both WM tasks. Table 6 summarizes reliability estimates for each separate sub-section of the measures, as well as for the complete set of items in each measure.

As seen in Table 6, reliability estimates for the complete set of items in each measure, in comparison to their sub-sections, were higher. The sub-sections had smaller reliability estimates because of the smaller number of items. Reliability estimates of the complete sets of items for all measures, except for the MK questionnaire and WM 1 task, were all above .9. The MK questionnaire and WM 1 (Shapebuilder) had reliability estimates of .72 and .73, respectively. However, the reliability estimate for Shapebuilder in the current study was still higher than the ones in previous studies. For example, Sprenger et al. (2013) reported the reliability estimate of .63 for Shapebuilder in their study. The higher reliability estimate for Shapebuilder in the current study can be due to a larger sample size.

Table 6. Reliability Estimates

Task	K	α	Rasch Reliability
IELTS 1	10	.75	.63
IELTS 2	10	.76	.68
IELTS 3	10	.72	.62
IELTS 4	9	.77	.68
IELTS Total	39	.90	.89
VBT 1	20	.67	.62
VBT 2	19	.77	.75
VBT 3	15	.84	.76
VBT 4	18	.78	.73
VBT Total	72	.92	.92
VDT 1	70	.92	.91
VDT 2	85	.93	.92
VDT Total	155	.97	.96
VK	227	.98	.97
SCT	53	.91	.85
GJT	71	.90	.91
SK Total	124	.95	.94
MK_PE	4	.77	.68
MK_DA	3	.77	.68
MK_PK	2	.84	.77
MK_MT	3	.61	.62
MK_PS	6	.68	.69
MK Total	18	.75	.72
AX 1	6	.91	.85
AX 2	15	.93	.92
AX Total	21	.95	.94
		Split-half Reliability	
WM 1 (Shapebuilder)		.73	
WM 2 (Blockspan)		.81	

Although measures such as MK_MT and MK_PS had low reliability estimates (likely because of the small number of their items), the rest of the measures had an acceptable level of reliability. However, as the main method of data analysis in the

current study was SEM, a method that partials out the negative effect of measurement error, the low reliability of some of the measures should not be a point of concern.

5.2 Descriptive Statistics

After item and reliability analyses, as well as deleting misfitting items, Rasch person ability logits for all measures except WM were generated. These would then be used as data for subsequent analyses. For WM measures, raw total scores generated by the computer were used. First, descriptive statistics were computed. As can be seen in Table 7, skewness and kurtosis of all measures (except kurtosis for GJT) were within the acceptable range of +/- 3 (Bachman, 2004; Tabachnick & Fidell, 2007). In addition to examining skewness and kurtosis, Kolmogorov–Smirnov and Shapiro–Wilk statistical tests were conducted to examine whether the assumption of univariate normality of the data was met. These tests, along with a visual examination of histograms, revealed that this assumption was indeed met.

Multivariate normality was also assessed, and results of the multivariate skewness and kurtosis tests revealed that the assumption was not met (Skewness: z -score = 22.02, p -value = .000; Kurtosis: z -score = 7.28, p -value = .000; Skewness and Kurtosis: Chi-square = 537.9, p -value = .000). For this reason, to fit correlational and structural equation models to the data in the CFA and SEM analyses, the Robust Maximum Likelihood (RML) method was used.

Table 7. Descriptive Statistics

	Min	Max	Mean	SD	Variance	Skewness	Kurtosis
IELTS 1	-5	3	.11	1.71	2.91	-.18	-.07
IELTS 2	-4	4	.04	1.84	3.4	.1	.07
IELTS 3	-3	4	.15	1.56	2.44	.36	.15
IELTS 4	-4	4	.21	1.98	3.91	.33	-.19
IELTS Total	-2.57	5.34	.061	1.42	2.02	1.12	1.79
VBT 1	-4	4	.08	1.15	1.33	.66	2.09
VBT 2	-5	5	.08	1.48	2.19	.13	1.03
VBT 3	-4	5	.02	1.81	3.27	.67	.68
VBT 4	-6	3	-.19	1.82	3.33	-.8	.87
VBT Total	-2.81	3.74	0	1.18	1.39	.71	.39
VDT 1	-3	4	-.02	1.11	1.22	.56	.87
VDT 2	-4	4	-.05	1.07	1.14	.51	.88
VDT Total	-3.19	3.86	0	1.01	1.02	.74	1.25
VK	-3	4	0	1.01	1.02	.8	1.08
SCT	-3	4	.09	1.38	1.91	.56	.48
GJT	-3	4	0	1.03	1.05	1.3	3.92
SK Total	-2.62	3.79	0	1.03	1.06	.94	1.63
MK_PE	-4	4	-.08	.97	.94	-.82	4
MK_DA	-5	4	.09	1.53	2.34	.33	.89
MK_PK	-4	5	-.42	2.87	8.21	.24	-.85
MK_MT	-5	4	.62	1.93	3.72	-.02	-.23
MK_PS	-2	4	.03	.97	.95	.89	2.26
MK Total	-.8	1.23	0	.38	.14	.61	.38
AX 1	-5	4	.06	1.43	2.03	.12	1.52
AX 2	-5	5	0	1.32	1.73	-.14	2.3
AX Total	-2.74	1.62	0	.89	.8	-.72	.38
WM 1 ⁷	1	27	12.05	4.03	16.27	.24	.9
WM 2	2	28	10.79	3.86	14.89	.82	2.09
WM Total	6.65	43.6	22.72	6.59	43.47	.02	.33

⁷ In order to make the metric of the two WM tasks comparable to the rest of the tasks, scores from these two tasks were divided by 100. This change of metric was necessary for the SEM analysis. SEM programs such as Lisrel fail to complete the parameter estimation, if there is a large discrepancy between the metric of measured variables.

Univariate and multivariate outliers were detected by assessing z scores and Mahalanobis distance, respectively. There were no multivariate outliers in the data set. Only three cases (participants) in the Shapebuilder data set were detected as univariate outliers. The Shapebuilder scores for these three cases were set as missing data. Missing values were then imputed in SPSS by means of the full information maximum likelihood estimation procedure, and mean scores per case were calculated. Descriptive statistics for Shapebuilder were then calculated after the data imputation. There were no missing values in the data.

5.3 Correlation Analysis

In order to explore the relationships between the measured variables of the current study, Pearson Product-Moment Correlation coefficients were computed. As seen in Table 8, all variables except for six (highlighted in the table) had statistically significant correlations with the four measures of listening ability. The PS and MT sub-sections of the MK questionnaire did not correlate significantly with two and four measures of listening ability, respectively. Additionally, there were several other non-significant correlations among other measured variables.

In addition, the correlations among the total scores of measured variables were computed. As seen in Table 9, all measures had a significant correlation with the listening ability score.

Table 8. Correlations among Separate Measured Variables

Task	I1	I2	I3	I4	B1	B2	B3	B4	D1	D2	SCT	GJT	MT	DA	PK	PE	PS	WM1	WM2	AX1	AX2	
I1	1																					
I2	.52**	1																				
I3	.6**	.6**	1																			
I4	.56**	.61**	.65**	1																		
B1	.53**	.54**	.61**	.6**	1																	
B2	.45**	.46**	.58**	.56**	.6**	1																
B3	.52**	.53**	.64**	.69**	.63**	.73**	1															
B4	.39**	.43**	.5**	.51**	.48**	.57**	.59**	1														
D1	.56**	.56**	.66**	.68**	.64**	.7**	.74**	.71**	1													
D2	.58**	.57**	.63**	.67**	.65**	.71**	.77**	.74**	.86**	1												
SCT	.62**	.61**	.61**	.64**	.7**	.56**	.64**	.5**	.59**	.61**	1											
GJT	.55**	.52**	.61**	.68**	.64**	.58**	.72**	.66**	.73**	.72**	.73**	1										
MT	.1	.04	.07	0	.07	.07	.11	.14*	.03	.11	.1	.06	1									
DA	.15**	.17**	.13*	.21**	.11	.12	.17**	.15*	.15*	.15*	.18**	.19**	.32**	1								
PK	.42**	.42**	.46**	.41**	.43**	.35**	.5**	.44**	.35**	.42**	.49**	.48**	.27**	.19**	1							
PE	.36**	.28**	.33**	.37**	.35**	.29**	.4**	.36**	.38**	.4**	.41**	.47**	-.03	-.02	.39**	1						
PS	.08	.16**	.12*	.11	.08	.18**	.19**	.14*	.13*	.13*	.11	.15*	.36**	.51**	.14*	-.17*	1					
WM1	.27**	.23**	.17**	.24**	.05	.05	.16*	.05	.21**	.17**	.26**	.17**	.14*	.03	.19**	.14*	-.03	1				
WM2	.28**	.29**	.15*	.22**	.03	.03	.15*	.06	.19**	.18**	.28**	.19**	.07	.07	.22**	.16*	.03	.73**	1			
AX1	.4**	.41**	.4**	.43**	.36**	.34**	.44**	.38**	.38**	.39**	.42**	.47**	.15*	.35**	.51**	.33**	.29**	.16*	.17**	1		
AX2	.43**	.45**	.43**	.43**	.35**	.33**	.41**	.4**	.37**	.42**	.46**	.45**	.15*	.42**	.57**	.38**	.25**	.18**	.2**	.82**	1	

** . Correlation is significant at the 0.01 level (2-tailed). * . Correlation is significant at the 0.05 level (2-tailed).

Table 9. Correlations among Composite Scores of the Measured Variables

Task	IELTS Total	VBT Total	VDT Total	SK	MK	WM	AX
IELTS Total	1						
VBT Total	.78**	1					
VDT Total	.8**	.87**	1				
SK	.77**	.78**	.76**	1			
MK	.37**	.44**	.36**	.42**	1		
WM	.31**	.14*	.2**	.25**	.14*	1	
AX	.52**	.47**	.43**	.51**	.56**	.19**	1

** . Correlation is significant at the 0.01 level (2-tailed). * . Correlation is significant at the 0.05 level (2-tailed).

5.4 Regression Analysis

It is important to make the results of the current study comparable to those from several major previous studies on the role of internal factors of L2 listening (e.g., Mecartty, 2000; Stæhr, 2009; Vandergrift & Baker, 2015). Therefore, prior to the SEM analysis, a set of regression analyses was conducted. Rasch person ability logits for the total set of listening test items were used as the dependent variable, and Rasch person ability logits from the other measures (excluding WM measures) were used as independent variables. For the WM measures, the composite score from the two tasks was used. The results of a backward stepwise multiple regression, summarized in Table 10, revealed that all variables except for MK explained a statistically significant amount of variance in the listening scores. However, evaluation of the variance inflation factor (VIF) index revealed that it was possible that the assumption of multicollinearity was not met. VIFs for VBT, VDT and SK were above 2.5, and according to Allison (1999), VIFs above 2.5 should be considered as a sign of concern for multicollinearity.

Table 10. Multiple Regression Results for all Independent Variables

Variable	B	SE (B)	β	<i>p</i>	VIF
VBT	.26	.09	.22	.003	5.21
VDT	.45	.1	.34	.000	4.64
SK	.37	.08	.27	.000	2.96
WM	.03	.01	.12	.001	1.55
MK	-.21	.15	-.06	.167	1.12
AX	.24	.07	.15	.000	1.69

To improve the regression model and address the issue of multicollinearity, results of the two VBT and VDT measures were combined, and the Rasch person ability logits for a new independent variable, labeled VK, was generated. Then, a new regression

model with the listening test scores as the dependent variable and VK, SK, WM, MK and AX as independent variables was tested. In this new model, MK was still not a significant variable of L2 listening ability.

For this reason, MK was deleted from the regression model to examine whether the remaining variables remained significant. As can be seen in Table 11, the rest of the independent variables remained statistically significant variables of L2 listening ability after MK was deleted. In the new model, VIFs for VK and SK were slightly above 2.5, again suggesting high correlations between these two independent variables. However, 2.5 is a strict threshold, and many statistical resources (e.g., Lomax & Hahs-Vaughn, 2012) consider VIF above 10 as a sign of severe multicollinearity.

Table 11. Multiple Regression Results without MK

Variable	B	SE (B)	β	<i>p</i>	VIF
VK	.74	.07	.52	.00	2.6
SK	.37	.08	.27	.00	2.9
WM	.03	0	.11	.00	1.08
AX	.21	.06	.13	.00	1.37

Comparing R squares between the two included models (i.e., one with MK, and one without MK) showed that deletion of MK from the regression model did not impact the amount of variance in the listening scores explained by the independent variables. In Table 12, Model 1 refers to the one with all variables, and Model 2 refers to the one without MK. Both models explained 74% of the variance in the listening test scores.

Table 12. Regression Models Comparison

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.86	.74	.73	.73
2	.86	.74	.73	.74

In order to explore whether MK by itself was a significant variable of listening ability, a simple regression model with MK as the sole independent variable was tested. Table 13 shows that MK was a significant variable. Also, Table 13 shows that 14% of the listening test scores variance was explained by MK.

Table 13. Simple Regression Results for MK

Variable	B	SE (B)	β	<i>p</i>	R Square	Std. Error of the Estimate
MK	1.41	.22	.37	.000	.14	1.32

5.5 Confirmatory Factor Analysis (CFA)

CFA is a special case of the general family of SEM, and it can be used to determine whether the measured scores (observed variables) group together under predetermined latent variables or factors. CFA provides measurement regression weights (loadings) for each observed variable, which show the strength of association between the observed variables and the latent factor. CFA also tests whether these weights are significantly different from zero. Additionally, CFA can be used to examine the magnitude of associations among factors. These factors should reflect theoretically motivated constructs. CFA can also be used to examine the extent to which the latent factors are distinct from each other.

In the current study, to evaluate and compare the CFA models, a profile of model fit tests and indices recommended by Hu & Bentler (1999) and Mueller & Hancock (2008) was used. Chi square, with its degrees of freedom and *p*-value, was checked. For a good model fit, the chi-square should not be statistically significant at a .05 level. However, in large samples and complex models, a chi-square is usually significant and not very informative. For this reason, descriptive measures or model fit indices are

preferred. In the current study, the following model fit indices and criteria to evaluate them were used: the standardized root mean square residual (SRMR < .08), the root mean square error of approximation (RMSEA < .06), the comparative fit index (CFI > .95), the normal fit index (NFI > .90), the non-normed fit index (NNFI > .95), and the goodness-of-fit statistic (GFI > .90). Also, in order to compare the fit of nested models, in addition to comparing model fit indices, a formal chi-square difference test ($\Delta \chi^2_{(df_1 - df_2)} = \chi^2_{df_1} - \chi^2_{df_2}$) was computed.

To examine associations among latent factors and to assess whether tasks in the current study (i.e., tests and questionnaires) measured separate latent factors, a set of first-order CFAs were performed.

First, through testing CFA Model 1, associations among latent factors were examined. In Model 1, the latent factors of VB, VD, SK, MK, WM and AX were correlated with each other as well as with the latent factor of L2 listening ability. Table 14, summarizes the fit indices for Model 1.

Table 14. Summary of Fit Indices for CFA Model 1

Index	CFI	NFI	NNFI	GFI	RMSEA	SRMR	Chi-square
Criterion	≥ .95	≥ .90	≥ .95	≥ .90	≤ .06	≤ .08	None significant
Model 1	.99	.97	.98	.91	.06	.05	* $\chi^2 = 425.9, df = 168$

Except for the significant chi-square, the rest of model fit indices showed that CFA Model 1 fit the data well. In this model, the correlations among all the latent factors were statistically significant. Table 15 summarizes these correlation coefficients.

Table 15. Correlation Coefficients among Latent Variables

Task	Listening	VB	VD	SK	MK	WM	AX
Listening	1						
VB	.9	1					
VD	.88	.95	1				
SK	.91	.9	.84	1			
MK	.71	.69	.52	.77	1		
WM	.33	.16	.21	.3	.32	1	
AX	.59	.5	.46	.56	.85	.22	1

Because of the high magnitude of correlations among the linguistic factors (i.e., VB, VD and SK), several separate CFA models were tested to examine how distinct these factors were. First, to examine whether the two factors of VB and VD were distinct, two alternative CFA models were tested. In CFA Model 2, which was a two-factor model, the two factors of VB and VD were separate but correlated. In CFA Model 3, which was a single-factor model, all six measured variables of VB and VD were loaded on a single factor. Table 16 summarizes fit indices for these two models.

Table 16. Summary of Fit Indices for CFA Models 2 and 3

Index	CFI	NFI	NNFI	GFI	RMSEA	SRMR	Chi-square
Criterion	≥ .95	≥ .90	≥ .95	≥ .90	≤ .06	≤ .08	None significant
Model 2	1	1	1	.99	0	0	$\chi^2 = 4.64, df = 8$
Model 3	.99	.99	.99	.97	.08	.02	$*\chi^2 = 22.8, df = 9$

Models 2 and 3 both had acceptable fit indices suggesting that they fit the data well. However, the chi-square for Model 2 was smaller and not statistically significant, while it was larger and statistically significant for Model 3. This suggests that Model 2 fit the data better than Model 3. In addition, the chi-square difference test between Models 2 and 3 ($\Delta\chi^2 = 18.16, df = 1, p\text{-value} = .000$) was statistically significant, which once again shows that Model 2 fit the data better than Model 3.

The better fit of Model 2 may suggest that the two factors of VB and VD were measured distinctly enough in the current study; however, the high correlation between two factors in Model 2 was .94 and statistically significant. This weakens the evidence for the distinctness of the two constructs. However, the statistical plausibility of Model 2, and its superiority in comparison with Model 3 suggests that the two factors of VB and VD can be entered in subsequent SEM models as two distinct predictor factors of the latent L2 listening factor. However, due to the strong correlation between VB and VD, the SEM analysis may reveal that one of these factors is redundant and cannot significantly predict success in L2 listening.

To examine the extent to which factors of VB and VD were distinct from the SK factor, three alternative first-order CFA models were tested. In Model 4, there were three separate but correlated factors of VB, VD and SK. In Model 5, all the measured variables of vocabulary knowledge, i.e., breadth and depth, were loaded on one single factor, and this factor correlated with the SK factor. In Model 6, all the measured variables of VB, VD and SK were loaded on a single factor. Table 17, summarizes fit indices for these three models.

Table 17. Summary of Fit Indices for CFA Models 4, 5 and 6

Index	CFI	NFI	NNFI	GFI	RMSEA	SRMR	Chi-square
Criterion	≥ .95	≥ .90	≥ .95	≥ .90	≤ .06	≤ .08	None significant
Model 4	.99	.99	.99	.96	.07	.02	* $\chi^2 = 42.73$, $df = 17$
Model 5	.98	.98	.98	.94	.09	.03	* $\chi^2 = 62.06$, $df = 19$
Model 6	.97	.97	.96	.91	1.3	.04	* $\chi^2 = 105.8$, $df = 20$

In terms of model fit indices, Models 4, 5 and 6 fit the data well. However, Model 4 had the smallest chi-square, and as shown in Table 18, the chi-square difference tests

comparing Model 4 with Models 5 and 6 revealed the superiority of Model 4. The statistical plausibility of Model 4, and its superiority to Models 5 and 6, suggested that three factors (VB, VD and SK) can be entered in subsequent SEM models as three distinct predictor latent factors of L2 listening construct.

Table 18. Chi-square Difference Tests Comparing Models 4, 5 and 6

Model 4: $\chi^2 = 42.73$, $df = 17$	Model 5: $\chi^2 = 62.06$, $df = 19$	$\Delta\chi^2 = 19.33$, $df = 2$, $p\text{-value} = .00$
	Model 6: $\chi^2 = 105.8$, $df = 20$	$\Delta\chi^2 = 63.07$, $df = 3$, $p\text{-value} = .00$

Finally, to examine the distinctness of the cognitive factors MK and WM, two first-order CFA models were tested. In Model 7, the two factors were separate but correlated, and in Model 8, all measured variables of MK and WM loaded on a single factor. Table 19 summarizes the fit indices for these two models.

Table 19. Summary of Fit Indices for CFA Models 7 and 8

Index	CFI	NFI	NNFI	GFI	RMSEA	SRMR	Chi-square
Criterion	$\geq .95$	$\geq .90$	$\geq .95$	$\geq .90$	$\leq .06$	$\leq .08$	None significant
Model 7	.96	.95	.91	.98	.08	.05	$*\chi^2 = 93.16$, $df = 13$
Model 8	.49	.48	.23	.82	2.3	.16	$*\chi^2 = 206.62$, $df = 14$

Evaluation of fit indices revealed that Model 7 fit the data well. However, Model 8 did not fit the data. The correlation between the two factors in Model 7 was .25 and statistically significant. The small correlation between the two factors and acceptable fit indices for Model 7, contrasted with unacceptable fit indices for Model 8, suggested that MK and WM were distinct in the current study.

5.6 Structural Equation Modeling (SEM)

SEM combines CFA and regression to model how latent factors are related to each other (Mueller & Hancock, 2008; Flora & Curran, 2004). An advantage for using SEM is that measurement error in the observed variables is reduced. The latent factors resulting

from SEM analyses consist of the variance that each of its measured variables have in common. Therefore, measurement error is partialled out, and the latent factors may be considered free of measurement error. In addition, SEM provides a better basis for comparing the relative significance of independent factors because they all can be entered into the model simultaneously, and unlike multiple regression, the arbitrary order of entering independent factors into the model does not influence the results.

In the current study, the dependent latent factor was L2 listening ability, and in the SEM analyses, it was regressed on the independent latent factors of VB, VD, SK, MK, WM and AX. The strength of the relationships between the dependent and independent latent factors is indicated by structural regression weights and significance tests. To evaluate and compare the SEM models in the current study, the same profile of fit tests and indices as described for CFA was used.

The aim of SEM analysis was to answer the current research questions about the role of VK and SK, as well as their relative significance in explaining success in L2 listening comprehension. However, before testing the overall SEM models (in which the latent factor of L2 listening ability was regressed on factors VB, VD, SK, MK and WM), a set of separate SEM models was tested. In these models, the latent variable of L2 listening ability was regressed on each separate independent latent variable. This preliminary step was taken to examine whether each independent factor was individually a significant variable of L2 listening ability. Table 20 summarizes the fit indices for these SEM models. SEM Models 1 to 6 had VB, VD, SK, WM, and AX as single independent factors of the L2 listening factor.

Table 20. Summary of Fit Indices for SEM Models 1 to 6b

Index	CFI	NFI	NNFI	GFI	RMSEA	SRMR	Chi-square
Criterion	≥ .95	≥ .90	≥ .95	≥ .90	≤ .06	≤ .08	None significant
Model 1	.99	.99	.99	.97	.06	.03	$\chi^2 = 33.99, df=19$
Model 2	1	1	1	.99	0	.01	$\chi^2 = 7, df=8$
Model 3	.99	.99	.99	.98	.07	.02	$\chi^2 = 19.04, df=8$
Model 4	.99	.98	.98	.98	.06	.04	$\chi^2 = 17.06, df=8$
Model 5	1	1	1	.99	.01	0	$\chi^2 = 4.87, df=8$
Model 6a	.86	.84	.8	.86	.16	.16	$*\chi^2 = 182.13, df=26$
Model 6b	.97	.95	.94	.96	.08	.06	$*\chi^2 = 40.73, df=13$

Model 6a had MK as the single variable of L2 listening ability; however, the fit indices were poor, and the PS and MT sub-sections of the MK questionnaires did not significantly load on the MK factor. For these reasons, PS and MT were deleted in Model 6b. In comparison to Model 6a, Model 6b fit indices improved and were within the acceptable range. In addition, the chi-square difference test between Models 6a and 6b ($\Delta\chi^2 = 141.4, df = 13, p\text{-value} = .000$) showed that Model 6b fit the overall data better. In Model 6b, all measured variables loaded on the MK factor significantly. Therefore, Model 6b was kept, and the MK factor had three measured variables in subsequent SEM analyses. Figure 3 shows SEM Models 1 to 5 and 6b.

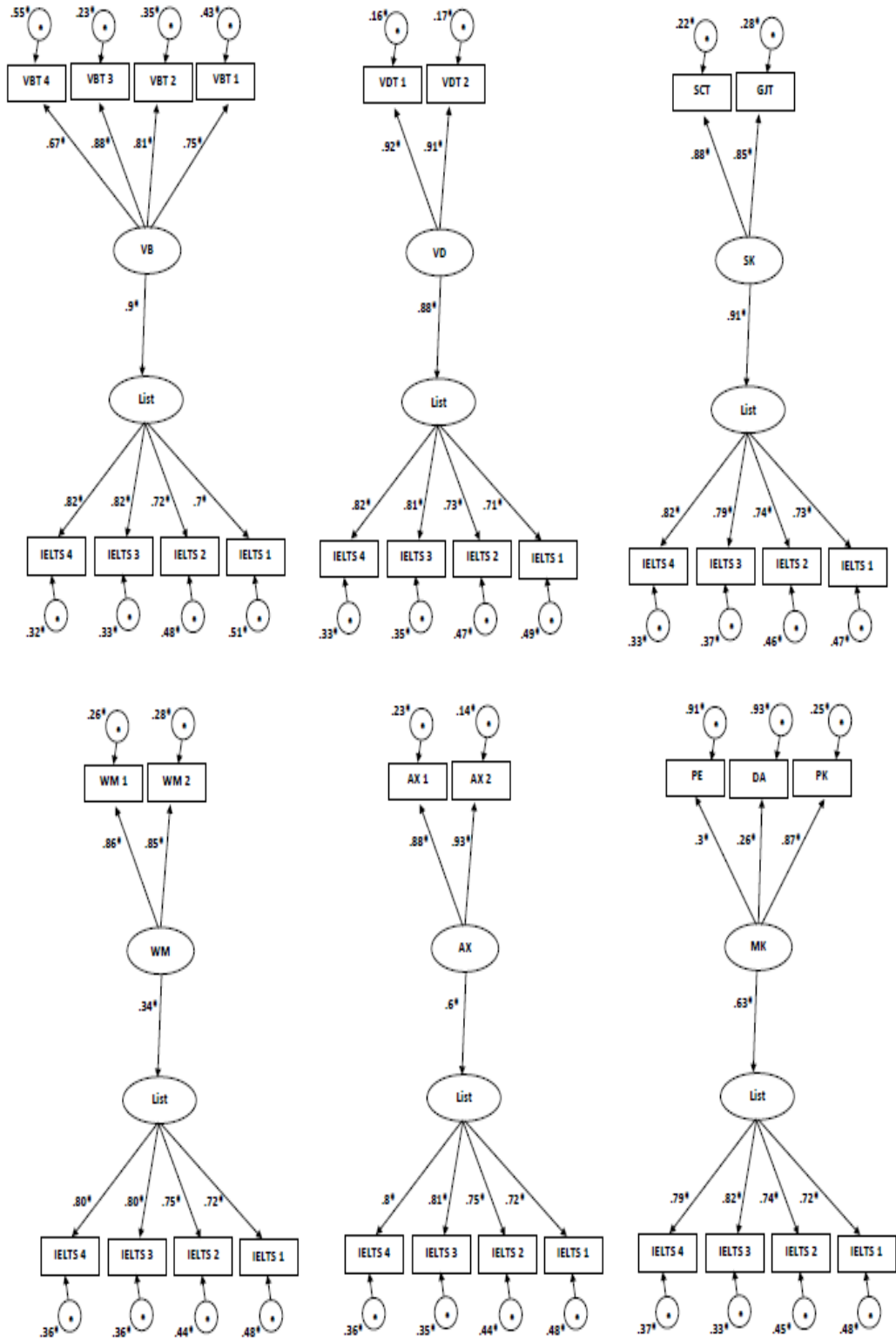


Figure 3. SEM Models 1 to 5, and 6b.

Several SEM models with all of the latent independent variables were then tested. First, in Model 7, L2 listening ability was regressed on the latent independent factors of VB, VD, SK, MK and WM. As can be seen in Table 21, fit indices for SEM Model 7 were within the acceptable range. However, in this model, the regression paths of the VB and WM independent factors to the L2 listening latent factor were not statistically significant.

Table 21. Summary of Fit Indices for SEM Models 7, 8 and 9

Index	CFI	NFI	NNFI	GFI	RMSEA	SRMR	Chi-square
Criterion	≥ .95	≥ .90	≥ .95	≥ .90	≤ .06	≤ .08	None significant
Model 7	.99	.97	.98	.92	.06	.05	* $\chi^2 = 194.21, df = 104$
Model 8	.98	.97	.98	.90	.06	.05	* $\chi^2 = 266.76, df = 136$
Model 9	.98	.96	.97	.92	.07	.06	* $\chi^2 = 171.92, df = 79$

In the next step, the AX independent factor was added to Model 8. However, in this model, the L2 listening factor was not directly regressed on the AX predictor factor. Instead, in light of theoretical accounts of the influence of AX on L2 listening, the WM factor was regressed first on the AX independent factor, followed by the L2 listening factor being regressed on the WM factor. Therefore, by testing Model 8 and comparing its results with Model 7, it could be determined whether accounting for the influence of AX on WM changes the relationship of WM and L2 listening ability.

In Model 8, as seen in Figure 4, AX was a significant latent variable factor for WM and explained 6% of the variance in the latent factor. It should be noted that sign of the regression path between AX and WM is negative in Model 8 and subsequent SEM models because original AX scores were used in this set of SEM analyses. In terms of model fit indices (seen in Table 21), Models 7 and 8 were almost identical. However, in Model 8, WM became a significant variable of L2 listening ability. VB, however, was still a non-significant independent latent variable in Model 8.

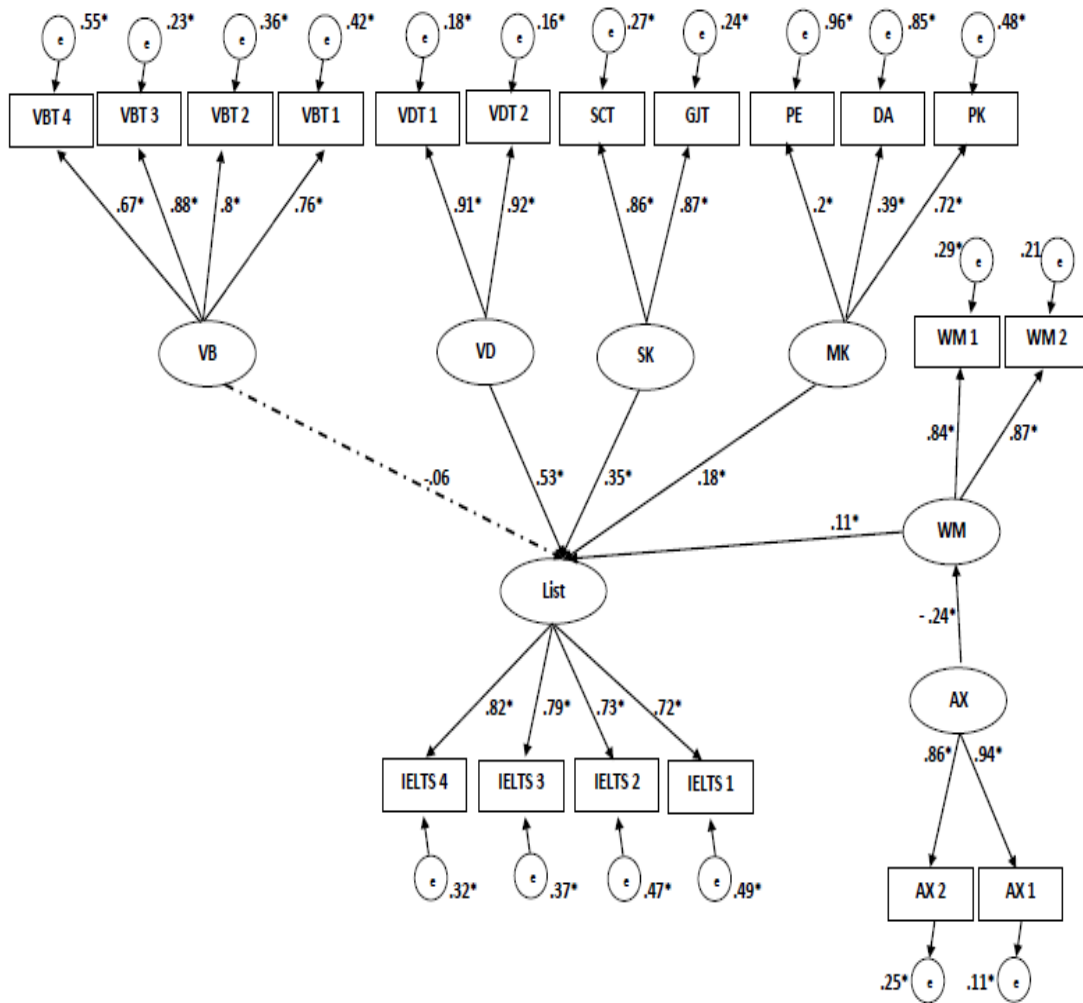


Figure 4. SEM Model 8

Because in Model 8, VB was still not a significant variable of L2 listening ability, in Model 9 it was deleted as an independent latent variable. As seen in Figure 5, after deleting VB, all independent factors in Model 9 were statistically significant. As shown in Table 21, Models 8 and 9 had almost identical fit indices. This suggests that the deletion of VB did not lead to any changes in the structural relationships among latent factors from a statistical point of view.

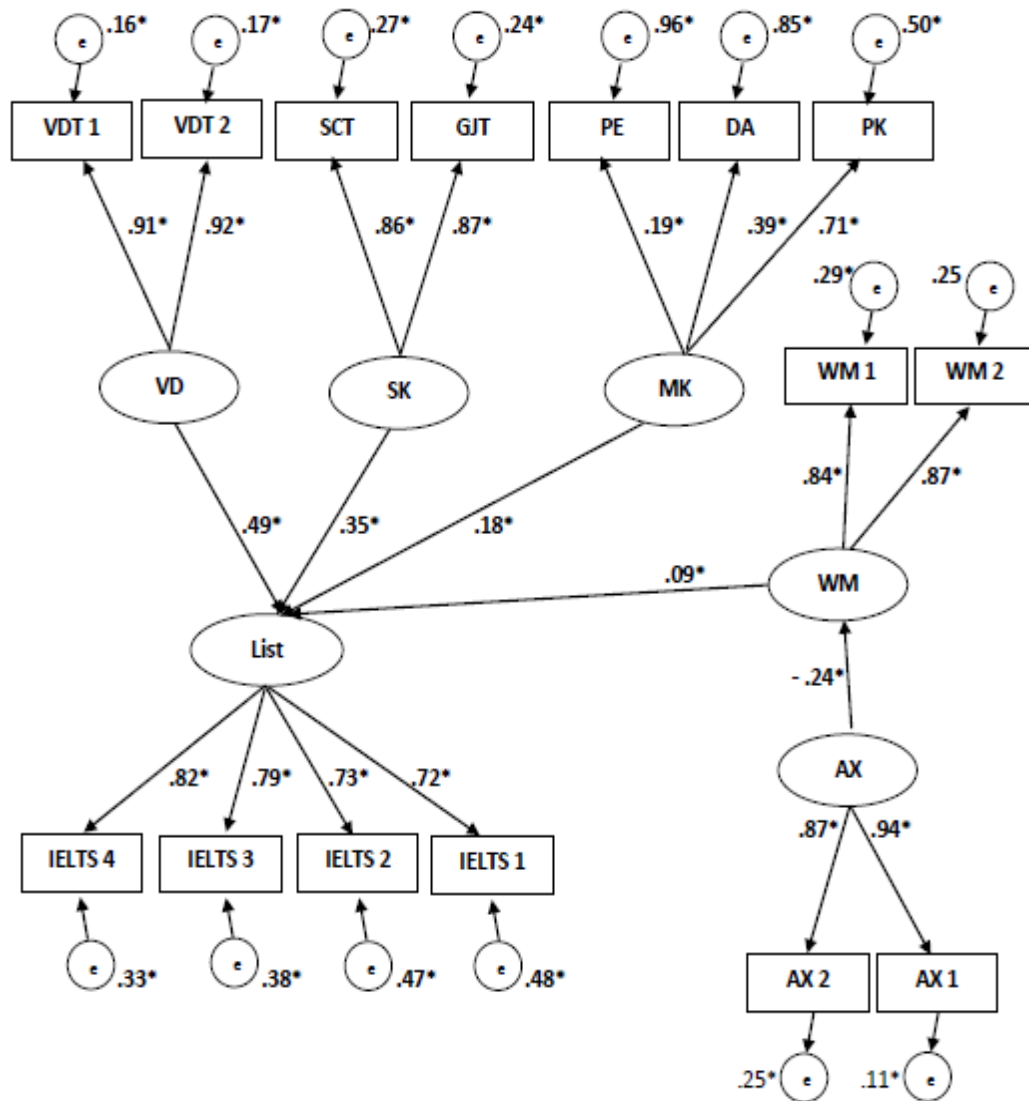


Figure 5. SEM Model 9

The VB factor had a high correlation with the VD factor, and because the VD factor could better explain success in L2 listening ability, VB turned into a redundant independent factor with a non-significant regression path to the L2 listening factor. This makes deletion of VB from the SEM models justifiable from a statistical point of view; however, from a substantive point of view, VB should not be altogether removed.

VB and VD factors were highly correlated, so these two factors could be considered as a single factor. Therefore, in Model 10, the measured variables of VB and VD were loaded on a single factor of VK, and then the L2 listening factor was regressed on it. Evaluation of fit indices of Models 10 showed that the model fit the data well.

Table 22 summarizes fit indices for Model 10.

Table 22. Summary of Fit Indices for SEM Model 10

Index	CFI	NFI	NNFI	GFI	RMSEA	SRMR	Chi-square
Criterion	≥ .95	≥ .90	≥ .95	≥ .90	≤ .06	≤ .08	None significant
Model 10	.98	.97	.98	.89	.07	.06	* $\chi^2 = 298.88$, $df = 141$

Although Model 10 fit the data well, to answer the main research questions of the current study, a more parsimonious model was desired. To examine the relative significance of VK and SK in explaining success in L2 listening comprehension, SEM Model 11 was tested. Model 11 was a modified version of Model 10, in which the number of measured variables of the latent factor of VK was reduced. Rasch person ability logits of the entire VBT and VDT measures were used as the two measured variables of the independent latent factor of VK. This way, both VK and SK latent independent factors each had two measured variables. Evaluation of fit indices revealed that Model 11 fit the data well, as Table 23 indicates.

Table 23. Summary of Fit Indices for SEM Model 11

Index	CFI	NFI	NNFI	GFI	RMSEA	SRMR	Chi-square
Criterion	≥ .95	≥ .90	≥ .95	≥ .90	≤ .06	≤ .08	None significant
Model 11	.98	.96	.97	.92	.07	.06	* $\chi^2 = 182.15$, $df = 79$

In Model 11, as can be seen in Figure 6, all regression paths between measured variables and their latent factors, as well as the ones between independent latent factors and the L2 listening latent factor, were statistically significant.

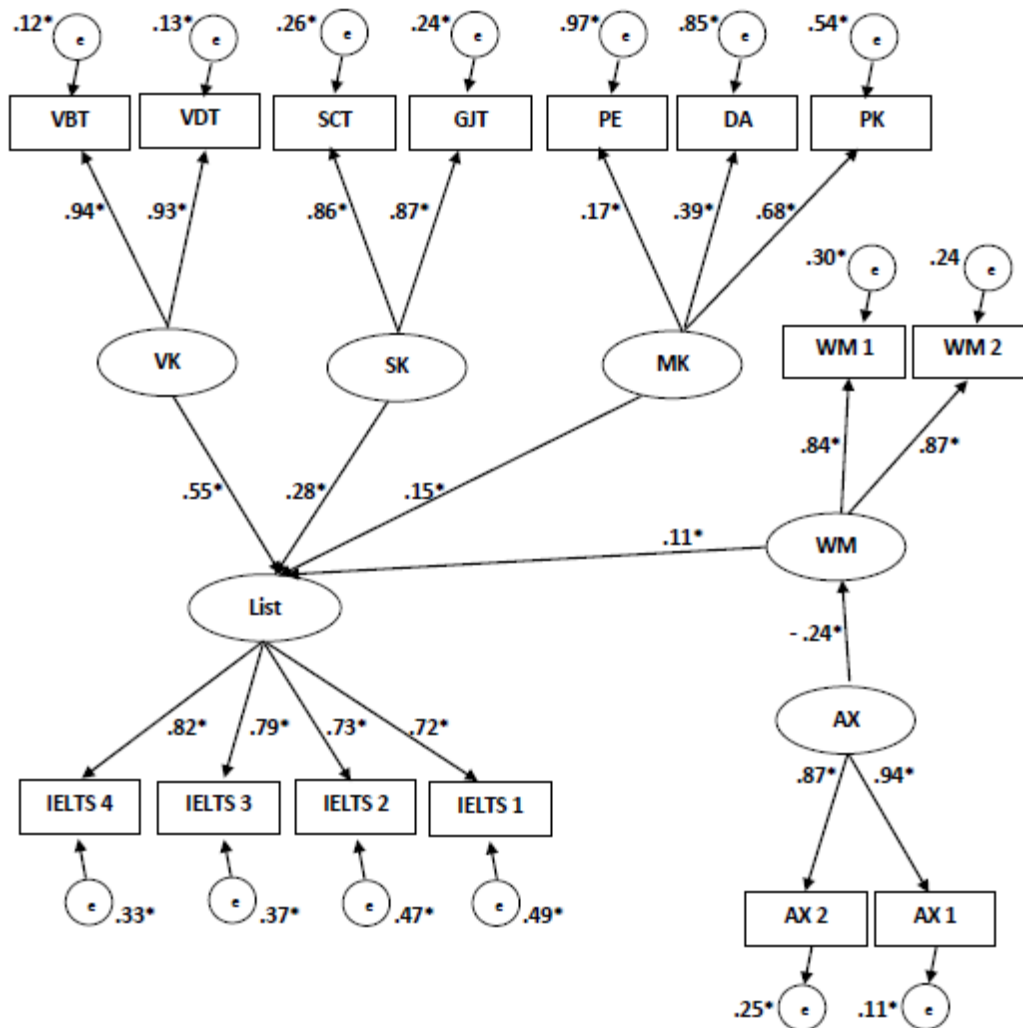


Figure 6. SEM Model 11

As the most inclusive, but also the most parsimonious model, Model 11 was employed for subsequent analyses. In order to examine whether each of the latent independent variables in Model 11 explained a significant amount of variance in the L2 listening latent variable, four alternative models to Models 11 were tested. In each of these alternative models, the structural regression path from an independent variable was constrained as carrying a zero regression weight. The chi-square test was employed to evaluate any difference in fit between each of these more constrained models and Model 11. If constraining the regression weight for an independent variable at zero does not lead to deterioration of model fit, that independent variable adds little or nothing to the explanation of individual differences in L2 listening ability (see, e.g., Schoonen, Hulstijn, & Bossers, 1998).

The regression weights for VK, SK, MK and WM were respectively set to zero in Models 12 to 15. Table 24 summarizes the chi-squares for these models, as well as the results of the chi-square difference tests between these models and Model 11.

Table 24. Chi-square Difference Tests Comparing SEM Models 11 to 15

Model 11: $\chi^2 = 182.15$, $df = 79$	Model 12: $\chi^2 = 194.85$, $df = 80$	$\Delta\chi^2 = 12.7$, $df = 1$, p -value = .00
	Model 13: $\chi^2 = 187.29$, $df = 80$	$\Delta\chi^2 = 5.14$, $df = 1$, p -value = .02
	Model 14: $\chi^2 = 186.24$, $df = 80$	$\Delta\chi^2 = 4.09$, $df = 1$, p -value = .04
	Model 15: $\chi^2 = 190.83$, $df = 80$	$\Delta\chi^2 = 8.68$, $df = 1$, p -value = .00

The alternative models to Model 11 all had larger chi-squares, and the results of the chi-square difference tests between these alternative models and Model 11 were statistically significant. This means that each of these latent independent variables significantly added to the explanation of individual differences in L2 listening ability. For

this reason, all independent variables were kept, and Model 11 was used for examining the relative significance of VK and SK in L2 listening.

Parameter estimates (e.g., regression weights) of VK and SK and the amount of variance they uniquely explained in the L2 listening latent variable were compared. The beta values (standardized loadings) in Model 11 for VK and SK were .55 and .28, respectively. Both beta values were statistically significant. This shows that, descriptively, VK accounted for individual differences in L2 listening better than SK. This more descriptively significant role of VK than SK was also revealed in the amount of unique variance these two factors explained in the L2 listening factor. All the independent latent variables jointly explained 90% of variance in the L2 listening ability factor, and VK and SK uniquely accounted for 6% and 2% of the explained variance, respectively. The unique contribution to individual differences in L2 listening ability for MK and WM factors was each 1%.

Therefore, descriptively, it seems that VK plays a more significant role than SK in explaining success in L2 listening ability. However, these results should be tested statistically. In order to do so, an alternative SEM model to Model 11 was tested. In Model 16, the regression paths of VK and SK to L2 listening ability were set to be equal. This means that in Model 16, the significance of VK and SK in explaining success in L2 listening ability was considered equal. If the chi-square of Model 16 increased in comparison to Model 11, and if the chi-square difference test between the two models was significant, it would mean that the descriptive difference between VB and SK is also statistically significant. On the other hand, if the chi-square difference test between Models 11 and 16 is non-significant, there is no statistical evidence to conclude that VK

plays a more significant role than SK in explaining success in L2 listening ability. Figure 5 shows Model 16.

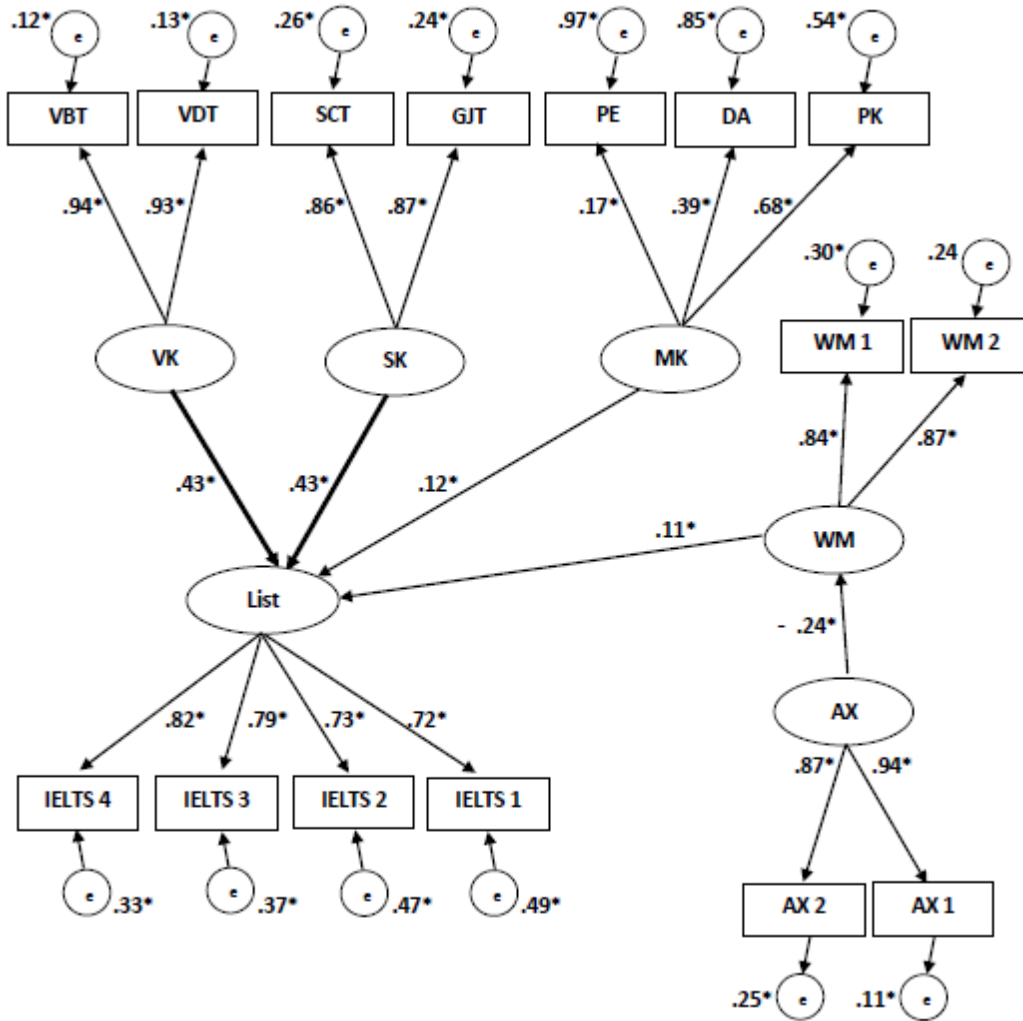


Figure 7. SEM Model 16

As seen in Figure 7, after setting the regression paths of VK and SK to the L2 listening ability latent factor as equal, the regression weights for both independent factors were estimated to be .43 and statistically significant.

Table 25 summarizes fit indices for Model 16. In terms of fit indices, Models 11 and 16 are almost identical; however, the chi-square for Model 16 is slightly larger.

Table 25. Summary of Fit Indices for SEM Model 12

Index	CFI	NFI	NNFI	GFI	RMSEA	SRMR	Chi-square
Criterion	≥ .95	≥ .90	≥ .95	≥ .90	≤ .06	≤ .08	None significant
Model 16	.98	.96	.97	.91	.07	.06	* $\chi^2 = 183.04$, $df = 80$

A chi-square difference test between Models 11 and 16 was then conducted. The results ($\Delta\chi^2 = .89$, $df = 1$, $p\text{-value} = .35$) showed that these two models were not statistically significantly different from each other. These results suggest that although VK plays a more important role than SK in explaining success in L2 listening from a descriptive point of view, the difference between the two independent factors is not statistically significant.

Chapter 6: Discussion of Results and Conclusions

6.1 Discussion of Results

The main purpose of the current study was to examine the role of vocabulary knowledge (VK) and syntactic knowledge (SK) in L2 listening comprehension, as well as their relative significance. The role of these two linguistic factors has been recognized in theoretical models of L2 listening; however, given the under-researched nature of L2 listening, limited empirical evidence is available to support these theoretical accounts. Previous studies were limited both quantitatively and qualitatively.

Unlike previous studies, the current project employed assessment tasks to measure *aural* and *proceduralized* VK and SK. In terms of VK, to avoid under-representing the construct, measures of both *breadth* (VB) and *depth* (VD) were included.

Additionally, the current study examined the role of VK and SK by accounting for individual differences in two important cognitive factors of L2 listening: metacognitive knowledge (MK) and working memory (WM). Also, to reveal the significant role of VK and SK more fully, the current study accounted for the negative impact of anxiety on WM, and in turn, on L2 listening. Finally, through SEM analysis the current study examined the role of VK and SK in L2 listening at the *construct level*, free from the influence of measurement error.

In the present chapter, a synthesis of results is discussed to answer the research questions. Answers to the preliminary research questions justified the inclusion of the variables into the statistical analyses, which subsequently led to answers for the main research questions.

6.1.1 Preliminary research question 1:

Are VB, VD, SK, MK, WM and anxiety associated with L2 listening ability?

Following theoretical and empirical findings in L2 listening, it was expected that all the above variables would be associated with L2 listening in a statistically significant way. Results of correlation analyses at both measurement and construct levels, as summarized in Tables 9 and 15, showed that this expectation was met. Linguistic variables (in comparison with cognitive and anxiety variables) were found to correlate more strongly with L2 listening ability.

At the measurement level, the correlation of L2 listening to VB and VD was .78 and .8, respectively. These correlations at the construct level were .9 and .88, respectively. The correlations among the constructs, obtained from a CFA, were higher because the effect of measurement error was partialled out. These results are both similar to and different from Staehr (2009). This study is the only other empirical investigation of the relative significance of VB and VD in explaining success in L2 listening. Similarly to the current study, Staehr found correlations between measures of VB and VD and the L2 listening measure to be strong and statistically significant.

However, compared to the current study, the order of the strength of associations in Staehr (2009) was in the opposite direction. In Staehr's study, the correlations between the VB and VD measures and the listening measure were .7 and .65, respectively. However, differences in the magnitude of associations between VB, VD and L2 listening measures were so small in both studies that no conclusion on the relative significance of the two can yet be drawn.

In the present study, the correlation between SK and L2 listening was .77 at the measurement level, and .91 at the construct level. On this point, the only available study is Mecartty (2000). Without distinguishing between VB and VD, Mecartty examined the relationship between SK, VK and L2 listening. The L2 listening measure and measures of SK and VK in the study correlated at a .26 and .38 level, respectively. These considerably smaller correlations in Mecartty's study are most probably due to unstandardized measures and higher measurement error.

The correlation between the L2 listening measure and measures of MK and WM in the current study was .37 and .31. At the construct level, these correlations were .71 and .33. These results are quite similar to Vandergrift and Baker (2015), who found respective correlations of .23 (MK) and .37 (WM).

Anxiety moderately correlated with L2 listening both at the measurement and construct levels. These correlations were .52 and .59, and it should be noted that they were positive because responses to the anxiety questionnaires were flipped.

To sum up and respond to the first preliminary research question, results of the current study showed that VB, VD, SK, MK, WM and anxiety were all statistically significantly associated with L2 listening. With regards to the magnitude of the associations, linguistic variables were more strongly correlated with L2 listening than non-linguistic variables. In terms of the strength of associations between VB, VD and SK with L2 listening, the differences were so small as to indicate that the linguistic variables had nearly identical levels of association with L2 listening. These results provide preliminary evidence in support of the inclusion of all the variables from the initial analysis in subsequent analyses.

Correlational results reveal very little about the predictive power of variables. For this reason, to justify the inclusion of all the predictor variables of L2 listening ability in the current study, a set of regression and SEM analyses were conducted.

6.1.2 Preliminary research question 2:

Does MK make a significant contribution to L2 listening ability?

Results of regression analysis indicated that MK was a significant variable of L2 listening as a single variable, but not when entered into the models along with other variables. This non-significant relationship may have been due to non-significant correlations between two sub-sections of the MK questionnaire (i.e., MT and PS) and the L2 listening measures.

Results of SEM analyses confirmed this possibility. In SEM Model 6a, all indicator variables of MK were loaded on a single independent latent factor, and the L2 listening factor was regressed on it. Results showed that the model did not fit the data well, and the regression path from the two MK measures of MT and PS to the MK latent factor was not statistically significant. SEM Model 6b did not include these two indicators of the MK latent factor, and it fit the data well. Additionally, when the MK independent factor was entered into SEM Model 11, it was a significant variable of L2 listening ability. In SEM Model 11 (the primary model used in the study), the regression path between the MK independent factor and L2 listening was .15 and statistically significant. In Model 11, PK had the largest standardized loading on the MK latent factor.

According to these results, it can be concluded that MK is a significant variable of L2 listening success, and it is justifiable to include it in the main SEM models of the

study. These results and conclusion are in line with findings from previous studies. Several previous studies had found a significant role for MK in L2 listening ability. Both Vandergrift et al. (2006) and Vandergrift and Baker (2015) found that PK (Person Knowledge) was the strongest variable of L2 listening among all sub-components of MK. In other words, learner perceptions of self-ability to regulate or control listening processes was best at explaining L2 listening success.

6.1.3 Preliminary research question 3:

Does WM make a significant contribution to L2 listening? Does anxiety influence the effect of WM?

The results of the multiple regression analysis revealed that both WM and anxiety were significant variables of L2 listening ability. The standardized beta coefficients (β), however, showed that WM was the weakest variable among all measured variables. The results of the preliminary SEM analyses also showed that WM and anxiety were each individual significant variables of L2 listening at the construct level.

The evaluation of SEM Model 7, however, showed that WM was not a significant variable of L2 listening ability. Drawing on empirical and theoretical evidence for the negative influence of anxiety on WM, in SEM Model 8 effects of anxiety on WM were accounted for. Subsequently, WM became a significant variable of L2 listening. In the primary SEM model for the overall data in the study (Model 11), the regression path from the WM independent factor to the L2 listening factor had a standardized loading of .11. This loading was also statistically significant.

These significant results of WM as a variable of L2 listening stand in contrast to what Andringa et al. (2012) and Vandergrift and Baker (2015) found. In both of these

two studies, although WM had a significant association with L2 listening, it was not a significant predictor. Aside from other methodological shortcomings, the nonsignificant results for the role WM in Andringa et al. (2012) and Vandergrift and Baker (2015) may have resulted from them not accounting for anxiety effects on WM. Studies in cognitive psychology (e.g., Eysenck et al., 2007) have shown the detrimental effect of anxiety on both executive function and storage aspects of WM. Thus, anxiety can mask individual differences in WM capacity, resulting naturally in nonsignificant results. If individuals with larger WM capacity experience anxiety during an L2 listening comprehension task, they cannot make an effective use of that capacity. On the other hand, individuals with less WM who do not experience anxiety can make the fullest use of their WM capacity.

These results led us to conclude that WM is a significant variable of L2 listening, and it is justifiable to be included in the main analyses. On the hand, results showed that the role of WM in L2 listening can be negatively influenced by the effects of anxiety to an extent that its role may not be observable. For this reason, in SEM Model 11, the WM independent factor was regressed on the anxiety factor prior to the L2 listening factor being regressed on the WM independent factor.

6.1.4 Preliminary research question 4:

Do VB and VD factors make a significant contribution to L2 listening? If so, which one is the better predictor?

Results of the multiple regression analysis showed that both VD and VB were significant variables of L2 listening ability. As for the magnitude of the relationships, the comparison of standardized beta coefficients (β) showed that the measure of VD was a

stronger variable than VB. However, to fully answer this research question, a set of SEM analyses was conducted.

The preliminary SEM analyses showed that when VB and VD were set as individual variables of L2 listening success, both were statistically significant variables. In SEM Models 1 and 2, VB and VD were each significant variables of L2 listening. Both models fit the data well, and the statistically significant regression weights for VB and VD were .9 and .88, respectively. However, to examine whether the two were significant variables of L2 listening when included with other variables, SEM Models 7 and 8 were evaluated.

In SEM Model 7, the regression path from the VD independent factor to the L2 listening factor was statistically significant, whereas the same relationship for VB was non-significant. For this reason, in SEM Model 8, the VB independent factor was deleted, and the fit of this model was compared with Model 7. The result of the comparison showed that the deletion of VB did not influence the fit of Model 8 negatively, which may imply that VB is not a significant variable of L2 listening.

However, these statistical results need more cautious interpretations. The results of correlation and regression analyses along with a CFA showed that VB and VD were highly correlated, and the two constructs were not distinctly measured. Collinearity between the two factors rendered VB redundant with respect to VD. However, this does not mean that VB is not a significant variable of L2 listening. It means that the two factors of VD and VB overlapped significantly in their ability to explain L2 listening success, and VD was a stronger variable.

These results are in contrast with Staehr (2009), who reported that VB was a significant variable of L2 listening, while VD was not. Staehr's results, however, could have been the artifact of the hierarchical regression analysis employed. In hierarchical multiple regression analysis, the researcher determines the order of entry of the variables. Staehr argued that because VB is the basic dimension of VK, it should be entered into the regression model first. By entering VB into the model first, a significant amount of variance in L2 listening scores was explained by this variable, and because measures of VB and VD were highly correlated, not much variance was left to be explained by VD. Opposite results could have been obtained if the order of variables entered was reversed.

In SEM analysis, on the other hand, the order of entering independent factors into the model does not influence results, and the relative significance of variables can be examined concurrently. For this reason, according to the results of the SEM analyses in the current study, more confidence is lent to the argument that VD is a stronger variable of L2 listening ability than VB. This conclusion is better understood along the developmental approach to conceptualization of acquisition and accumulation of VK.

In the developmental approach, vocabulary acquisition is conceptualized as an incremental process in which knowledge of a word develops along a continuum of mastery ranging from zero knowledge to the full mastery of a lexical item (Read 2000). Milton (2009) argued from this perspective that VD "really seems to appear only after a sizable vocabulary breadth has been attained" (p. 169). This does *not* mean that VB develops independently or completely prior to VD. Rather, as Read (2004) explained, it means that "as learners expand the absolute number of words that they have some understanding of, they will also be learning more about words that they encounter or use

frequently” (p. 221). This suggests that the development of VD depends on the development of VB and requires more exposure to and frequent use of words.

Schmitt (2010) also pointed out the differences between VB and VD development. Schmitt suggested that developing VB can be achieved by explicit and intentional teaching and learning tasks, but developing VD requires massive exposure to the L2 and takes much more effort and time. Schmitt (2010) acknowledged that it is true that VB is the most foundational aspect of VK; however, much more must be known about words if they are to be used in real-time. This suggests that the development of VD is essential for VK to be useful in online language use tasks such as L2 listening comprehension.

Therefore, as L2 learners need more time and exposure to develop VD in comparison to VB, measures of VD are better at distinguishing VK among them. Also, since VD plays a more significant role in language use, it seems plausible to accept that VD can better explain individual differences in L2 listening ability.

Additionally, the superiority of VD in explaining individual differences in L2 listening can be explained from a measurement perspective. In order to perform well on measures of VD such as the one used in the current study, learners must know basic meaning of words as well as nuances of meaning. Learners must also be familiar with how a particular word is associated and collocated with other words. This makes the VD measures much more difficult than the VB measures, and in turn, better at differentiating learners in terms of overall VK. As the result, measures of VD and the latent factor of VD could explain individual differences in L2 listening ability better than VB. However,

these results raise the question as to the extent that VB and VD should be considered separate *dimensions* of VK, which is addressed in preliminary research question 5.

6.1.5 Preliminary research question 5:

Should VB and VD be considered distinct predictors of L2 listening in the current study, or should they be combined as a single predictor factor?

In the current study, there were several types of evidence suggesting VB and VD were not distinct dimensions of VK. At the measurement and construct levels, the correlation between VB and VD was .87 and .95, respectively. CFA Models 2 and 3 also suggested that VB and VD were not distinct. The statistically significant correlation between the two factors in CFA Model 2 was .94. The acceptable fit of CFA Model 3, in which all measures of VB and VD were loading on a single factor, also showed no distinction between the two. Results of SEM Model 10 also showed the two aspects of VK as indistinct. In this model, measures of VB and VD loaded on a single latent factor of VK, and the model fit the data well. In the regression analysis, collinearity clearly existed between the measures of VB and VD. These results together led us to conclude that VB and VD are not distinct dimensions of VK, at least as measured in the current study. For this reason, it was decided that VB and VD should be combined as a single predictor factor of VK in subsequent analyses related to the main research questions.

The lack of separation between VB and VD can be partially understood by examining the way these two constructs are measured. In VB measures, learners identify a word or a phrase which is the best synonym or definition for a particular word. However, knowledge of synonyms is also considered one aspect of VD. On the other hand, in VD measures, learners choose synonyms and collocations for a particular word.

To complete the task, learners must know the basic meaning of a given word, as well as the basic meaning of the synonym and collocation choices. Knowing basic meaning of words is considered VB. In terms of current methods of measurement, there seems to be no easy way to assess learners' VK separately in these two dimensions.

The dimension approach to defining VK has been mostly used for assessment purposes. This approach has a simplifying effect of breaking the complex construct of VK into its more manageable components (Schmitt, 2010). However, in research settings, it seems that available VK tests fall short in measuring these two dimensions distinctly.

Read (2004) stated that the breadth versus depth metaphor has served the rhetorical purpose of encouraging researchers to look beyond conventional vocabulary test items that require learners to indicate their VK in its most basic level. However, it seems that the dimension approach does not reflect the true nature of VK and its development. Thus, the two dimensions of VB and VD are not able to be adequately operationalized in a distinct way in assessment.

To define VK, an alternative view to the dimension approach is a network building perspective (Henriksen, 1999). In this view, VK development is the growth of a lexical network by incorporating new words into the mental lexicon and gaining knowledge of how to link a new word to or distinguish it from related words. This view assumes that as a learner's vocabulary size expands, newly-learned words need to be accommodated within an already existing mental lexical network. This approach is different from the dimension approach, as it does not focus on the learning of individual words. Rather, it explores the development of links between sets of words in the mental lexicon. VD is thus considered to be knowledge of how individual words are linked, as

well as the ability to distinguish semantically related words. In this way, VB and VD are not distinct aspects of vocabulary development; rather, they should be considered different levels of VK development (Read, 2004).

It can be concluded that if VB and VD were distinct dimensions of VK, VD would be a better predictor of success in L2 listening. Measures of VD tap a more nuanced level of VK, and for this reason, they are better at distinguishing levels of VK. However, there was no empirical evidence in the current study to consider VB and VD separate. For this reason, measures of VB and VD were combined to create a single predictor variable of VK.

6.1.6 Main research question 1:

Does VK make a significant contribution to L2 listening?

In order to answer this question, measures of VB and VD were combined in the multiple regression analysis to generate a single measure of VK. When the VK measure was entered into the regression model along with the other variables, results revealed that VK was a significant variable of L2 listening ability. A comparison of standardized beta coefficients (β) showed that VK was the strongest variable among all included variables. At the construct level, SEM Model 11 was examined. In this model, the magnitude of the standardized regression path between the VK independent factor and the L2 listening factor was .55 and statistically significant. These results led us to conclude that VK is a strong and significant variable of L2 listening ability. These results are also supported by findings in previous studies showing that VK plays a critical role in understating aural input.

Staehr (2009) found that measures of both VB and VD were strongly correlated with L2 listening comprehension, with measures of VK as significant predictors of L2 listening ability. Mecarty (2000) also found that VK was strongly correlated with L2 listening ability, with VK as a significant predictor of L2 listening ability. Vandergrift and Baker (2015) also found a strong association between VK and L2 listening ability. Even when including effects of several other variables, Vandergrift and Baker (2015) found that VK was the most important internal factor of L2 listening.

The important role of VK in L2 listening has similarly been acknowledged in componential theoretical models of L2 listening (Buck, 2001; Rost, 2013), and can be explained within process-oriented models of listening comprehension. According to these models, for successful listening comprehension to take place, L2 listeners rely on different knowledge sources through top-down and bottom-up processes. Successful listening comprehension is then the result of a complex interaction between these two types of processes (Vandergrift, 2007). Proceduralized aural VK facilitates bottom-up processing of aural input, allowing L2 learners to use top-down processing for successful L2 listening comprehension.

There is empirical evidence showing that L2 learners with limited proceduralized VK primarily employ a bottom-up approach for constructing meaning in listening comprehension (Lynch, 1998; Rubin, 1994). Heavy reliance on low-level cues can lead listeners with limited proceduralized VK to experience partial or complete break-downs in L2 listening comprehension (Buck, 2001).

L2 listeners with limited proceduralized VK tend to approach aural input through mental word-by-word translation (Vandergrift, 2003b). Such an approach is detrimental

to online comprehension of aural input, since aural input is ephemeral. L2 listeners do not have enough time to translate individual words in aural input. Moreover, limited VK limits the ability of L2 listeners to know the meaning of a large portion of words in the input, making word-by-word translation even more challenging.

Additionally, focusing on low-level cues in aural input does not allow L2 listeners with limited proceduralized VK to access contextual information. Contextual information can only be obtained if listeners pay attention to top-level cues in aural input. If listeners fail to recognize and assign meaning to certain words in the aural input through low-level processing, they cannot access relevant contextual information by drawing on top-level cues to construct an adequate meaning representation of aural input (Rost, 2002).

Additionally, L2 listeners with limited proceduralized VK focus exclusively on bottom-up processing. As such, they deprive themselves of the opportunity to employ cognitive and metacognitive strategies to compensate for inefficient or incomplete processing of the bottom-level cues (Staehr, 2009). As Bonk (2000) explained, successful employment of cognitive resources and metacognitive strategies depend on the quality and quantity of information obtained through low-level processing.

Detrimental effects of limited proceduralized VK on L2 listening comprehension have been supported by results of studies investigating sources of difficulty in listening among L2 learners. For example, Goh (2000), Hasan (2000) and Kelly (1991) found that limited VK was the most important obstacle in successful L2 listening comprehension. The importance of VK in L2 listening can also be discussed from the lexical coverage perspective. Studies such as Staehr (2009) and van Zeeland and Schmitt (2012) showed that for successful L2 listening comprehension, VK for at least 95 percent of aural text

lexical coverage is needed. Clearly, greater VK leads to a higher degree of aural text coverage, which subsequently increases the ability of L2 listeners to successfully handle heavy linguistic processing demands in listening comprehension.

Findings regarding the significant role of VK in L2 listening lend credence to the possibility of a L2 VK threshold for successful L2 listening. The Linguistic Threshold Hypothesis (Clarke, 1979; 1980) assumes that listeners need to attain a certain level of L2 linguistic knowledge before they can efficiently transfer L1 skills to L2 listening comprehension. Transferring L1 listening skills to the task of L2 listening comprehension increases the efficiency of aural input processing and assists L2 listeners in overcoming comprehension breakdowns. This transfer also helps L2 listeners optimize top-down processing and make better use of contextual information for more successful comprehension (Vandergrift & Baker, 2015). Considering this view in light of VK as a primary component of linguistic knowledge, it seems reasonable to assume that L2 listeners without sufficient L2 VK are often lacking the facilitative and compensatory effect of L1 listening skills.

Results in the current study, coupled with findings in previous empirical and theoretical investigations on the role of VK in L2 listening, lead us to conclude that recognizing words and assigning meaning to them is central to L2 listening ability. As a result, in order to be more successful at L2 listening, learners need to expand VK both quantitatively and qualitatively. They can then process bottom-level cues of aural input more efficiently, which would then enable them to focus on top-level cues to access contextual information. VK also assists L2 listeners in transferring L1 listening skills into L2 listening.

6.1.7 Main research question 2:

Does SK make a significant contribution to L2 listening?

Multiple regression analysis showed that the SK measure was a significant variable of L2 listening ability. The comparison of standardized beta coefficients (β) revealed that after VK, SK was the second strongest variable of success in L2 listening comprehension. In order to answer this research question at the construct level, SEM Model 11 was examined.

In SEM Model 11, the magnitude of the standardized regression path between the SK independent factor and the L2 listening factor was .28 and statistically significant. The comparison of the standardized regression paths in SEM Model 11 revealed that, at least descriptively, SK was the second-strongest variable of L2 listening ability (after VK) among all other variables. These results stand in contrast with what Mecartty (2000) found.

Mecartty (2000) investigated the relative significance of VK and SK in L2 listening comprehension. It was reported that although measures of both VK and SK were correlated with the measure of L2 listening, only VK could explain success in L2 listening. The measure of SK in Mecartty's (2000) study did not make any significant contribution to explaining variance in L2 listening test scores. However, as previously discussed, Mecartty's results could be related to methodological shortcomings. The measures tapped written and controlled SK, whereas tapping aural proceduralized SK is more appropriate to measure successful L2 listening comprehension. In addition, Mecartty (2000) used morphosyntactic target structures in constructing SK measures. Inaccurate or incomplete processing of morphosyntax does not necessarily lead to

miscomprehension. Mecartty (2000) also used hierarchical regression analysis, and entering VK as the first predictor variable into the model biased results against SK.

The current study employed tasks that tapped aural proceduralized SK. In order to develop appropriate measures, target structures for which difficulty in processing would lead to miscomprehension were chosen. Importantly, in the present study, the use of SEM did not allow bias for or against the predictive role of any variable.

The significant role for SK in predicting success in L2 listening found in the current study can be explained through theoretical accounts, both in componential and processing-oriented models. In componential models proposed by Buck (2001) and Rost (2013), SK has been outlined as one of the major linguistic components of L2 listening ability. However, these models do not explain how SK plays its role in listening comprehension. On the other hand, process-oriented models like Field's (2013) explain why SK is needed for the comprehension of propositions in L2 listening.

In Field's (2013) model, listening comprehension is broken down into the following five stages: input decoding, lexical search, parsing (low-level or bottom-up processes), meaning construction and discourse representation (high-level or top-down processes).

According to Field (2013), constructing idea units or understating propositions takes place at the *parsing* stage. Parsing is defined as the establishment of relationships between the meaning of individual words and whole utterances. After recognizing individual words and assigning meaning to them, L2 listeners must combine the words together to comprehend complete utterances or idea units. As a result of parsing, listeners typically keep the meaning of idea units or the gist of utterances in mind, and forget the

actual words and syntax of an utterance. Listeners combine the gist of individual idea units in order to comprehend the gist of the whole aural text. Like other low-level or bottom-up L2 listening processes (i.e., input decoding and lexical search), parsing also draws upon stored linguistic knowledge. Input decoding and lexical search primarily rely on phonological and lexical knowledge. Parsing, on the other hand, heavily draws on SK.

Underlying processes of L2 listening comprehension are online, and listeners constantly create hypotheses about what they will hear. These hypotheses can be at the level of the phoneme, word or whole utterance. To formulate hypotheses at the whole utterance level, listeners must impose a syntactic pattern on the utterance, which can be done only in a piecemeal fashion because aural input is heard in real time (Field, 2013).

Therefore, for successful parsing to take place, listeners match provisional hypotheses about aural input at the utterance level against subsequent evidence. Aural proceduralized SK can assist L2 listeners to more accurately anticipate upcoming syntactic structures in aural input. In case of a mismatch between the hypothesis and the aural evidence, aural proceduralized SK can assist L2 listeners to quickly formulate new hypotheses (Field, 2009). A line of indirect evidence for the importance of SK in L2 listening comes from studies on the effects of syntactic modification on the comprehensibility of input.

For example, Long (1985) showed that nonnative speakers performed significantly better in comprehending a modified lecture with less syntactically complex sentences. Cervantes and Gainer (1992), Chiang and Dunkel (1992) and Teng (2001) also found that EFL learners' comprehension was more successful when listening to passages with less syntactically complex passages. Since there is a limit to WM capacity, less

syntactically complex sentences can be parsed more efficiently. On the other hand, learners with more aural proceduralized SK can parse utterances more efficiently, leaving WM with more room for processing at other stages.

The assumption of the Linguistic Threshold Hypothesis (Clark, 1979) can also hold true about SK. SK is a main component of linguistic knowledge, and for this reason, it seems plausible to assume that learners should reach a certain level of L2 SK before being able to transfer L1 skills to the task of L2 listening comprehension.

6.1.8 Main research question 3:

What is the relative significance of VK and SK in explaining success in L2 listening ability?

The comparison of standardized regression weights in the multiple regression analysis showed that VK was a stronger variable of L2 listening ability than SK. In order to answer this question at the construct level, SEM Models 11 and 16 were evaluated.

In Model 11, the relative significance of these two linguistic factors was examined descriptively. In this model, the standardized regression paths from VK and SK to the L2 listening factor were .55 and .28, respectively. Both were also statistically significant. Therefore, in terms of standardized regression weights, VK played a more important role than SK in explaining success in L2 listening. This result was confirmed by the amount of unique variance VK and SK explained in the L2 listening factor. VK uniquely accounted for six percent of the total variance in the L2 listening factor, while SK uniquely accounted for two percent. However, these results were at the descriptive

statistics level, and in order to examine whether VK significantly explained L2 listening success better than SK, SEM Model 16 was tested.

In SEM Model 16, the regression paths of VK and SK to L2 listening ability were set to be equal, and the comparison between Models 11 and 16 revealed no statistically significant difference in the chi-squares. This suggests that the difference between VK and SK independent factors was not statistically significant. However, this non-significant difference could be due to lack of statistical power. For example, a larger sample size or less complex SEM model may have provided more statistical power to estimate the true difference between the contributions of these two linguistic factors. Current results do not provide clear conclusions about the relative significance of VK and SK in L2 listening. Liao (2007), the only other study that indirectly compared the relative significance of VK and SK in L2 listening, also did not present conclusive evidence in this regard.

Liao (2007) also used SEM analysis to investigate the relative significance of VK and SK in L2 listening, cautiously concluding that VK plays a more significant role in L2 listening than SK. This cautious conclusion was due to two limitations in the SEM analysis. First, Liao did not regress the L2 listening factor directly on the VK and SK predictor factors. Instead, the measures of VK and SK were loaded on a second-order predictor factor labeled lexico-grammatical knowledge. Then, the listening factor was regressed on this single predictor factor. Based on the loadings of the VK and SK measures on the lexico-grammatical factor, Liao indirectly compared the relative significance of VK and SK in the L2 listening factor. The results would be more interpretable if the listening factor had been directly regressed on VK and SK predictor

factors. Also, unlike the current study, Liao did not test whether differences in the magnitude of loadings led to statistically significant differences in the amount of contribution VK and SK made to L2 listening ability.

Taken together, the results of the current study and Liao (2007) suggest that the question of the relative significance of VK and SK in L2 listening remains open. As Mecartty (2000) explained, it is difficult to ascertain relative contributions of VK and SK in L2 listening due the complex interplay between VK and SK. Buck (1990) and Rost (2013) also argued that L2 listening comprehension is a massively parallel interactive process, taking advantage of information from a large number of sources. As a result, it may be impossible to conclusively determine the relative significance of its individual underlying factors. Additionally, the measurement instruments purported to measure VK and SK constantly overlap, making these linguistic knowledge sources difficult to isolate to examine their relative significance. However, the results of the current study do allow use to confidently conclude that both VK and SK are strong and significant variables of L2 listening.

Relying on the possibly low-powered descriptive results, we can cautiously conclude that VK plays a more significant role than SK in explaining success in L2 listening. This cautious conclusion is in line with several theoretical arguments and empirical evidence that assign a more important role to VK than SK in L2 listening.

According to Liao (2007), the primary purpose of listening is to understand meaning. For this reason, listening comprehension may involve more VK than SK. Due to the fleeting nature of aural input and the limited capacity of WM, L2 listeners may focus more on lexical than syntactic cues in the input (Buck, 2001). The input processing

model of VanPatten (2002) also postulates that, to accommodate the limited capacity of WM, L2 learners tend to process input for meaning before form.

Conrad (1985) investigated whether nonnative listeners paid more attention than natives to syntactic information as opposed to semantic information in listening. Results showed that native listeners primarily used semantic units to process spoken input, while nonnative listeners tended to direct more attention to syntactic information. However, as language proficiency increased, L2 listeners also relied more on semantic cues from the text and less on syntactic cues to process the aural message. Therefore, if L2 listeners pay more attention to semantic than syntactic information in the aural text, it seems plausible to assume that VK plays a more significant role than SK in successful L2 listening comprehension. In addition, studies such as Goh (2000), Hasan (2000) and Kelly (1991) found that limited VK was a more important obstacle in successful L2 listening comprehension than limited SK.

6.2 Conclusions

The central finding of the present study was that both vocabulary and syntactic knowledge play a significant role in explaining success in L2 listening. Regarding vocabulary knowledge, it was found that quality or depth of vocabulary knowledge makes a more significant contribution to L2 listening ability than does breadth of vocabulary knowledge. This result leads us to conclude that for successful L2 listening, much more than word recognition or assigning basic meaning to words is required. To be successful at L2 listening comprehension, learners need to develop vocabulary

knowledge to the extent that words become fully integrated into the mental lexicon network.

At bottom-up processing of aural input, different stages of comprehension are supported by various linguistic knowledge sources, and at the “lexical search” stage, lexical knowledge plays a fundamental role. During the lexical search, word segmentation and recognition, as well as assigning meaning to the recognized words forms the basis for aural comprehension (Field, 2013, Rost, 2013). By this account, breadth of vocabulary knowledge, which is defined as ability to recognize words and assign at least a superficial meaning to them, should be considered the most fundamental aspect of lexical knowledge in aural comprehension.

However, for efficient bottom-up processing of aural input, much more than just recognizing words and assigning a superficial meaning to them is required. Deeper vocabulary knowledge that involves knowledge of word forms, lemmas, parts of speech and collocations plays a significant role in successful aural comprehension (Rost, 2013). For example, knowledge of collocations, which is one of the key aspects of depth of vocabulary knowledge, facilitates bottom-up processing of aural input (Schmitt, 2010).

Knowledge of collocations is a part of knowledge of formulaic language, and it has been shown that collocations and other formulaic language are processed more quickly and efficiently than non-formulaic language (e.g., Conklin & Schmitt, 2008). As more efficient processing of formulaic language assists L2 listeners to compensate for a limited WM capacity, it can be argued that knowledge of collections (among other aspects of depth of vocabulary knowledge) is essentially connected with functional, fluent and communicative language use (Schmitt, 2010).

Therefore, although breadth of vocabulary knowledge enables L2 listeners to recognize words and have a partial knowledge of their meaning, without depth of vocabulary knowledge, efficient and successful processing of aural input may be challenging and lead to break downs in comprehension. If L2 listening comprehension is an example of functional, fluent and communicative use of language, depth of vocabulary knowledge seems to contribute to the successful comprehension more than breadth of vocabulary knowledge. It is important to note that the conclusion about the more significant role of depth of vocabulary knowledge does *not* undermine the importance of breadth of vocabulary knowledge.

As explained before, word recognition and meaning association forms the foundation of vocabulary knowledge, and is one of the key components of linguistic knowledge in drawing on bottom-up processes of L2 listening. Moreover, the expansion of vocabulary size can potentially lead to the development of depth of vocabulary knowledge. However, increasing vocabulary breadth should not result in learners ignoring the importance of depth of vocabulary knowledge. Therefore, increasing vocabulary knowledge both quantitatively and qualitatively should be recognized as a central way to improve L2 listening ability.

The current study also found a significant role for syntactic knowledge in explaining success in L2 listening. Although the role of this linguistic component has been acknowledged in theoretical models of L2 listening ability, almost no prior empirical evidence has supported a significant role for it.

Current results showed that, along with vocabulary knowledge, syntactic knowledge is one of the key linguistic sources L2 learners draw on in bottom-up

processing of aural input. Proceduralized syntactic knowledge is essential for effective parsing of aural input in order to create idea units and understand propositions in L2 listening comprehension.

It should be noted that the current study found a significant role for syntactic knowledge in L2 listening while accounting for the effect of other important factors such as vocabulary knowledge, metacognitive knowledge and working memory. In the final SEM models of the current study, the syntactic knowledge factor was a significant variable of L2 listening ability even when other factors, especially vocabulary knowledge, explained a significant amount of variance in the listening ability factor. These results, reinforces the importance of proceduralized syntactic knowledge in L2 listening ability, although the role of this linguistic factor has been ignored in L2 listening research.

As to the relative significance of vocabulary knowledge and syntactic knowledge, results in the current study do not allow us to reach a clear conclusion. Although vocabulary knowledge played a more significant role than syntactic knowledge in descriptive results, the difference in magnitude of importance was not statistically significant. According to theoretical accounts and empirical evidence presently and from previous studies, it can be tentatively concluded that vocabulary knowledge plays a more important role than syntactic knowledge in L2 listening.

It should also be noted that there is limited set of syntactic structures in each language and mastering these structures takes much less time and effort compared to learning a vast number of words and other lexical units. Although not tested empirically, it can be logically assumed that syntactic knowledge plays a more significant role in L2

listening among lower proficiency learners. As learners' proficiency grows, compared to lexical knowledge, individual differences in syntactic knowledge diminishes faster. For this reason, individual differences in vocabulary knowledge remain as a more important factor in explaining individual differences in L2 listening ability even at higher levels of proficiency.

The relative significance of these two linguistic components is somewhat of a less important issue, since both were found to play an important role in L2 listening success. Syntactic knowledge and vocabulary knowledge are highly correlated constructs; as such, teasing them apart completely in empirical studies is nearly impossible. Therefore, the important implication of the current findings is that *both* vocabulary and syntactic knowledge should be improved both qualitatively and quantitatively for more successful L2 listening comprehension to occur.

Proceduralized syntactic and vocabulary knowledge promote fluent and efficient bottom-up processing, which in turn frees up sufficient working memory space for top-down processing of aural input. As a result, L2 listeners can complete these tasks: access the contextual information of the aural input by paying attention to top-level cues; activate prior knowledge about the topic; effectively use cognitive and metacognitive strategies; and transfer L1 listening skills to the L2 listening task for more successful L2 listening comprehension.

In interpreting these conclusions, one important issue merits close attention. The design of the current study was cross-sectional, and for this reason, no one-way causal inference with regards to the relationship between L2 linguistic knowledge and L2 listening ability can be drawn. This relationship is bi-directional, and these constructs

typically co-vary. Listening is the main channel of input for learning new vocabulary and syntactic structures, and frequent exposure to words and syntactic structures in aural input can lead to proceduralization and automatization of linguistic knowledge. On the other hand, developing proceduralized vocabulary and syntactic knowledge (whether or not through listening) leads to improved listening ability. Although current conclusions do not imply a one-way causal relationship between L2 linguistic knowledge and listening ability, they do not discourage such conclusions particularly in the context of current study.

EFL classes in Iranian formal education system start at the age of ten and continue to the end of high school. These EFL classes are still taught based on the grammar-translation method by which the primary purpose of instruction is developing grammatical and vocabulary knowledge of learners. Iranian students typically start taking communicative EFL classes outside of the formal education system only after graduating from high school. This means that these students are typically exposed to a large amount of grammar and vocabulary before having the opportunity to develop language use skills such as listening. Therefore, considering syntactic and vocabulary knowledge as factors that explain success in L2 listening can be done with more confidence in the context of the current study compared to the ones conducted in the context of learning a language as a *second* language. In an *ESL* context (rather than an EFL context), for example, learners are exposed to a lot of naturalistic input and have the opportunity to develop their L2 linguistic knowledge and language use skills concurrently.

The evidence of a significant contribution of both vocabulary and syntactic knowledge to success in L2 listening has important pedagogical implications.

Awareness of the importance of vocabulary and syntactic knowledge in L2 listening ability is fundamental to a theoretically grounded pedagogy of listening comprehension. There is a general consensus in L2 listening research literature that listening instruction has favored the development of top-down processes by focusing its attention on the development of cognitive and metacognitive strategies at the expense of developing bottom-up processes (e.g., Field, 2001; Rost, 2002).

Developing bottom-up processes requires developing L2 learners' knowledge in vocabulary and syntax, among other L2 linguistic components. The results of the current study revealed that these two linguistic factors contribute to L2 listening skill significantly more than cognitive factors such as metacognitive knowledge and working memory. In addition, working memory capacity is theoretically not amenable to change through instruction, and metacognitive strategies can be imported from L1 listening given high enough levels of proficiency. For these reasons, L2 listening instruction should focus on the development of underlying linguistic components of listening.

It is important to note that the current study revealed that *proceduralized aural* knowledge of L2 vocabulary and syntax significantly contributes to explaining success in L2 listening. In order to complete the vocabulary tests in the current study, learners listened to the item words only once, and answered within a short time limit. In order to mark the correct response, they had to recognize words aurally and assign meaning or make associations with other words quickly. For the syntactic measures, learners heard item sentences only once and decided on grammaticality or responded to comprehension questions within a short time limit. These measures thereby assessed learners'

proceduralized aural vocabulary and syntactic knowledge, in a way corresponding more to the nature of online L2 listening.

Therefore, instruction for the development of vocabulary knowledge and syntactic knowledge should be carried out in a way which promotes proceduralized aural knowledge of the two constructs. This point is particularly necessary to teaching and learning languages in foreign rather than second language contexts. The current study was carried out in the context of EFL, in which the focus of instruction is primarily on the development of controlled written linguistic knowledge. In addition, EFL learners do not typically have exposure to naturalistic input, while rarely practicing the L2 in online communicative situations.

6.3 Limitations and Directions for Future Research

Several issues not addressed in the current study warrant more in-depth investigations. Conclusions were limited first by the variables included in the current study. Previous studies (e.g., Andringa et al., 2012; Vandergrift & Baker, 2015) have shown that other internal factors such as background knowledge, L1 listening ability, L1 vocabulary knowledge, L2 auditory discrimination ability, and intelligence or reasoning ability play a significant role in explaining success in L2 listening. Future studies on the role of vocabulary and syntactic knowledge and their relative significance in L2 listening should include as many variables as possible.

As mentioned above, one of the key variables not accounted for in the current study is the role of schematic or background knowledge. Background knowledge about the topic of listening passages can significantly influence the performance of test takers

and change results about the relative significance of vocabulary and syntactic knowledge in L2 listening. In the current study, the listening sub-section of IELTS test was used as the measure of L2 listening ability. As IELTS is a general English proficiency test, topics of its listening passages are chosen in a way that the general population of test takers have enough familiarity with them. Future studies should examine how individual differences in background knowledge about listening passages can influence the relative significance of vocabulary and syntactic knowledge in L2 listening.

Results and conclusions in the current study are also limited to the context of learning English as a *foreign* language by Persian speakers. These results may change in the context of learning other languages by other groups of language learners, especially in the context of learning a new language as a *second* language. In learning a new language as a foreign language, the main channel of input is visual, and learners are primarily exposed to written language samples through reading. On the other hand, in the context of learning a new language as a second language, the main input channel is aural, and learners are exposed to many aural language samples. This difference in the primary channel of input may result in a difference in the relative significance of vocabulary and syntactic knowledge in L2 listening.

Also, results and conclusions in the current study are limited to *Persian* EFL learners. Except for scientific and technological terms, there is a very limited number of cognates shared between Persian and English. In the current study, none of the English target words in the vocabulary tests were cognates in Persian. Therefore, the results of the current study, especially in terms of the role of vocabulary knowledge, may not be

directly generalized to L2 listeners whose first language share a lot of cognates with their second language.

In addition, measures in the current study tapped *aural* vocabulary and syntactic knowledge. The measurement modality effect is an important research issue.

Comparisons between results obtained from visual and aural measures of vocabulary and syntactic knowledge can shed light on the nature of linguistic knowledge that is required for successful L2 listening. In addition, the strong relationships among variables in the current study could have been an artifact of shared aural measurement modality. Future studies can examine the modality effect by including both aural and visual measures of vocabulary knowledge and syntactic knowledge and examine to what extent the current findings can be attributed to the modality of the measures.

Given the limited sample size in the current study, SEM multi-sample analysis was not conducted. In this kind of analysis, learners can be divided into several L2 listening ability levels, and the contribution of vocabulary knowledge and syntactic knowledge to L2 listening can be examined across different levels. Such analysis can provide indirect evidence of the role of vocabulary and syntactic knowledge in L2 listening from a developmental perspective.

Finally, the current study had an exploratory and correlational design. Future studies can go beyond this cross-sectional design and employ a longitudinal mixed methods design to explore whether L2 listening ability develops as a function of improved L2 vocabulary and/or syntactic knowledge. Advanced statistical procedures such as latent growth curve models could be usefully implemented to address these issues.

Appendix A

VBT Sample Items

1. see: They <saw it>.
 - a closed it tightly
 - b waited for it
 - c looked at it
 - d started it up
2. time: They have a lot of <time>.
 - a money
 - b food
 - c hours
 - d friends
3. period: It was a difficult <period>.
 - a question
 - b time
 - c thing to do
 - d book
4. figure: Is this the right <figure>?
 - a answer
 - b place
 - c time
 - d number
5. poor: We <are poor>.
 - a have no money
 - b feel happy
 - c are very interested
 - d do not like to work hard
6. microphone: Please use the <microphone>.
 - a machine for making food hot
 - b machine that makes sounds louder
 - c machine that makes things look bigger
 - d small telephone that can be carried around
7. nil: His mark for that question was <nil>.
 - a very bad
 - b nothing
 - c very good
 - d in the middle
8. pub: They went to the <pub>.
 - a place where people drink and talk
 - b place that looks after money
 - c large building with many shops
 - d building for swimming
9. circle: Make a <circle>.
 - a rough picture
 - b space with nothing in it
 - c round shape
 - d large hole
10. dig: Our dog often <digs>.
 - a solves problems with things
 - b creates a hole in the ground
 - c wants to sleep
 - d enters the water

Appendix B
VDT Sample Item

1. beautiful

<input type="checkbox"/> enjoyable <input type="checkbox"/> expensive <input type="checkbox"/> free <input type="checkbox"/> loud	<input type="checkbox"/> education <input type="checkbox"/> face <input type="checkbox"/> music <input type="checkbox"/> weather
---	--

2. bright

<input type="checkbox"/> clever <input type="checkbox"/> famous <input type="checkbox"/> happy <input type="checkbox"/> shining	<input type="checkbox"/> colour <input type="checkbox"/> hand <input type="checkbox"/> poem <input type="checkbox"/> taste
---	--

3. calm

<input type="checkbox"/> open <input type="checkbox"/> quiet <input type="checkbox"/> smooth <input type="checkbox"/> tired	<input type="checkbox"/> cloth <input type="checkbox"/> day <input type="checkbox"/> light <input type="checkbox"/> person
---	--

4. natural

<input type="checkbox"/> expected <input type="checkbox"/> helpful <input type="checkbox"/> real <input type="checkbox"/> short	<input type="checkbox"/> foods <input type="checkbox"/> neighbours <input type="checkbox"/> parents <input type="checkbox"/> songs
---	--

5. fresh

<input type="checkbox"/> another <input type="checkbox"/> cool <input type="checkbox"/> easy <input type="checkbox"/> raw	<input type="checkbox"/> cotton <input type="checkbox"/> heat <input type="checkbox"/> language <input type="checkbox"/> water
---	--

6. general

<input type="checkbox"/> closed <input type="checkbox"/> different <input type="checkbox"/> usual <input type="checkbox"/> whole	<input type="checkbox"/> country <input type="checkbox"/> idea <input type="checkbox"/> reader <input type="checkbox"/> street
--	--

7. bare

<input type="checkbox"/> empty <input type="checkbox"/> heavy <input type="checkbox"/> uncovered <input type="checkbox"/> useful	<input type="checkbox"/> cupboard <input type="checkbox"/> feet <input type="checkbox"/> school <input type="checkbox"/> tool
--	---

8. acute

<input type="checkbox"/> hidden <input type="checkbox"/> often <input type="checkbox"/> rich <input type="checkbox"/> sharp	<input type="checkbox"/> angle <input type="checkbox"/> hearing <input type="checkbox"/> illness <input type="checkbox"/> stones
---	--

9. common

<input type="checkbox"/> complete <input type="checkbox"/> light <input type="checkbox"/> ordinary <input type="checkbox"/> shared	<input type="checkbox"/> boundary <input type="checkbox"/> circle <input type="checkbox"/> name <input type="checkbox"/> party
--	--

10. complex

<input type="checkbox"/> angry <input type="checkbox"/> difficult <input type="checkbox"/> necessary <input type="checkbox"/> sudden	<input type="checkbox"/> argument <input type="checkbox"/> passengers <input type="checkbox"/> patterns <input type="checkbox"/> problem
--	--

Appendix C

GJT Sample Items

Number	Structure	Sentence
1	Active-C	My father was cut the grass in the yard every Saturday afternoon.
2	Object RC-NC	Last week, the computer, which I bought from France, got broken.
3	Passive Object RC-NC	I read the book which was written by a Russian writer.
4	Passive-NC	This song was made long before the band became famous.
5	Subject RC-C	My sister, who she majored in computer science, can use a computer well.
6	Object RC-NC	The man whom I saw him yesterday is from the United States.
7	Hypothetical-NC (2)	If I had more time, I would not worry about the project.
8	Causative -NC	Did she have you cleaned up her messy room?
9	Passive-NC	All the dishes had been washed before I arrived home.
10	Causative -C	My older brother made me cleans his room for many years.
11	Passive Object RC-NC	I saw the girl who she was offered to work as a nurse.
12	Hypothetical-NC (3)	If you had arrived earlier, we will not have called you.
13	Causative -C	I had John fix the old car yesterday for free.
14	Passive-NC	The picture took by a famous photographer from France.
15	Object RC-NC	Yesterday I called my friend, who I had met in New York last week.
16	Passive Object RC-NC	I met the new student who he was recently admitted to the program.
17	Causative -NC	Does your teacher make you do homework during the weekends?
18	Subject RC-C	My new boss, who he is very nice, lives in a small apartment.
19	Active-C	Students reviewed all the course materials before the final test.
20	Subject RC-C	My older sister, who I live with, knows a lot about cars.

Appendix D

SCT Sample Items

Number	Structure	Sentence
1	Subject RC-C	Sarah, who is the sister of Tom, knows a lot about cars. Q: Does Tom know a lot about cars?
2	Hypothetical-NC (2)	If they had enough money, they would not worry about their trip. Q: Are they worried about the costs of their trip?
3	Object RC-NC	The girl, who Mike called yesterday, is reading the book. Q: Did the girl call Mike?
4	Active-C	Sarah is going to call Tommy to tell him about the party. Q: Is Sarah going to call Tommy?
5	Causative -C	Mike had John fix his car yesterday for free. Q: Did Mike fix John's car?
6	Subject RC-C	Her new teacher, who looks exactly like John, lives in a boat. Q: Does John live in a boat?
7	Passive-NC	The tall man is being hit by the small boy in the parking lot. Q: Is the small boy hitting the tall man?
8	Passive Subject RC-NC	The gift, which was sent by Mary, arrived at Tom's office. Q: Did Mary receive the gift?
9	Hypothetical-C (2)	If she were faster, she could play on a good football team. Q: Does she play on a good team?
10	Passive Subject RC-NC	Tommy, who was welcomed by Sarah, is very calm and friendly. Q: Is Tommy friendly?
11	Active-C	My brother always pats my father on his back to show his support. Q: Does his father pat his brother?
12	Causative -NC	Sarah cannot have her babysitter pick up her children from school. Q: Does Sarah need to pick up her children?
13	Object RC-NC	Robert liked the town, which Mike hated for many years. Q: Did Mike hate the town?
14	Causative -C	Mike got John to paint his new apartment. Q: Did Mike paint John's apartment?
15	Subject RC-C	Their daughter, who is Tom's classmate, is very smart. Q: Is their daughter very smart?
16	Hypothetical-C (3)	If Sarah had called me yesterday, I would have ignored her. Q: Did Sarah call yesterday?
17	Causative -C	Their father made John work in the yard over the weekend. Q: Did John work during the weekend?
18	Hypothetical-NC (3)	If you had worked harder, you would not have failed the test. Q: Did he fail the test?
19	Passive Subject RC-NC	Her sister, who was loved by Tom at school, called Robert. Q: Did Tom call Robert?
20	Active-C	Mary painted John while he was sitting on a bench. Q: Did John paint Tom?

Appendix E

MALQ Questionnaire (English Version)

Type scale	Strategy or belief/perception					
Planning-evaluation	1. Before I start to listen, I have a plan in my head for how I am going to listen.	1	2	3	4	5 6
Directed attention	2. I focus harder on the text when I have trouble understanding.	1	2	3	4	5 6
Person knowledge	3. I find that listening in French is more difficult than reading, speaking, or writing in French.	1	2	3	4	5 6
Mental translation	4. I translate in my head as I listen.	1	2	3	4	5 6
Problem-solving	5. I use the words I understand to guess the meaning of the words I don't understand.	1	2	3	4	5 6
Directed attention	6. When my mind wanders, I recover my concentration right away.	1	2	3	4	5 6
Problem-solving	7. As I listen, I compare what I understand with what I know about the topic.	1	2	3	4	5 6
Person knowledge	8. I feel that listening comprehension in French is a challenge for me.	1	2	3	4	5 6
Problem-solving	9. I use my experience and knowledge to help me understand.	1	2	3	4	5 6
Planning/evaluation	10. Before listening, I think of similar texts that I may have listened to.	1	2	3	4	5 6
Mental translation	11. I translate key words as I listen.	1	2	3	4	5 6
Directed attention	12. I try to get back on track when I lose concentration.	1	2	3	4	5 6
Problem-solving	13. As I listen, I quickly adjust my interpretation if I realize that it is not correct.	1	2	3	4	5 6
Planning/evaluation	14. After listening, I think back to how I listened, and about what I might do differently next time.	1	2	3	4	5 6
Person knowledge	15. I don't feel nervous when I listen to French.	1	2	3	4	5 6
Directed attention	16. When I have difficulty understanding what I hear, I give up and stop listening.	1	2	3	4	5 6
Problem-solving	17. I use the general idea of the text to help me guess the meaning of the words that I don't understand.	1	2	3	4	5 6
Mental translation	18. I translate word by word, as I listen.	1	2	3	4	5 6
Problem-solving	19. When I guess the meaning of a word, I think back to everything else that I have heard, to see if my guess makes sense.	1	2	3	4	5 6
Planning/evaluation	20. As I listen, I periodically ask myself if I am satisfied with my level of comprehension.	1	2	3	4	5 6
Planning/evaluation	21. I have a goal in mind as I listen.	1	2	3	4	5 6

MALQ Questionnaire (Persian Version)

1	پیش از این که شنیدن را شروع کنم، در ذهن خود برنامه مشخصی برای گوش کردن دارم.
2	وقتی در درک متن شنیداری مشکل دارم، بیشتر حواسم را جمع می‌کنم و روی متن و موضوع تمرکز می‌کنم.
3	به نظر من مهارت شنیدن در زبان انگلیسی مشکل تر از مهارت‌های خواندن، صحبت کردن یا نوشتن است.
4	آنچه می‌شنوم را بصورت همزمان در ذهنم ترجمه می‌کنم.
5	از کلماتی که معنی آنها را می‌دانم استفاده می‌کنم تا معنی کلماتی که نمی‌دانم را حدس بزنم.
6	وقتی در حال گوش کردن، حواسم پرت می‌شود، دوباره سریعاً تمرکز خود را به دست می‌آورم.
7	در حال گوش کردن، چیزی که از متن می‌فهمم را با آنچه از قبل درباره‌ی موضوع می‌دانم مقایسه می‌کنم.
8	احساس می‌کنم درک شنیداری در زبان انگلیسی برای من چالش برانگیز و مشکل است.
9	در حال گوش کردن، از دانش و تجربه‌ام برای درک موضوع کمک می‌گیرم.
10	پیش از شنیدن به یک متن، به متن‌های مشابهی که ممکن است قبلاً درباره آن موضوع شنیده باشم فکر می‌کنم.
11	در حال گوش کردن، کلمه‌های مهم و کلیدی متن را در ذهنم ترجمه می‌کنم.
12	هنگامی که تمرکزم را از دست می‌دهم، سعی می‌کنم به سرعت دوباره حواسم را جمع کنم.
13	در حال گوش کردن، اگر بفهمم که برداشت من از موضوع صحیح نیست، سریعاً آن را اصلاح می‌کنم.
14	پس از اتمام شنیدن، به شیوه و کیفیت گوش کردنم فکر می‌کنم. بعد فکر می‌کنم که احتمالاً دفعه آینده چه شیوه متفاوتی برای گوش کردن به کار خواهم برد.
15	وقتی به زبان انگلیسی گوش می‌کنم، دستپاچه نمی‌شوم.
16	اگر در درک چیزی که می‌شنوم مشکل داشته باشم، قید مطلب را می‌زنم و دیگر گوش نمی‌کنم.
17	از ایده‌ی کلی و موضوع متن برای حدس زدن معنای کلماتی که نمی‌دانم یا نمی‌فهمم استفاده می‌کنم.
18	در حال گوش کردن، متن را کلمه به کلمه در ذهنم ترجمه می‌کنم.
19	وقتی معنی کلمه‌ای را حدس می‌زنم، دوباره به کل آنچه شنیده‌ام فکر می‌کنم تا ببینم آیا حدسم درست بوده است.
20	در حال گوش کردن، مرتباً از خود می‌پرسم که آیا از سطح درک شنیداری خود رضایت دارم یا نه.
21	وقتی دارم به متنی گوش می‌کنم، در ذهنم هدف مشخصی برای گوش کردن دارم.

Appendix F

Anxiety Questionnaire 1

1	Taking English listening comprehension tests does not scare me. امتحانات درک شنیداری به زبان انگلیسی باعث ترس و استرس در من نمی شوند.
2	Listening to native English speakers make me feel uneasy and confused. وقتی که به افرادی که انگلیسی زبان مادریشان است گوش میکنم، استرس میگیرم و گیج می شوم.
3	I have usually been at ease listening to native English speakers in class. من معمولاً در کلاس به افرادی که انگلیسی زبان مادریشان است مشکلی ندارم و احساس راحتی میکنم.
4	I almost never get uptight while taking English listening comprehension tests. من تقریباً هیچ وقت احساس اضطراب و استرس نمیکنم و وقتی که در یک امتحان درک شنیداری به زبان انگلیسی شرکت میکنم.
5	I get really uptight while taking English listening comprehension exams. من دچار اضطراب و استرس شدید می شوم و وقتی که در یک امتحان درک شنیداری به زبان انگلیسی شرکت میکنم.
6	I get a sinking feeling when I think of trying to complete a difficult listening comprehension exercise. من دچار اضطراب و استرس شدید می شوم و وقتی که به انجام دادن یک تمرین مشکل درک شنیداری به زبان انگلیسی فکر میکنم.
7	My mind goes blank and I am unable to think clearly when my instructor asks me a question in English. وقتی که معلم سئوالی را به زبان انگلیسی از من می پرسد، هول می شوم و نمیتوانم به راحتی فکر کنم.
8	I am afraid of doing English listening comprehension exercises when I know that they will be graded. از انجام تمرینات درک شنیداری به زبان انگلیسی دچار ترس و اضطراب میشوم و وقتی که می دانم قرار است بابت آن نمره ای دریافت کنم.
9	Just thinking about trying to understand a native English speaker makes me nervous. حتی فکر کردن به اینکه قرار است به فردی که انگلیسی زبان مادریش است گوش کنم و متوجه منظورش شوم به من استرس می دهد.

Anxiety Questionnaire 2

1	<p>I get nervous if a listening passage is read only once during English listening tests.</p> <p>اگر در طول یک آزمون درک شنیداری به زبان انگلیسی، یک متن تنها یک بار پخش شود، دچار استرس و نگرانی میشوم.</p>
2	<p>When a person speaks English very fast, I worry that I might not understand all of it.</p> <p>وقتی که فردی زبان انگلیسی را با سرعت بالا صحبت میکند، نگران میشوم که نتوانم که تمام موضوع را متوجه نشوم.</p>
3	<p>I am nervous when I am listening to English if I am not familiar with the topic.</p> <p>وقتی که به زبان انگلیسی گوش میکنم، اگر با موضوع آشنا نباشم، احساس استرس و نگرانی میکنم.</p>
4	<p>If I let my mind drift even a little bit while listening to English, I worry that I will miss important ideas.</p> <p>وقتی که به یک متن در زبان انگلیسی گوش میکنم، اگر حواسم برای مدت کوتاهی پرت شود، نگران میشوم که شاید دیگر مفهوم اصلی متن را متوجه نشوم.</p>
5	<p>When I'm listening to English, I am worried when I can't watch the lips or facial expression of a person who is speaking.</p> <p>وقتی که به یک متن در زبان انگلیسی گوش میکنم، اگر صورت و لبهای گوینده را نبینم، احساس نگرانی میکنم.</p>
6	<p>During English listening tests, I get nervous and confused when I don't understand every word.</p> <p>در طول یک آزمون درک شنیداری به زبان انگلیسی، اگر معنی تک تک لغاتی را که میشنوم ندانم، احساس استرس و سردگمی میکنم.</p>
7	<p>I feel uncomfortable in class when listening to English without the written text.</p> <p>در کلاس انگلیسی اگر به چیزی گوش کنم بدون اینکه متن کتبی آن را ببینم، دچار استرس میشوم.</p>
8	<p>I feel confident when I am listening in English.</p> <p>به توانایی خود در درک شنیداری به زبان انگلیسی اعتماد به نفس دارم.</p>
9	<p>I get worried when I have little time to think about what I hear in English.</p> <p>وقتی که به زبان انگلیسی گوش میکنم، اگر وقت کافی برای فکر کردن به آنچه می شنوم نداشته باشم، احساس استرس و نگرانی میکنم.</p>
10	<p>I get worried when I can't listen to English at my own pace.</p> <p>وقتی که به زبان انگلیسی گوش میکنم، اگر سرعت پخش متن سریع تر از سرعت مطلوب من باشد، احساس استرس و نگرانی میکنم.</p>
11	<p>I get upset when I'm not sure whether I understand what I am listening to English.</p> <p>اگر مطمئن نباشم که آیا آنچه را به زبان انگلیسی شنیدیم، درست متوجه شدم یا نه، احساس ناراحتی و استرس میکنم.</p>
12	<p>If a person speaks English very quietly, I am worried about understanding.</p> <p>اگر کسی با صدای آرام با من انگلیسی صحبت کند، نگران می شوم که آیا درست او را متوجه می شوم یا نه.</p>
13	<p>I have no fear of listening in English as a member of an audience.</p> <p>وقتی در جمعی حضور دارم، اصلاً نگران توانایی خود در گوش کردن به و درک زبان انگلیسی نیستم.</p>
14	<p>I am nervous when listening to an English speaker on the phone or when imagining a situation where I listen to an English speaker on the phone.</p> <p>وقتی که پای تلفن با کسی انگلیسی صحبت میکنم (یا حتی وقتی که به این موقعیت صحبت پای تلفن فکر میکنم)، احساس استرس و نگرانی میکنم.</p>
15	<p>I feel tense when listening to English as a member of a social gathering or when imagining a situation where I listen to English as a member of a social gathering.</p> <p>وقتی که به یک مهمانی میروم، از اینکه همه انگلیسی صحبت کنند (یا حتی وقتی که به رفتن به چنین مهمانی فکر میکنم)، احساس استرس و نگرانی میکنم.</p>
16	<p>Listening to new information in English makes me uneasy.</p> <p>وقتی که کسی اطلاعات جدیدی را شفاهن به زبان انگلیسی به من میدهد، دچار استرس میشوم.</p>
17	<p>I get annoyed when I come across words that I don't understand while listening to English.</p> <p>وقتی که به زبان انگلیسی گوش میکنم، اگر در متن کلماتی باشد که معنی آن را ندانم، دچار استرس می شوم.</p>
18	<p>It frightens me when I cannot catch a key word of an English listening passage.</p> <p>وقتی که به زبان انگلیسی گوش میکنم، از اینکه متوجه یک کلمه کلیدی و مهم نشوم، دچار اضطراب و نگرانی میشوم.</p>

Bibliography

- Adolphs, S. & Schmitt, N. (2003). Lexical coverage and spoken discourse. *Applied Linguistics*, (24) 4, 425-38.
- Al-jasser, F. (2008). The effect of teaching English phonotactics on the lexical segmentation of English as a foreign language. *System*, 36, 94–106.
- Andringa, S., Olsthoorn, N., van Beuningen, C., Schoonen, R., & Hulstijn, J. (2012). Determinants of Success in Native and Non-Native Listening Comprehension: An Individual Differences Approach. *Language Learning*, 62 (s2), 49-78.
- Alderson, J. R. (1993). The relationship between grammar and reading in an English for academic purposes test battery. In *A new decade of language testing research: Selected papers from the 1990 Language Testing Research Colloquium* (pp. 203-219).
- Allison, P. D. (1999). *Logistic regression using the SAS system: Theory and application*. Cary, NC: SAS Institute Inc.
- Anderson, J. R. (1995). *Cognitive psychology and its implications* (4th ed.). New York: Freeman.
- Anderson, J. R. (2009). *Cognitive psychology and its implications* (7th ed.). New York: Worth.
- Atkins, S. M., (2011). Working memory assessment and training. Dissertation, University of Maryland, Digital Repository at University of Maryland [2012-02-17T07:03:07Z].
- Atkins, S.M., Harbison, J.I., Bunting, M.F., Teubner-Rhodes, S., & Dougherty, M.R. (2009). Measuring working memory with automated block span and automated letter-number sequencing. Poster presented at the 50th Annual Meeting of the Psychonomic Society, Boston, MA, USA.
- Atkins, S. M., Harbison, J. I., Bunting, M. F., & Dougherty, M. R. (2012). Using online games to assess cognitive ability: Validity of two web-deployable measures of cognitive ability. *A technical report*.

- Atkins, S.M., Sprenger, A.M., Colflesh, G.J.H., Briner, T.L., Buchanan, J.B., Chavis, S.E., ... & Dougherty, M.R. (2014). Measuring working memory is all fun and games: A four-dimensional spatial game predicts cognitive task performance. *Experimental Psychology*, *61*, 417-438.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F. (2004). *Statistical analyses for language assessment*. Cambridge, England: Cambridge University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. New York: Oxford University Press.
- Baddeley, A. D. (2000). The episodic buffer: a new component of working memory?. *Trends in cognitive sciences*, *4*(11), 417-423.
- Baddeley, A. D. (2001). Is working memory still working?. *American Psychologist*, *56*(11), 851.
- Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of communication disorders*, *36*(3), 189-208.
- Baddeley, A. D. (2012). Working memory: theories, models, and controversies. *Annual review of psychology*, *63*, 1-29.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of learning and motivation*, *8*, 47-89.
- Bae, J., & Bachman, L. F. (1998). A latent variable approach to listening and reading: Testing factorial invariance across two groups of children in the Korean/English two-way immersion program. *Language Testing*, *15*(3), 380-414.
- Beglar, D. (2010) A Rasch-based validation of the Vocabulary Size Test. *Language Testing*, *27* (1), 101-18.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

- Bloomfield, A., Wayland, S. C., Rhoades, E., Blodgett, A., Linck, J., & Ross, S. (2011). What makes listening difficult. Factors affecting second language listening comprehension. *A technical report*. Retrieved from <http://www.casl.umd.edu/node/127>
- Bonk, W. J. (2000). Second language lexical knowledge and listening comprehension. *International journal of listening*, 14(1), 14-31.
- Brindley, G., & Slatyer, H. (2002). Exploring task difficulty in ESL listening assessment. *Language Testing*, 19(4), 369-394.
- Brunfaut, T. & Révész, A. (2011). *EAP listening task difficulty: The impact of task variables, working memory and listening anxiety*. Paper presented at the Language Testing Research Colloquium. Ann Arbor, MI, June 23–25.
- Bond, T. G., & Fox, C. M. (2013). *Applying the Rasch model: Fundamental measurement in the human sciences*. Psychology Press.
- Buck, G. (1990). *The Testing of Second Language Listening Comprehension*. Unpublished PhD Thesis. University of Lancaster.
- Buck, G. (2001). *Assessing Listening*. Cambridge: Cambridge University Press.
- Buck, G., & Tatsuoka, K. (1998). Application of the rule-space procedure to language testing: Examining attributes of a free response listening test. *Language Testing*, 15(2), 119–157.
- Buck, G., Tatsuoka, K., Kostin, I., & Phelps, M. (1997). The sub-skills of listening: Rule-space analysis of a multiple-choice test of second language listening comprehension. In *Current developments and alternatives in language assessment: Proceedings of LTRC* (Vol. 96, pp. 589-624).
- Byrne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Call, M. E. (1985). Auditory Short-Term Memory, Listening Comprehension, and the Input Hypothesis. *Tesol quarterly*, 19(4), 765-781.
- Cervantes, R., & Ganier, G. (1992). The effects of syntactic simplification and repetition on listening comprehension. *Tesol Quarterly*, 26(4), 767-770.

- Chamot, A. U. (2005). Language learning strategy instruction: Current issues and research. *Annual Review of Applied Linguistics*, 25, 112-130.
- Chapelle, C. A. (1998). Construct definition and validity inquiry in SLA research. In L. F. Bachman and A. D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 32–70). Cambridge: Cambridge University Press.
- Chiang, C. S., & Dunkel, P. (1992). The effect of speech modification, prior knowledge, and listening proficiency on EFL lecture learning. *TESOL quarterly*, 26(2), 345-374.
- Clark, M., Wayland, S., Osthus, P., Brown, K. G., Castle, S., & BA, A. R. (2014). The effects of note taking on foreign language listening comprehension. *A technical report*. Retrieved from <http://www.govtilr.org/publications/notetaking.pdf>
- Clarke, M. A. (1979). Reading in Spanish and English: Evidence from adult ESL students. *Language learning*, 29(1), 121-150.
- Clarke, M. A. (1980). The short circuit hypothesis of ESL reading—or when language competence interferes with reading performance. *The Modern Language Journal*, 64(2), 203-209.
- Conrad, L. (1985). Semantic versus syntactic cues in listening comprehension. *Studies in Second Language Acquisition*, 7(01), 59-69.
- Conway, A. R., & Engle, R. W. (1994). Working memory and retrieval: a resource-dependent inhibition model. *Journal of Experimental Psychology: General*, 123(4), 354-373.
- Cowan, N. (2008). What are the differences between long-term, short-term, and working memory?. *Progress in brain research*, 169, 323-338.
- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American psychologist*, 12(11), 671.
- Daller, H., Milton, J., & Treffers-Daller, J. (2007). Editors' introduction. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 79-92). Cambridge: Cambridge University Press.

- Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, 3(4), 422-433.
- DeKeyser, R.M., & Koeth, J. (2010). Cognitive Aptitudes for Second Language Learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2, pp. 395-406). New York: Routledge.
- Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, M. (2008). German children's comprehension of word order and case marking in causative sentences. *Child Development*, 79(4), 1152-1167.
- Dunkel, P. (1991). Listening in the native and second/foreign language: Toward an integration of research and practice. *Tesol Quarterly*, 25(3), 431-457.
- Dunn, D. M., & Dunn, L. M. (2007). *Peabody picture vocabulary test: Manual*. Pearson.
- Elkhafaifi, H. (2005). Listening comprehension and anxiety in the Arabic language classroom. *The modern language journal*, 89(2), 206-220.
- Ellis, R. (1994). Factors in the incidental acquisition of vocabulary from oral input: a review essay. *Applied Language Learning*, 5(1), 1-32.
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current directions in psychological science*, 11(1), 19-23.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological review*, 102(2), 211.
- Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition & Emotion*, 6(6), 409-434.
- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, 7, 336-353.
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive psychology*, 47(2), 164-203.
- Feyten, C. (1991). The power of listening ability: an overlooked dimension in language acquisition. *The Modern Language Journal*, 75 (ii), 173-180.

- Field, J. (1998). Skills and strategies: Towards a new methodology for listening. *ELT Journal*, 52(2), 110–118.
- Field, J. (2001). Finding one's way in the fog: Listening strategies and second-language learners. *Modern English Teacher*, 9, 29–34.
- Field, J. (2004). An insight into listeners' problems: too much bottom-up or too much top-down?. *System*, 32(3), 363-377.
- Field, J. (2008). *Listening in the Language Classroom*. Cambridge: Cambridge University Press.
- Field, J. (2009). A cognitive validation of the lecture-listening component of the IELTS listening paper. *IELTS research reports*, 9, 17-65.
- Field, J. (2013). Cognitive validity. In A. Geranpayeh & L. Taylor (Eds.). *Examining listening: Research and practice in assessing second language* (pp. 77–151). Cambridge: Cambridge University Press.
- Flowerdew, J. (Ed.). (1994). *Academic listening: Research perspectives*. Cambridge: Cambridge University Press.
- Freedle, R., & Kostin, I. (1999). Does the text matter in a multiple-choice test of comprehension? The case for the construct validity of TOEFL's minitalks. *Language testing*, 16(1), 2-32.
- Fountain, R. L., & Nation, I. S. P. (2000). A vocabulary-based graded dictation test. *RELC Journal*, 31(2), 29-44.
- Gilmore, A. (2004). A comparison of textbook and authentic interactions. *ELT Journal*, 58(4), 363–374.
- Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching*, 40, 97–118.
- Goh, C. (1998). How ESL learners with different listening abilities use comprehension strategies and tactics. *Language Teaching Research*, 2(2), 124-147.
- Goh, C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, 28, 55–75.

- Goh, C. (2002). Exploring listening comprehension tactics and their interaction patterns. *System*, 30(2), 185-206.
- Goh, C. (2005). Second language listening expertise. In Jonson, K. (Ed.), *Expertise in second language learning and teaching* (pp. 64-84). New York: Palgrave Macmillan.
- Graham, S., & Macaro, E. (2008). Strategy Instruction in Listening for Lower-Intermediate Learners of French. *Language learning*, 58(4), 747-783.
- Granena, G. (2013). Reexamining the robustness of aptitude in second language acquisition. In Granena, G., & Long, M. (Eds.). *Sensitive periods, language aptitude, and ultimate L2 attainment* (Vol. 35) (pp. 179-204). USA: John Benjamins Publishing.
- Gyllstad, H. (2005). Words that go together well: Developing test formats for measuring learner knowledge of English collocations. *The Department of English in Lund: Working Papers in Linguistics, Volume 5*.
- Gyllstad, H. (2007). *Testing English collocations: Developing receptive tests for use with advanced Swedish learners*. Unpublished PhD Thesis. Lund University.
- Haigh, M., & Stewart, A. J. (2011). The influence of clause order, congruency, and probability on the processing of conditionals. *Thinking & Reasoning*, 17(4), 402-423.
- Harding, L., Alderson, J. C., Brunfaut, T., (2015). Diagnostic assessment of reading and listening in a second or foreign language: Elaborating on diagnostic principles. *Language Testing*, 1-20.
- Harley, B. (2000). Listening strategies in ESL: Do age and L1 make a difference? *TESOL Quarterly*, 34, 769-776.
- Hasan, A. S. (2000). Learners' perceptions of listening comprehension problems. *Language Culture and Curriculum*, 13(2), 137-153.
- Henning, G. (1990). *A study of the effects of variation of short-term memory load, reading response length, and processing hierarchy on TOEFL listening comprehension item performance*. (TOEFL Research Reports RR- 33). Princeton, NJ: Educational Testing Service.

- Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in second language acquisition*, 21(02), 303-317.
- Hoaglin, D. C., Iglewicz, B., & Tukey, J. W. (1986). Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 81(396), 991-999.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1-55.
- Hu, M. & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a foreign language*, 13(1), 403-30.
- Johnson, K. (1996). *Language Teaching and Skill Learning*. UK: Blackwell.
- Joreskog, K. G., & Sorbom, D. (2012). LISREL 9.1 [computer software]. Lincolnwood, IL: Scientific Software International.
- Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, 44(02), 137-166.
- Kane, M. J., Conway, A. R., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. *Variation in working memory*, 21-48.
- Kass, R. A., & Tinsley, H. E. A. (1979). Factor analysis. *Journal of Leisure Research*, 11, 120-138.
- Kelly, P. (1991). Lexical Ignorance: The Main Obstacle to Listening Comprehension with Advanced Foreign Language Learners. *Iral*, 29(2), 135-49.
- Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*, 11(02), 261-271.
- Lee, D., & Hatesohl, D. (1993). *Listening: Our most used communication skill*. Retrieved from <http://extension.missouri.edu/publications/DisplayPub.aspx?P=CM150>

- Liao, Y. F. (2007). Investigating the construct validity of the grammar and vocabulary section and the listening section of the ECCE: Lexico-grammatical ability as a predictor of L2 listening ability. *SPAAN FELLOW, 1001*, 37.
- Lin, M. (2006). *The effects of note-taking, memory and rate of presentation on EFL learners' listening comprehension*. Unpublished doctoral dissertation, La Sierra University, California.
- Linacre, J.M. (2016). Winsteps® (Version 3.92.0) [Computer Software]. Beaverton, Oregon: Winsteps.com. Retrieved January 1, 2016. Available from <http://www.winsteps.com/>
- Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic bulletin & review, 21*(4), 861-883.
- Lomax, R. G., & Hahs-Vaughn, D. L. (2013). *Statistical concepts: A second course*. Routledge.
- Long, M. H. (1985). Input and second language acquisition theory. In S. M. Gass, & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 377-393). Rowley, MA: Newbury House Publishers.
- Lund, R. J. (1991). A comparison of second language reading and listening comprehension. *Modern Language Journal, 73*, 32-40.
- Lynch, T. (1998). Theoretical perspectives on listening. *Annual Review of Applied Linguistics, 18*, 3-19.
- MacIntyre, P. D. (1995). How does anxiety affect second language learning? A reply to Sparks and Ganschow. *The Modern Language Journal, 79*(1), 90-99.
- MacIntyre, P. D., & Gardner, R. C. (1991). Language Anxiety: Its Relationship to Other Anxieties and to Processing in Native and Second Languages. *Language learning, 41*(4), 513-534.
- MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language learning, 44*(2), 283-305.

- McDonald, J. L., (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, 55(3), 381–401.
- McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research*, 19 (6), 741-760.
- McNamara, T. F. (1996). *Measuring second language performance* (p. 165). London: Longman.
- Meara, P. (1996a). The dimensions of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Competence and performance in language learning* (pp. 35–53). Cambridge: Cambridge University Press.
- Meara, P. & Milton, J. (2003). *X_Lex, The Swansea Levels Test*. Newbury: Express.
- Mecartty, F. H. (2000). Lexical and Grammatical Knowledge in Reading and Listening Comprehension by Foreign Language Learners of Spanish. *Applied*, 11(2), 323-348.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational researcher*, 23 (2), 13-23.
- Mills, N., Pajares, F., & Herron, C. (2006). A reevaluation of the role of anxiety: Self-efficacy, anxiety, and their relation to reading and listening proficiency. *Foreign language annals*, 39(2), 276-295.
- Milton, J. (2009). *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters.
- Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. *Communicative proficiency and linguistic development: Intersections between SLA and language testing research*, 211-232.
- Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. *L2 vocabulary acquisition, knowledge and use: new perspectives on assessment and corpus analysis*, 57.

- Milton, J., & Alexiou, T. (2009). Vocabulary size and the common European framework of reference for languages. *Vocabulary studies in first and second language acquisition*, 194-211.
- Milton, J. & Hopkins, N. (2006). Comparing phonological and orthographic vocabulary size: do vocabulary tests underestimate the knowledge of some learners. *The Canadian Modern Language Review*, 63(1), 127-147.
- Milton, J., & Treffers-Daller, J. (2013). Vocabulary size revisited: the link between vocabulary size and academic achievement. *Applied Linguistics Review*, 4(1), 151-172.
- Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language . In R. Chacón-Beltrán , C. Abello-Contesse , M. Torreblanca-López , & M. López-Jiménez (Eds.), *Further insights into non-native vocabulary teaching and learning* . Clevedon, UK: Multilingual Matters.
- Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions four general conclusions. *Current directions in psychological science*, 21(1), 8-14.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex frontal lobe tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49–100.
- Mueller, R. O., & Hancock, G. R. (2008). Best practices in structural equation modeling. In Osborne J (Ed.), *Best practices in quantitative methods* (pp. 488-508). Thousand Oaks, CA: Sage.
- Nation, I.S.P. (1983). *Testing and teaching vocabulary*. Guidelines 5: 12–25.
- Nation, I.S.P. (1990). *Teaching and Learning Vocabulary*. Boston, MA: Heinle and Heinle.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening?. *Canadian Modern Language Review*, 63(1), 59-82.

- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The language teacher*, 31(7), 9-13.
- Nation, I.S. P. & Waring. R. (1997). 'Vocabulary size, text coverage and word lists' in N. Schmitt and M. McCarthy (Eds.): *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge University Press, pp. 6–19.
- Nielson, K. B. (2014). Can planning time compensate for individual differences in working memory capacity?. *Language Teaching Research*, 18(3), 272-293.
- Nizonkiza, D. (2011). The relationship between lexical competence, collocational competence, and second language proficiency. *English Text Construction*, 4(1), 113-145.
- Ranta, E. (2009). Syntactic features in spoken ELF–learner language or spoken grammar. *English as a lingua franca: Studies and findings*, 84-106.
- Rapee, R. M. (1993). The utilisation of working memory by worry. *Behaviour Research and Therapy*, 31(6), 617-620.
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language testing*, 10(3), 355-371.
- Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. *Validation in language assessment*, 41-60.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge university press.
- Read, J. (2004). 7. Research in teaching vocabulary. *Annual Review of Applied Linguistics*, 24, 146-161.
- Read, J. (2007). Second language vocabulary assessment: Current practices and new directions. *IJES, International Journal of English Studies*, 7(2), 105-126.
- Real, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, 57(1), 1-23.
- Richards, J. C. (1983). Listening comprehension: Approach, design, procedure. *TESOL quarterly*, 17(2), 219-240.

- Robertson, E. K., & Joanisse, M. F. (2010). Spoken sentence comprehension in children with dyslexia and language impairment: The roles of syntax and working memory. *Applied Psycholinguistics*, 31(01), 141-165.
- Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, 57(3), 348-379.
- Rost, M. (1990). *Listening in Language Learning*. Harlow: Longman.
- Rost, M. (2002). *Teaching and Researching Listening*. London, UK: Longman.
- Rost, M. (2005). L2 listening, in E. Hinkel (ed.) *Handbook of research in second language teaching and learning*. Mahwah, NJ: Erlbaum.
- Rost, M. (2006). Areas of research that influence L2 listening instruction. In E. Uso-Juan and A. Martinez-Flor (Eds.) *Current Trends in the Development and Teaching of the Four Language Skills*. (pp. 47–74). New York: Mouton de Gruyter.
- Rost, M. (2013). *Teaching and researching: Listening*. Routledge.
- Rubin, J. (1994). A review of second language listening comprehension research. *The modern language journal*, 78(2), 199-221.
- Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review*, 56 (2), 282-308.
- Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language learning*, 52 (3), 513-536.
- Qian, D. D. & Schedle, M. (2004). Evaluation of an in-depth vocabulary knowledge for assessing reading performance. *Language Testing*, 21, 28-52.
- Salsbury, T., Crossley, S. A., & McNamara, D. S. (2011). Psycholinguistic word information in second language oral discourse. *Second Language Research*, 27 (3), 343-360.
- Scarcella, R. C., & Oxford, R. L. (1992). *The tapestry of language learning: The individual in the communicative classroom*. Boston: Heinle & Heinle.

- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two new versions of the Vocabulary Levels Test. *Language testing*, 18(1), 55-88.
- Schoonen, R., Hulstijn, J., & Bossers, B. (1998). Metacognitive and language-specific knowledge in native and foreign language reading comprehension: An empirical study among Dutch students in grades 6, 8 and 10. *Language learning*, 48(1), 71-106.
- Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, 14, 369-369.
- Shiotsu, T., & Weir, C. J. (2007). The relative significance of syntactic knowledge and vocabulary breadth in the prediction of reading comprehension test performance. *Language Testing*, 24(1), 99-128.
- Siemer, M., & Reisenzein, R. (1998). Effects of mood on evaluative judgments: Influence of reduced processing capacity and mood salience. *Cognition and Emotion*, 12(6), 783-805.
- Schmitt, N. (2010). *Researching vocabulary*. Basingstoke, England: Palgrave Macmillan.
- Spielmann, G., & Radnofsky, M. L. (2001). Learning language under tension: New directions from a qualitative study. *The Modern Language Journal*, 85(2), 259-278.
- Stæhr, L. S. (2007). *The relationship between vocabulary knowledge and listening comprehension in English as a foreign language*. Copenhagen Business School.
- Stæhr, L. S. (2008). Vocabulary size and the skills of reading, listening and writing. *Language Learning Journal*, 36, 139-152.
- Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, 31 (04), 577-607.
- Statistics, I. I. S. (2011). Version 20. Armonk, NY: International Business Machines Corp.

- Sun, K. C. (2002). Investigation of English listening difficulties of Taiwan students. In *A paper presented at the Eleventh International Symposium on English Teaching—English Teachers' Association/ROC, Taipei.*
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Needham Heights, MA: Allyn and Bacon.
- Tatsuoka, K. K. (2009). *Cognitive assessment: An introduction to the rule space method.* Rutledge.
- Teng, H. C. (2001). Impact of Speech Modification on EFL Listening. *Proceedings of the PAC3 at JALT 2001*, 22-25.
- Thompson, I. (1995). Testing listening comprehension. *AATSEEL Newsletter*, 37, 24-31.
- Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules* (Vol. 1950). Cambridge, MA: MIT Press.
- Tyler, M. D. (2001). Resource consumption as a function of topic knowledge in nonnative and native comprehension. *Language Learning*, 51(2), 257-280.
- Urquhart, A. H., & Weir, C. J. (1998). *Reading in a second language: Process, product, and practice.* New York: Longman.
- Vandergrift, L. (1996). Listening Comprehension Strategies of Core French High School Students. *Canadian Modern Language Review*, 52(2), 200-23.
- Vandergrift, L. (1997). The strategies of second language (French) listeners: A descriptive study. *Foreign Language Annals*, 30, 387-409.
- Vandergrift, L. (2002). "It was nice to see that our predictions were right": Developing metacognition in L2 listening comprehension. *Canadian Modern Language Review*, 58, 555-75.
- Vandergrift, L. (2003a). Orchestrating strategy use: Towards a model of the skilled L2 listener. *Language Learning*, 53, 461-494.
- Vandergrift, L. (2003b). From prediction to reflection: Guiding students through the process of L2 listening. *Canadian Modern Language Review*, 59, 425-440.

- Vandergrift, L. (2004). Learning to listen or listening to learn. *Annual Review of Applied Linguistics*, 24, 3–25.
- Vandergrift, L. (2005). Relationships among motivation orientations, metacognitive awareness and proficiency in L2 listening. *Applied Linguistics*, 26, 70–89.
- Vandergrift, L. (2006). Second language listening: listening ability or language proficiency?. *The Modern Language Journal*, 90(1), 6-18.
- Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language teaching*, 40 (03), 191-210.
- Vandergrift, L. (2011). Second Language Listening. *Handbook of research in second language teaching and learning*, 2, 455.
- Vandergrift, L., & Baker, S. (2015). Learner Variables in Second Language Listening Comprehension: An Exploratory Path Analysis. *Language Learning*, 65(2), 390-416.
- Vandergrift, L., & Goh, C. M. (2012). *Teaching and learning second language listening: Metacognition in action*. New York: Routledge.
- Vandergrift, L., Goh, C., Mareschal, C., & Tafaghodtari, M. H. (2006). The Metacognitive Awareness Listening Questionnaire (MALQ): Development and validation. *Language Learning*, 56, 431–462.
- VanPatten, B. (2002). Processing instruction: An update. *Language Learning*, 52, 755–803.
- van Zeeland, H., & Schmitt, N. (2012). Lexical coverage in L1 and L2 listening comprehension: the same or different from reading comprehension?. *Applied Linguistics*, 34(4), 457- 479.
- Vogely, A. J. (1998). Listening comprehension anxiety: Students' reported sources and solutions. *Foreign Language Annals*, 31(1), 67-80.
- Ying-hui, H. (2006). An investigation into the task features affecting EFL listening comprehension test performance. *The Asian EFL Journal Quarterly*, 8(2), 33–54.

- Wayland, S., Saner, L., Ralph, A., O'Connell, S., Linck, J., Kramasz, D., & Gynther, K. (2013). *The Long and the Short of It: Passage Length and Information Density in Second Language Listening Comprehension* (Technical Report No. TTO 81434). Retrieved from the University of Maryland Center for Advanced Study of Language website:
<http://www.govtilr.org/Publications/Length,%20Density,%20and%20Working%20Memory.pdf>
- Wesche, M., & Paribakht, T. S. (1996). Assessing Second Language Vocabulary Knowledge: Depth Versus Breadth. *Canadian Modern Language Review*, 53 (1), 13-40.
- Widdowson, H. (2003). *Defining issues in English language teaching*. Oxford: Oxford University Press.
- Winke, P., & Gass, S. (2013). The Influence of Second Language Experience and Accent Familiarity on Oral Proficiency Rating: A Qualitative Investigation. *TESOL Quarterly*, 47(4), 762-789.
- Wright, B. D., Linacre, J. M., Gustafson, J. E., & Martin-Lof, P. (1994). Reasonable mean-square fit values. *Rasch measurement transactions*, 8(3), 370.
- Young, M. Y. C. (1997). A serial ordering of listening comprehension strategies used by advanced ESL learners in Hong Kong. *Asian Journal of English Language Teaching*, 7, 35-53.
- Zareva, A., Schwanenflugel, P., & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency: Variable sensitivity. *Studies in Second Language Acquisition*, 27(04), 567-595.