Individual differences in listening comprehension among adult Chinese learners of English

Yun Wang

Institute of Education
University of Reading

Thesis submitted for the degree of
Doctor of Philosophy

December 2018

Acknowledgements

My greatest gratitude goes to my first supervisor Professor Jeanine Treffers-Daller and my second supervisor Professor Suzanne Graham, for their enlightening instruction, stimulating comments, insightful guidance and above all their patience in instructing me to complete this PhD thesis. My supervisors provided great help in my applying for the Reading University Scholarship before I started the PhD study. They have also given me instructions to make presentations on either international conferences or PhD student conferences. They have encouraged me to take training courses out of my interests which were provided by Reading University or other universities so that I can improve myself overall academically. Without their help and encouragement, I would not have completed the PhD study. They not only provide inspiring suggestions on my study, but also offer prompt help in my life. It is their help and support that makes me feel confident in that I can complete the study. During the years of my PhD study, I have learned from my supervisors how to work as a professional researcher, a respectable scholar, and a knowledgeable teacher. I am very lucky to have them to be my supervisors.

I would like to express my thanks to the Graduate School and the Institute of Education, Reading University for all the prompt help from the staff during my study, e.g., Ms. Rachel Willis, Ms. Helen Apted, Dr. Holly Joseph, and Dr. Daisy Powell.

I would also like to extend my sincere appreciation to Dr. Stuwart McLean and Dr. Joshua Matthews for their generously agreeing to share the tests and the recording designed by them and used in their studies. My special thanks also go to Professor Hongchen Xu and Professor Lars-Erik Malmberg for their helpful advice on analysing data.

I am grateful to my colleagues for their support and help in my PhD study, and to all the students who participated in my study for their cooperation. I would like to thank Dr. Daguo Li, Dr. Xiaorong Zhang, Dr. Pengchong Zhang and Dr. Shi Zhang for their sincere help in my study and for their encouragement to overcome difficulties in my life. My thanks go to Professor Lynda O’Brien for her proofreading of my thesis.

Last but not the least, I am greatly indebted to my family, my husband Dr. Guanghui Tang and my daughter Miss Yingying Tang for their constant love and support that encourage me to go through those hard days. It is their love that motivates me to keep on going ahead and to be always confident in completing this study! I love them!

Abstract

The thesis focuses on the individual differences which can explain the variability in listening comprehension among adult Chinese learners of English. Although there is now a large number of studies on listening comprehension among second language (L2) learners, individual differences in listening have received less attention than individual differences in reading. The thesis aims to fill some of the gaps in our knowledge in this field of research.

Listening comprehension was measured with two different tests (the College English Test Band 4 listening section and the Cambridge Preliminary English Test listening section). Four groups of explanatory variables were included in the study: linguistic knowledge, sentence processing speed, cognitive factors and learners' use of English in daily life. Structural equation models of listening comprehension were built based on Andringa, Olsthoorn, Van Beuningen, Schoonen and Hulstijn (2012). This model was tested among 187 Chinese learners of English (one group in China, *N* = 147; the other group in the UK, *N* = 40).

The results indicate that the listening comprehension of learners in China was significantly lower than that of Chinese learners in the UK. Linguistic knowledge, frequency of English use in daily life and aural sentence processing speed were key predictors of listening in these groups. Phonological knowledge was the most important predictor of listening comprehension among the variables measuring linguistic knowledge and word recognition from speech explained variance in listening comprehension over and above the contribution of word segmentation from speech. When the two groups of learners were considered separately word recognition from speech was the most important predictor for learners in China whilst for learners in the UK, learners' grammar knowledge and the reasoning ability were key. Finally a comparison of both listening tests revealed that the listening section of the CET4 only measures learners' ability to comprehend information explicitly available in the text and not inferencing skills. Pedagogical implications for teachers and learners of English and test developers in China and in the UK are provided based on these findings.

Declaration of original authorship

I confirm that this is my own work and the use of all other materials has been properly and fully acknowledged.

Print name: Yun Wang

Date: 13/12/2018

Table of Contents

List of Figures

LIST OF TABLES

List of ABBREVIATIONS

PET	the Preliminary English Test
CET4	the College English Test Band 4
WST	the Word Segmentation Test
OPT	the Oxford Placement Test
WRS	the Test of Word Recognition from Speech
LVLT	the Listening Vocabulary Levels Test
TROG-2	the Test for Reception of Grammar version 2
SPS	Sentence Processing Speed Test
WMCF	Working Memory Forward Task in Chinese
WMCB	Working Memory Backward Task in Chinese
WMFE	Working Memory Forward Task in English
WMEB	Working Memory Backward Task in English
Raven's SPM	Raven's Standard Progressive Matrices
PBQ	Personal Background Questionnaire.

Introduction

Listening comprehension (LC) is a very complicated process because it involves integrating linguistic and non-linguistic knowledge during online processing. Linguistic knowledge covers phonology, morphology, vocabulary, syntax, semantics, and discourse structure, whereas non-linguistic knowledge includes knowledge of the contexts in which interactions take place and of specific facts (Buck, 2001). For second language (L2) learners and users, i.e., those who use a second language in their daily work or studies, this process is likely to be more difficult than for native speakers of a language. L2 learners and users generally have lower levels of competence in the language and are less familiar with the contexts within which they receive input.

As pointed out by Andringa, Olsthoorn, Beuningen, Schoonen, and Hulstijn (2012), there is now a considerable body of literature on LC among non-native speakers, but explaining individual differences in LC has not received substantial attention compared with studies which look into individual differences in reading comprehension. A better understanding of individual differences which determine L2 learners' success in comprehending speech input is therefore urgently needed because L2 learners find it hard to make progress in improving their LC (Graham, 2011).

It is particularly relevant to study LC among Chinese learners of English because Chinese learners find it very hard to understand English speakers (Goh, 2000; Y. Wang, 2008). This is possibly due to typological differences between Chinese and English (see Trenkic & Warmington, 2018, for an overview). The few available studies among Chinese learners are based on interviews or questionnaires which investigate what learners or teachers perceive to be the learners' listening problems. However, it is possible that there is a discrepancy between self-reported problems in listening and real listening problems in online listening processes. So far, very few empirical studies have been conducted to explore Chinese learners' online listening processes. Thus, it is very important to conduct an empirical study to identify the listening problems that Chinese learners encounter in real listening processes so that any interventions aimed at addressing these issues can be based on solid empirical evidence. The issue is particularly important for Chinese university students in the UK as their academic achievement is generally lower than that of students whose native language is English, although their non-verbal ability is the same (Trenkic & Warmington, 2018). The lower levels of competence in English among the Chinese students in Trenkic and Warmington's study explained 51% of the variance in students' grades on their university courses.

The available evidence shows that there are important differences between native and non-native speakers with respect to LC. Andringa et al. (2012) built a structural equation model of LC among native and non-native speakers of Dutch, with three groups of factors: learners' linguistic knowledge; their processing speed; and their cognitive ability. They found that for native speakers, linguistic knowledge and processing speed explained 91% of the variance in LC; while for L2 learners linguistic knowledge and reasoning ability explained 96% of such variance. As their model of LC was exceptionally successful in explaining variance in L2 LC, the current study was set up as a partial replication study of Andringa et al. (2012) with the aim of investigating whether a similar model could account for LC among Chinese learners of English.

As pointed out by Andringa et al. (2012), in addition to linguistic knowledge, the efficiency with which this knowledge can be applied (processing accuracy and speed) and general cognitive ability (working memory and general cognitive ability) may explain some of the variance in LC. The current study, therefore, also explores to what extent these three groups of factors can explain LC among adult Chinese learners of English. The design differs from that of Andringa et al. in that a variable which measures differences in usage of English is included. This was done because it was suggested in their discussion that this variable might explain the overall differences in LC between native speakers and L2 learners. To be able to study the impact of language use on listening in more detail, the current study compares Chinese learners in the UK with Chinese learners of English in China; the former are likely to have many more opportunities for using English than the latter, and therefore, there are likely to be important differences between the groups with respect to their ability to understand spoken English. Currently, a limited number of studies have explored the contribution of immersion in an L2 environment to the prediction of L2 learners' LC. Therefore, such empirical studies are greatly needed to fill the gap in our understanding of LC among non-native speakers.

My interest in carrying out the research is the result of my experience with the teaching of LC to Chinese learners of English. During my teaching, it often happened that my students raised the same issues related to their frustrations with English LC. Having learned English for so many years and having spent a lot of time and energy on English listening, they wondered why their LC remained so limited, why it was so difficult to improve their English

LC, or how they could improve this skill. Lack of exposure to authentic spoken English language might have been one of the reasons which led to their limited use of English in daily activities. However, there must be some other individual differences between those Chinese learners of English which explain differences in their ability to understand oral English and which impact their performance on English LC. As a teacher of English in China, and also as a researcher, I wonder which variables may explain success in LC among Chinese learners of English.

This thesis aims to address this and is organised into seven chapters. Chapter 1 presents the L1 LC processes, the similarities and differences between listening and reading comprehension, and the L2 LC processes. In Chapter 2, firstly, I review the literature on learner variables which affect listening comprehension in L2 learners. Then I review the literature on issues faced by Chinese learners of English in LC, present the aims of the current study, and pose the research questions the study aims to answer. Chapter 3 gives an overview of the methodology of the study, including its design, why specific participants are chosen, and what instruments are adopted to explore the research questions. Attention is also paid to the data collection procedure and preliminary data analyses, and ethical issues are discussed. Chapter 4 presents the three pilot studies conducted prior to the main study data collection. Chapter 5 presents, firstly, the results of reliability and means of each task, the correlations between each observed variable and the dependent variables for both the entire group and the subgroups, and the listening proficiency of the subgroups. Then, Section 5.2 presents the results of hierarchical regression analyses on listening comprehension models with each dependent variable. Next, Section 5.3 presents the results of assumptions for testing the hypothesized structural equation models of listening comprehension and the results of testing the models. In the final section, the research questions are answered. In Chapter 6, firstly, I compare studies on word segmentation from the speech stream. Then the models found in the current study and the other studies are compared. Next, the lower-level processes in L2 listening comprehension adopted in the current study and the evidence for Chinese learners' listening problems are discussed. In this chapter, the learner variables which can explain the success in L2 listening comprehension for the entire group and the learner variables which can explain why the listening proficiency in the two subgroups differs are also discussed. In the last section of this chapter, the two listening tests which are used to measure learners' listening proficiency in the study are compared. In the last chapter, the main findings of the current study and its limitations are summarised. The implications for pedagogy in the UK and in China, and the implications for test developers in HE in China are provided.

Chapter 1 L1 and L2 listening comprehension

Listening comprehension (LC) is a very complicated process because it involves integrating linguistic and non-linguistic knowledge during online processing. According to Cutler and Clifton (1999) in the processes of comprehension, L1 listeners have to use linguistic knowledge, particularly phonological knowledge, to process speech input. Since listening and reading comprehension are both receptive processes, they have some similarities. However, since listening involves processing acoustic input and reading involves processing visual input, there are also clear differences between the two types of comprehension.

This chapter is comprised of three sections. Section 1.1 presents the processes involved in L1 listening comprehension and is followed by discussion of the similarities and differences between listening comprehension (LC) and reading comprehension. In the final section, processes involved in L2 listening comprehension are presented.

1.1 L1 listening comprehension

LC is a complex process in which both linguistic and non-linguistic knowledge are involved. Of linguistic knowledge, the most important types are phonology, lexis, syntax, semantics, and discourse structure (Buck, 2001). Non-linguistic knowledge involves general knowledge about the world and knowledge about topics and contexts (Buck, 2001). LC is a result of interaction between aural input, types of linguistic knowledge, and general world knowledge. Listeners make use of helpful information to interpret speech input. When they receive it, they have to recognise acoustic signals, segment phonemes and words, and construct meaning both from words and from the broader discourse. In the process of comprehension, listeners have to pay attention to stress and intonation patterns of the language which may convey important information. For example, level tones are often used to show that additional information is about to be revealed, and falling intonation is often used to identify new information (Rost, 2016). Not only do listeners need knowledge of the language, but they also need the ability to apply the language to understanding words, processing idea units, and processing connected discourse. In addition to language knowledge, listeners also need to make use of world knowledge, past experience, future intentions, and intelligence, to interact with speech input and create text interpretation (Buck, 2001).

According to Anderson (1995), three stages are involved in LC: perception, parsing, and utilization. In terms of the processes of these stages, perception, or speech recognition, means to match aural input with the phonological and vocabulary system of the input language. In this stage, listeners may encounter problems of segmentation which are likely to happen either between phonemes within a word, or between words because speech is a stream of sounds and there are no obvious boundaries between words (Anderson, 1995). The second stage, parsing, is a process by which listeners transform words from speech input into a mental representation of combined word meaning. In the parsing process, listeners comprehend input with reference to semantic and syntactic structures. In the third stage, utilization, listeners take advantage of the mental representation of a sentence's meaning to comprehend speech input. At this stage, LC requires backward and forward inferences and connections (Anderson, 1995). Since perception is the base of comprehension, Cutler and Clifton (1999) developed a model of L1 LC which focuses on perception of aural input and shows how listeners decode this input with the help of linguistic knowledge (see Figure 1.1).

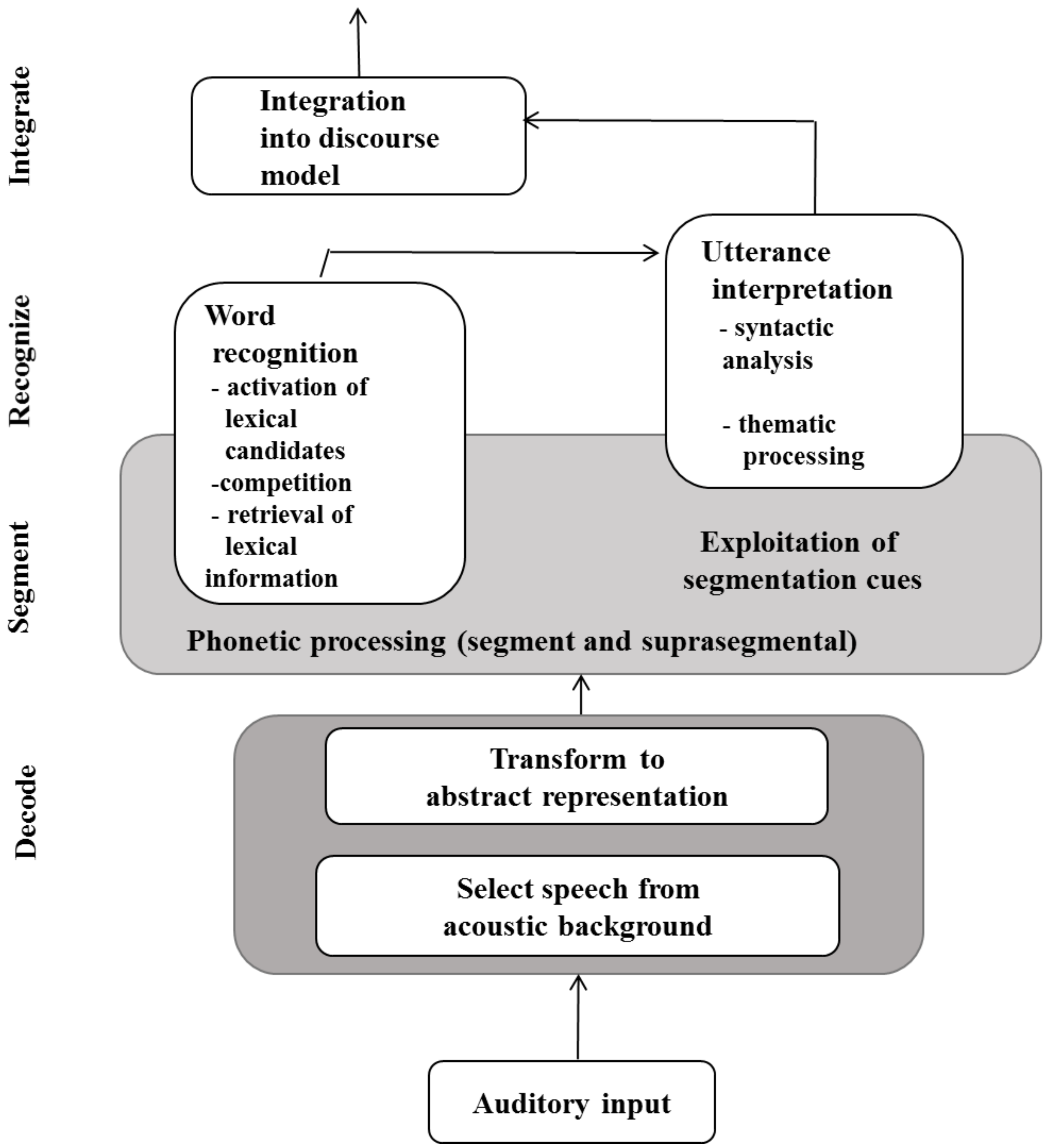


Figure 1.1 Cognitive model of LC processing (Cutler & Clifton, 1999, p. 124)

Cutler and Clifton (1999, p. 124) assume there are four stages in the process of LC:

- decoding of speech;
- segmenting continuous words and syntactic boundaries;
- recognising words and interpreting them syntactically and thematically;
- integrating words, syntactic analysis, and thematic processing into a discourse model.

In this model, Cutler and Clifton (1999) define decoding as transforming a time-varying input into a representation which consists of discrete elements. Linguists regard phonemes as the smallest elements in spoken language and speech consists of a series of phonetic segments. For example, the sound of the word *key* differs from the sound of the word *sea* because the former is comprised of the segments */k/* and */i/*, while the latter is comprised of the segments */s/* and */i/*. It is a crucial step in LC to decode a phonetic segment correctly because otherwise an incorrect interpretation will follow. For example, when a listener hears the sound sequence /maiˈtrein/, if it is decoded as the sounds /mai/ and /trein/, the sound will match *my train*; if it is decoded as the sounds /mait/ and /rein/, the sound will match *might rain* (Field, 2008). According to Cutler and Clifton (1999), word segmentation takes place during word recognition and utterance interpretation processes. In their model, suprasegmental structure refers to stress, intonation patterns, and pitch accent. It represents a level of organisation which is above the segmental level. English listeners make use of rhythm to segment spoken words because English rhythm is based on the contrast of stressed and unstressed syllables (Brown, 1990). Rhythmic beat consists of stressed syllables; meanwhile unstressed syllables occur between stressed syllables and are compressed as far as possible so that the next stressed syllable falls on the regular beat (Brown, 1990). In English, strong syllables are those which contain full vowels, e.g., *eye, pill*, *crypt*, and *scrounge* are all strong monosyllabic words; weak syllables are those which contain reduced vowels, e.g., the second syllable in *ion*, *scrounges*, *pillow,* and *cryptic* are weak syllables (Cutler & Norris, 1988). According to Cutler and Clifton (1999), another difficulty facing listeners attempting to recognise words is related to the temporal aspect of spoken words. Spoken words are presented over time: listeners hear the beginning first and the end last. The pronunciation of the sounds of the spoken word may activate lexical presentations of different words. For example, hearing the spoken word *steak* may activate words with similar pronunciations, e.g., *snake*, *stay,* or *stack*. Apart from identifying individual spoken words and their meanings, listeners also have to depend on a sequence of spoken words and on how a sentence is interpreted semantically when recognizing spoken words.

Since Cutler and Clifton (1999) focus on processing of the L1, it is not clear whether their model is suitable for L2 learners. It is likely that listening is more difficult for L2 learners than for native speakers of a language because L2 learners have difficulties with sound perception, recognition of words and comprehending the meaning they convey (Broersma & Cutler, 2011). Generally, L2 learners also have poorer lexical and syntactic knowledge than native speakers. The specific difficulties facing non-native listeners may vary depending on the listeners' first language and their L2, e.g., the difficulties faced by Chinese learners of English, which is the subject of the current study, might differ from learners of other L1s.

1.2 Similarities and differences between listening and reading comprehension

Listening and reading comprehension are both receptive processes. They both engage lower-order (bottom-up) processes which refer to the use of linguistic knowledge in the comprehension of written/aural input. They also both involve higher-order (top-down) processes in which knowledge sources are applied to the comprehension of language input (Field, 2008; Vandergrift & Baker, 2015) and to meaning-construction (Lee & Schallert, 1997; McLaughlin & Heredia, 1996). Higher-order processing refers to use of other sources of knowledge in reading/listening comprehension (e.g., world knowledge, past experience). Readers as well as listeners need to extract ideas and relate them to what has gone before, read between lines/adopt inferencing strategies to discover underlying meaning, and connect reading text/listening text to world knowledge. Furthermore, both listening and reading are influenced by factors such as L2 language proficiency and working memory (Roberts, 2012), and metacognition and motivation (Grabe, 2009; Vandergrift, 2005). Winne and Perry (2000) define metacognition as "the awareness learners have about their general academic strengths and weaknesses, cognitive resources they can apply to meet the demands of particular tasks, and their knowledge about how to regulate engagement in tasks to optimize learning processes and outcomes" (p. 533).

However, since listening involves processing acoustic input and reading involves processing visual input, there are also clear differences between the two types of comprehension. Compared with reading, listening is more demanding for a number of reasons. Firstly, while reading, readers have the advantage of a standardised spelling system in which blank spaces between each of the words help readers decide where one word ends and the next word begins. For listeners, by contrast, speech sounds vary from one utterance to another, even from one speaker to another, due to differences in accents, assimilation, and elision (Field, 2008). Assimilation is a phonological process which leads to sounds becoming more similar to each other in articulation or voice. In *light grey*, for example, the alveolar [t] is often assimilated to the velar [g] and becomes similar to a [k], so that the expression is sounded as *like grey*; in *white board* the alveolar [t] is often assimilated to the bilabial [b] and becomes similar to [p], so that the expression is sounded as *wipe board* (Field, 2008). Elision most commonly affects instances of [t] and [d] at the ends of words. For example, in *nex(t) spring* the [t] is often deleted so that speakers pronounce this expression as *[nek`sprɪŋ]*, and in *cou(ld) take* the sound [d] is often elided so that the expression is sounded as *[ku`teik]* (Field, 2008). A second reason why listening differs from reading is that readers can go back to check word recognition and overall understanding. This is almost impossible when listening because acoustic input is mostly transient and unfolds in time, which means that listeners cannot control the speed of input and have to hold more information in working memory (Field, 2008; Vandergrift & Baker, 2015). Therefore, working memory becomes a more important variable in listening than in reading (Vandergrift & Baker, 2015). Moreover, a reader's comprehension is based on well-organized sentences and structures because writers have enough time to plan what they are going to write and have time to choose correct words and structures. This is not the same for listeners. All speakers, in all circumstances use stress/intonation to convey meaning and often produce informal conversation which contains incorrect pronunciations, hesitations, and/or repetitions. Therefore, listeners must pay attention to prosodic features which may carry important communicative information (Vandergrift & Baker, 2015).

1.3 L2 listening comprehension

In order to show L2 listeners' speech decoding processes, Field (2008) provides an information-processing model which is similar to Cutler and Clifton's (1999) model (see Figure 1.2). In his model, Field (2008) makes use of the sentence *Do you speak English* to show the sequence of how a listener receives a series of speech input and then develops comprehension from the smallest sounds (phonemes) to syllables, words, chunks, and sentences. As can be seen in Figures 1.1 and 1.2, the two models share common processes of LC, although they use different terms. For example, in both models, there are processes of recognising phonemes, recognising syllable structure and stress, segmenting words, syntactic parsing (i.e. assigning a syntactic structure to a sentence) and forming decoding hypotheses. Differences between the two models are that Cutler and Clifton (1999) distinguish four levels which move from decoding to segmenting to recognising to integrating, but Field (2008) distinguishes seven levels which begin at phoneme level through syllable, word-form, chunk, syntax, intonation, and meaning levels. Since Field (2008) shows development of LC from smaller to bigger units, he separates *the word recognition level* in Cutler and Clifton's (1999) model into word-form and chunk levels. While intonation and stress are grouped into the suprasegmental process in Cutler and Clifton, they are grouped into a separate intonation level in Field (2008). Although both models reflect lower-level processes and the focus is on using linguistic knowledge sources, processing is different in L1 and L2. Rost (2016) points out that although bottom-up processing in L1 and L2 follows the same perceptual and decoding procedures, processing in the L2 seldom operates as smoothly as in the L1, especially over long stretches of listening. First-language listeners process speech input automatically and unconsciously, but processing is not automatized to the same degree in L2 listeners. Since L2 learners rarely develop the same high level of ability as in their L1 and, in most cases, L2 listeners have gaps in L2 knowledge, such gaps have significant impact on L2 LC (Buck, 2001). It is noted that there are major differences between novice listeners and expert listeners in that expert listeners command a set of highly automatic decoding routines and can match sounds in the input with words in their L2 vocabulary accurately, rapidly, and effortlessly (Field, 2008). Field (2008) argues that the decoding processes reflect L2 learners' need to familiarise themselves with unfamiliar characteristics of the L2 (e.g., its phonology, word forms, and grammatical structures). However, it is still not clear how listeners use L2 linguistic knowledge to process aural input at the perception stage of LC.

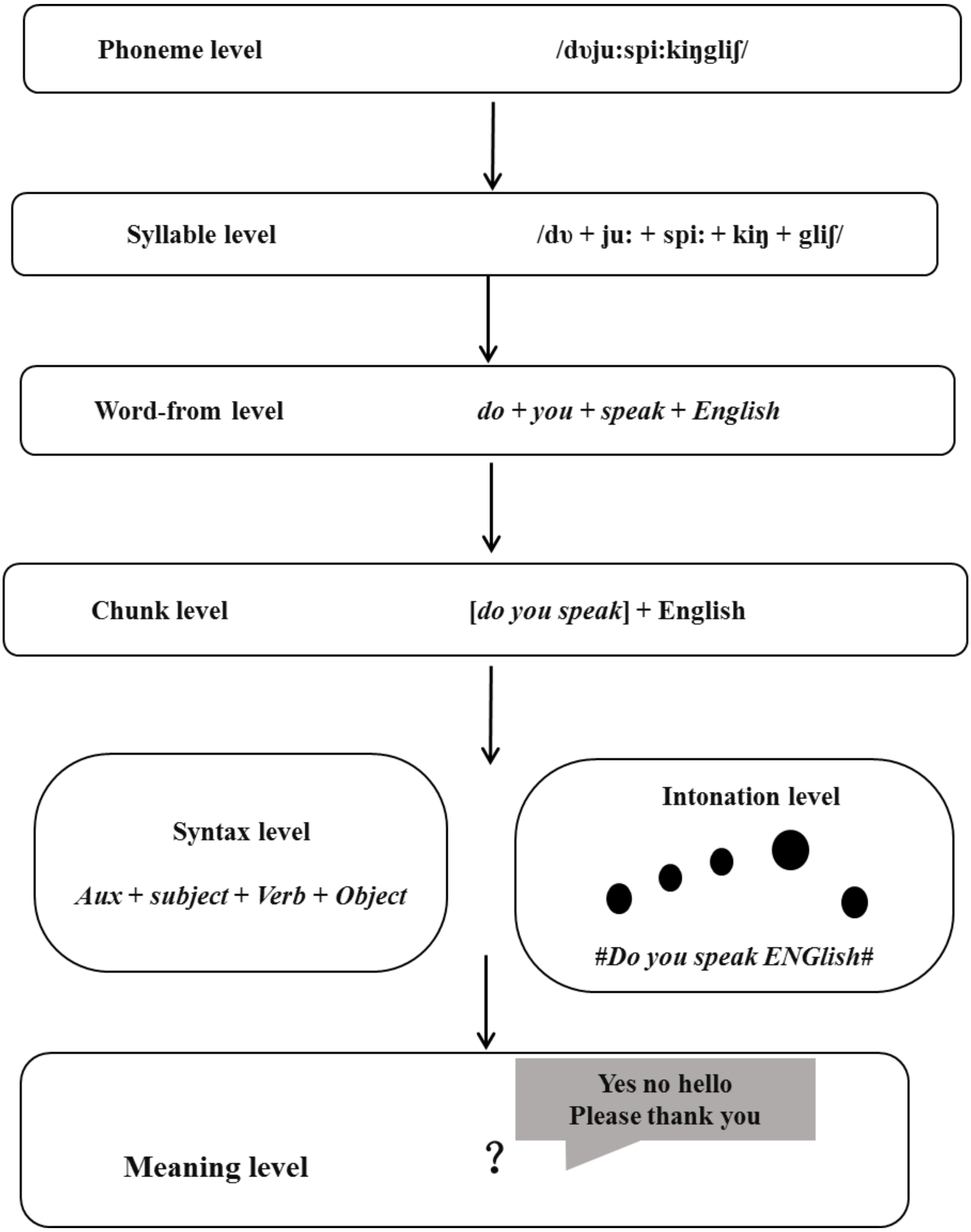


Figure 1.2 Levels of representations of a simple utterance (Field, 2008, p. 114)

Field (2013) put forward a new model of L2 LC which includes a model of lower-level processes and two models of higher-level processes in LC. This new model, which incorporates elements of Cutler and Clifton (1999) and Field (2008), has been adopted in the current study. According to Field (2013), this model is in accordance with Anderson (2000) as it covers the same three operations: *decoding*, *parsing* and *utilisation*. However, Field's new model also differs from Anderson's model in that there are now five distinct levels: the *decoding* level in Anderson's model is divided into an *input decoding* level and a *lexical search* level, and the *utilisation* level in Anderson (2000) is divided into the *meaning construction* level and the *discourse construction* level. The *parsing* level has not been subdivided any further. The first three levels in the model (input decoding, lexical search and parsing) belong to the lower-level processes in L2 LC (see Figure 1.3). Field (2013) assumes that the meaning construction level and the discourse construction level belong to the higher-level processes in L2 LC. These two levels are represented in Figure 1.4 (*meaning construction*) and Figure 1.5 (*discourse construction*). In this thesis, the terms "lower-order processes" and "lower-level processes" are used interchangeably. The terms "higher-order processes" and "higher-level processes" are also used interchangeably.

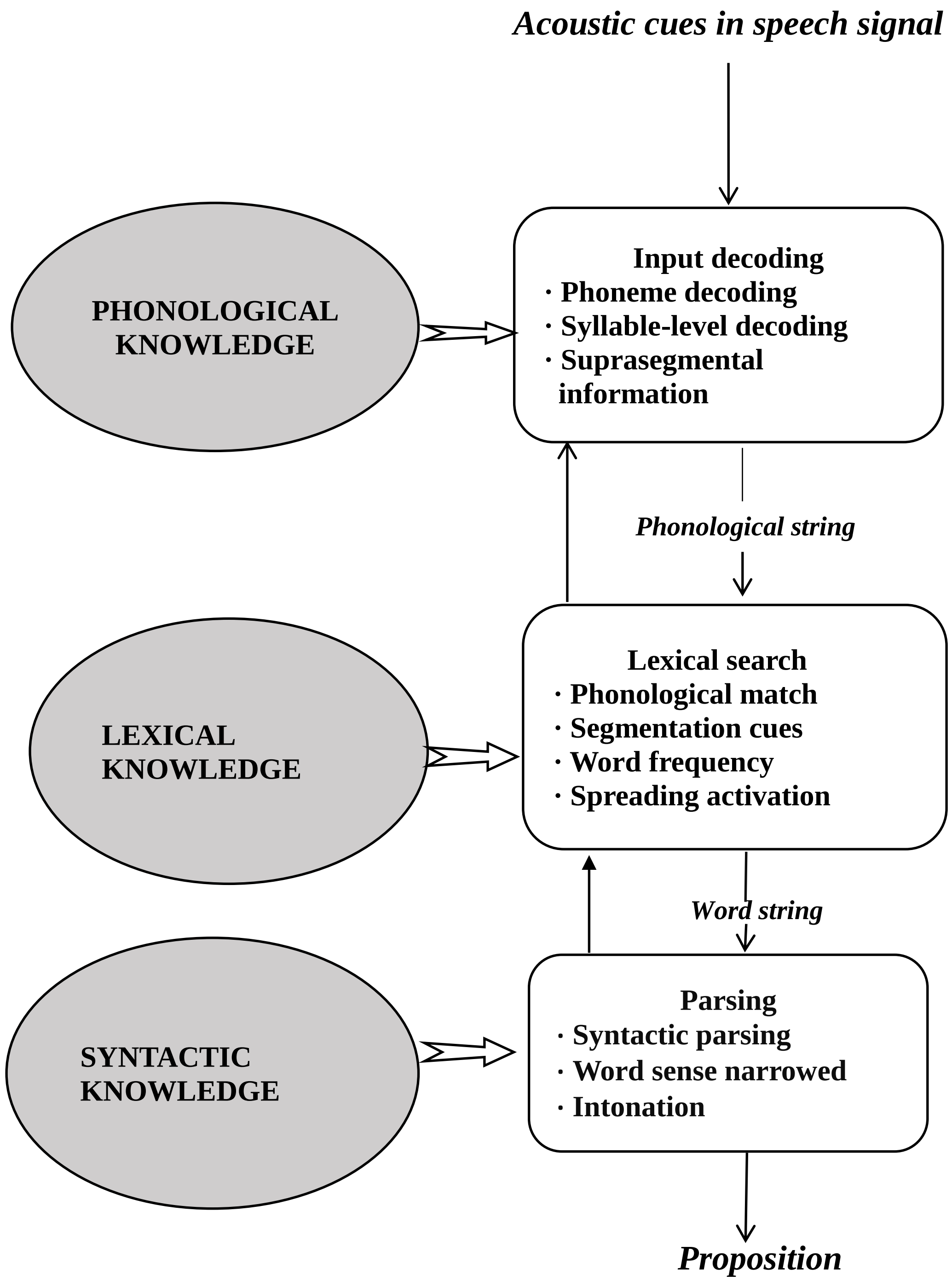


Figure 1.3 Model of lower-level processes in L2 LC (Field, 2013, p. 97)

According to Field (2013), in the model of lower-level processes (see Figure 1.3) the input decoding level also corresponds to the *decode* stage in Cutler and Clifton (1999) and the *phonological level* and the *syllable level* in Field (2008). At this level listeners make use of their phonological knowledge to help them transform speech into syllables. With the phonological string, listeners combine perception information and word segmentation cues to search for words. Listeners' lexical knowledge helps them make the best word match for what they hear. This is the lexical search level. The level corresponds to the *segment* stage and the lexical part of the *recognise* stage in Cutler and Clifton (1999), and to the *word-form* level and the *chunk* level in Field (2008). With phonological string and word string, listeners use their syntactic knowledge at the parsing level to help specify precise words and impose a syntactic pattern for LC. The parsing level is in accordance with the syntactic part of the *recognise* stage in Cutler and Clifton (1999) and with the *syntax* and *intonation* levels in Field (2008). Field (2013) argues that the new lower-level processes of LC share some features with Cutler and Clifton (1999). He divides the *decoding* level into two levels, as in Cutler and Clifton (1999): the first relates to the *phonological* level and the second to the *lexical* level. The new lower-level processes in LC have tentative features of LC with two-directional structures between input decoding, lexical searching and parsing. This differs from the unidirectional structures in Cutler and Clifton (1999) and Field (2008). According to Field (2013), L2 LC should be a tentative process in which listeners have to constantly update and revise their provisional hypotheses, at word, phrase, and clause levels, with the continuous speech input.

Although lower-order processing is important in L2 LC, successful comprehension is a result of interaction between lower-order processing and higher-order processing. Higher-order/top-down processing means the use of non-linguistic knowledge in the processes of LC. According to Buck (2001), L2 listeners have more problems than L1 listeners in LC due to insufficient linguistic and socio-cultural contextual knowledge. L2 listeners' visual information, general background knowledge, common sense, and their knowledge about the context of communication can help compensate for gaps in their linguistic knowledge in LC (Buck, 2001). Higher-level processing is associated with building meaning (meaning construction and discourse construction) (Field, 2013). It relies on using previous knowledge in processing a message rather than on individual sounds and words (Flowerdew & Miller, 2005). Field (2013) provides a model of meaning construction in L2 LC (see Figure 1.4) and a model of discourse construction in L2 LC (see Figure 1.5). The two models represent higher-level processing of listening which corresponds to the *utilization* stage in Anderson (2000). In the lower-level processing of listening, a proposition is produced as a result of decoding and parsing of speech input. Then, in the process of meaning construction, it is the listener's task to connect the proposition with the circumstances in which it was produced and to infer the speaker's intention so that meaning representation can be produced (Field, 2013). Clahsen and Felser (2006) argue that there are fundamental differences between language processing in native speakers and (adult) non-native speakers because non-native speakers make use of syntactic representations which are shallower and less detailed than those of native speakers. As can be seen in Figure 1.4, three types of knowledge source are assumed to support meaning construction in the processes of LC. The first of these is pragmatic knowledge, i.e., knowledge of the relationship between a listener's knowledge of pragmatic forms in the L2 and a speaker's intentions. External knowledge includes world knowledge, the speaker's knowledge, and situational knowledge. Finally, discourse representation is a listener's recall of the listening event from speech input and is possibly related to the listener's awareness of the current topic. As can be seen in the model (see Figure 1.4), in a process of meaning construction a listener has to not only make use of the three types of knowledge sources to comprehend a speaker's intentions, but also needs to use inference to comprehend what a speaker intends. Up to this point the listener's LC of the speech input has not ended; now the listener needs to make judgements about the information obtained and relate that to their perception of both the speaker's and their own, the listener's, goals. This level is discourse construction (see Figure 1.5).

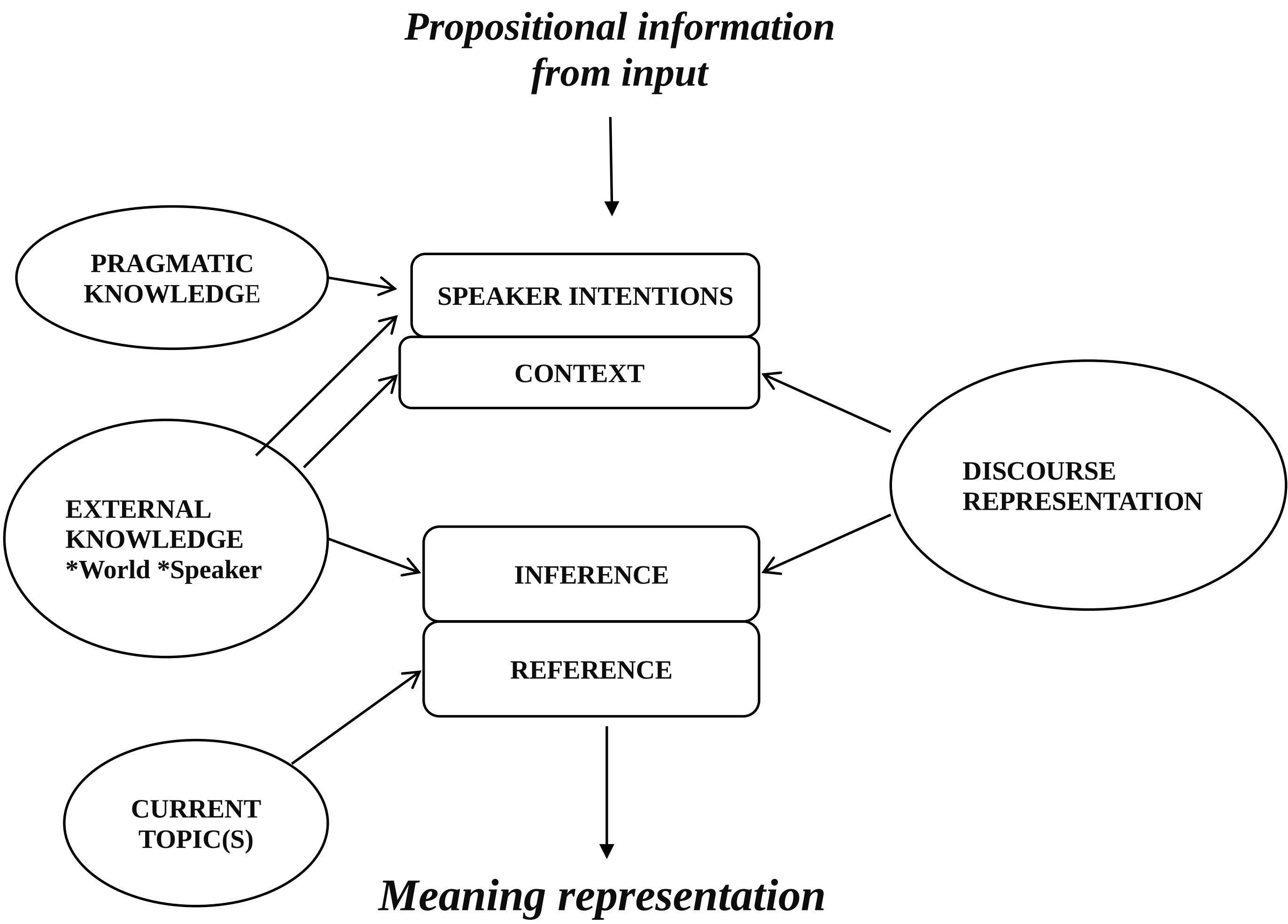


Figure 1.4 Model of meaning construction in listening (Field, 2013, p. 101)

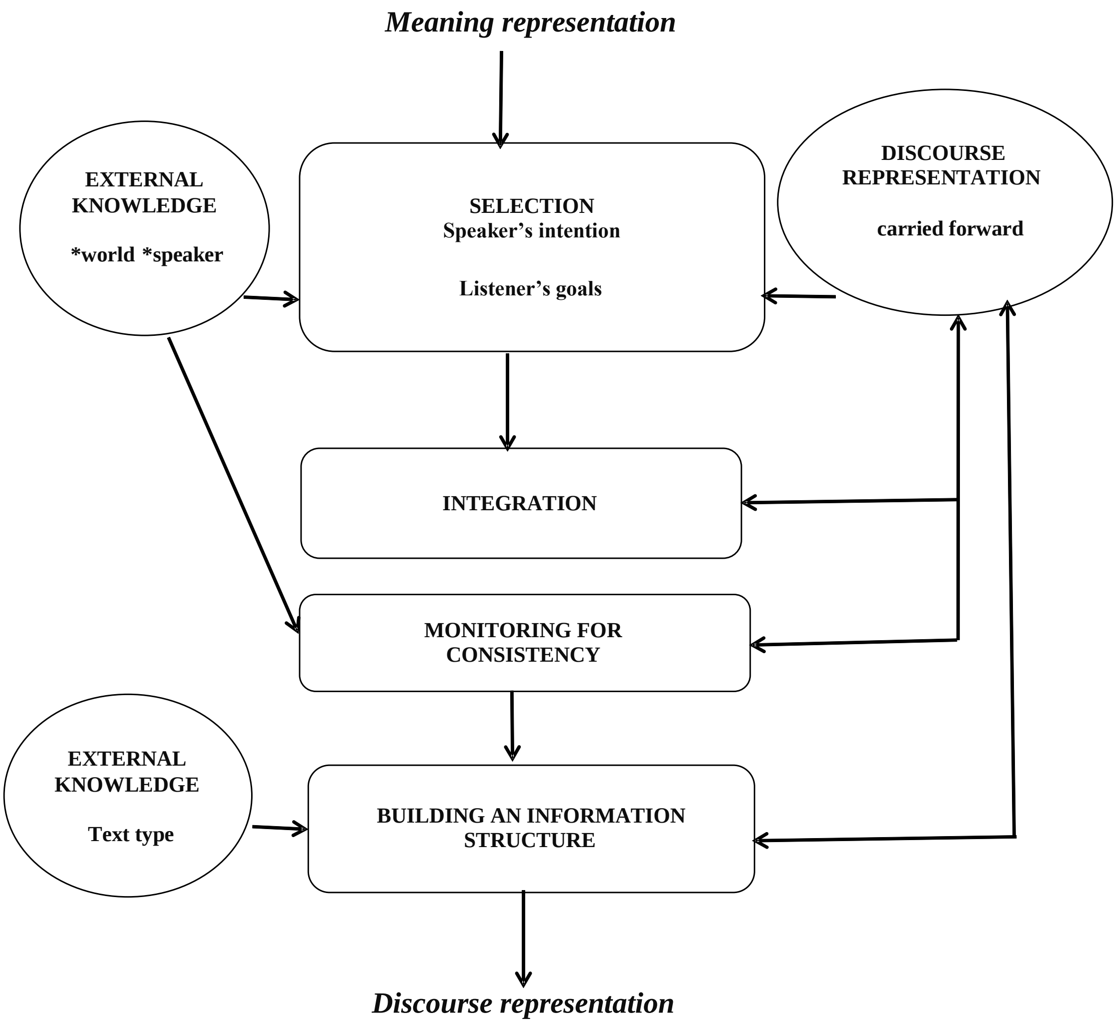


Figure 1.5 Model of discourse construction in listening (Field, 2013, p. 104)

Discourse construction is the second level of higher-level processing in LC. This level also corresponds to the *integrate* level in Cutler and Clifton (1999). According to Field (2013), discourse representation is related to a listener's recall of what has occurred. In the process of discourse construction, external knowledge, which includes world knowledge, knowledge of a speaker, and knowledge of a situation, helps a listener make judgements about the relevance of a new piece of information to the discourse. External knowledge helps the listener decide the relevance of their perception of the speaker's intentions and the listener's own goals. Discourse representation helps the listener make judgements of the new piece of information and integrate the new meaning representations into the developing discourse representation. External knowledge of world and the speaker, together with the discourse representation, help the listener monitor whether a new piece of information is consistent with what has already occurred. Since more information has been obtained for the discourse development, a listener has to make use of external knowledge of text types and discourse representation to decide the relative importance of each item. Based on this, an information structure is built and the discourse construction is produced.

As mentioned, Field (2013) built his model of lower-level processes in LC by drawing upon Cutler and Clifton (1999) and Field (2008). Those earlier models focus on the perception of speech input, or lower-level processing, with little representation of meaning construction or higher-level processing. Subsequently, Field (2013) provides three separate models of LC, one for the lower-level processing and two for the higher-level processing. The model of the lower-level processes reflects the tentative processes of perception and parsing of speech input, while the models of the higher-level processes point out the importance of meaning construction and discourse construction in L2 LC.

As explained in this section, language is processed simultaneously at different levels. Flowerdew and Miller (2005) argue that phonological, syntactic, semantic, and pragmatic information interact with higher-order processing. They hold that an important advantage of the interactive model of LC is that it allows for the possibility of individual variation in language processing. This is important because some individuals put more emphasis on lower-order processing, while others rely more on higher-order processing. Successful LC is assumed to be the result of a complex interaction between higher-order and lower-order cues (Staehr, 2009) and it is likely that the interaction is both compensatory and confirmatory (Graham & Macaro, 2008). According to Field (2008), compensatory processing means that a listener uses their outside information, including world knowledge, topic knowledge, and knowledge of the speaker, to fill in gaps in linguistic knowledge where decoding has been unsuccessful. For example, if a listener feels they have decoded a large proportion of aural input and is confident about the accuracy of their decoding, they will depend less on external information provided by context. On the other hand, if the listener does not feel confident about the accuracy of the decoding, they will rely more on information provided by context. Confirmatory processing means that when LC is relatively problem-free, a listener might use higher-order and lower-order cues to check whether the message has been understood.

In the interaction of L2 learners' lower-order and higher-order processing in LC, the accuracy of word recognition is important, as is the automaticity of word recognition skills to L2 learners (Segalowitz, 2010). Automatic processes require little or no attentional effort from learners and take relatively little time (Segalowitz, 1991). Buck (2001) points out the importance of L2 learners' automatic processing of aural input when he claims that speakers generally produce three words per second and that this leaves little time for L2 listeners to find precise meanings for each word, or even to recognise the structure of clauses. L2 learners' listening processes must be almost entirely automatic if listeners are to follow speakers who produce speech at a normal rate. Therefore, the more automatic a listener's processing is, the more efficient and faster the processing will be (Buck, 2001), and vice versa.

An advantage of Cutler and Clifton (1999) is that the model clearly describes the processes of how L1 learners comprehend speech input. The model is also based on detailed research into language processing (particularly phonetic/phonological processing) where the designers tested different aspects of it. One of the disadvantages of the model is that it was designed for L1 listeners. The original model is about L1 listening, although Cutler (2012) has added considerable information about L2 listening, and highlights the similarities and differences between the processes for both groups of learners. She suggests, for example, that L2 learners rely on procedures from the L1 in segmenting continuous speech and these are not helpful for L2 processing, particularly if there is a great discrepancy between, for example, the rhythm and phoneme sequence probabilities between the two languages. Because L2 learners rely on the L1 and generally have lower levels of competence in the language (e.g. a smaller vocabulary), have had less exposure to the language, and are less familiar with the contexts within which they receive input, L2 listeners are likely to be less efficient at L2 processing than native speakers.

One of the advantages of Field's (2013) model is that it incorporates Cutler and Clifton (1999) and Field (2008). Therefore, there are some overlaps between Field (2013) and the other two models, although Field (2013) is more applicable to L2 listeners. Field highlights that for L2 listeners the processing of language is tentative and listeners have to update and revise their provisional hypotheses at the levels of word, phrase and clause during the process of receiving speech input. This is less clear in other models of L2 listening. Another advantage of Field (2013) is that the model of lower-level processing distinguishes between language processing at three different levels, and these are connected to each other in both directions. Although a disadvantage of the model is that Field has not provided empirical evidence to support the model, the current study could help provide initial evidence for aspects of it.

In summary, this chapter introduced the concept of LC, i.e., LC is a result of interaction between linguistic knowledge and general world knowledge. It showed how L1 listeners decode aural input with the help of linguistic knowledge in four stages under the cognitive model of processing in LC (Cutler & Clifton, 1999). Since both listening comprehension and reading comprehension are receptive processes, the similarities and differences between the two types of comprehension were presented. The chapter explained why L1 LC and L2 LC are different: L2 learners rarely develop the same high level of ability as in their L1; and, in most cases, L2 listeners have gaps in L2 language knowledge (Buck, 2001) which might lead to lower automaticity in processing speech input. Since L2 learners have much poorer phonological, lexical, and syntactical knowledge than native speakers, the current study adopts the model of lower-level processes in L2 LC put forward by Field (2013). Although the study focuses on exploring how L2 learners' linguistic knowledge impacts their LC, the models of higher-level processes in LC put forward by Field (2013) were also presented.

Chapter 2 Variables affecting listening comprehension in L2 learners

To date, few studies have been conducted on the contribution of individual differences to explaining variance in L2 LC. In order to help identify gaps in our understanding of L2 LC, the literature on L2 learner variables is reviewed. Although L2 learners with different L1s may have some common listening problems, Chinese learners of English may have different listening problems due to the fact that Chinese and English have typological differences. Therefore, in this chapter, literature on L2 learner variables is reviewed in the first section. These variables include: learners' phonological knowledge, vocabulary knowledge and grammar knowledge; aural sentence processing speed; cognitive ability, comprising working memory capacity and reasoning ability; and frequency of the use of English in daily life. Issues for Chinese learners of English in LC are summarised in Section 2.2, and the final section introduces the current study and presents the research questions.

2.1 Learner variables

The available literature shows that researchers have conducted many studies on factors which impact L2 learners' LC, yet a limited number of studies have been conducted on learner variables which contribute to success in L2 LC (e.g., Andringa, Olsthoorn, Van Beuningen, Schoonen, & Hulstijn, 2012; Bonk, 2000; Goh, 2000; Graham, 2006, 2011; Graham, Santos, & Vanderplank, 2008, 2010). A better understanding of variables which affect LC in L2 learners is urgently needed because listening is a source of frustration and an area in which L2 learners find it hard to make progress (Graham, 2011). Outside the classroom, listening is very important to those who use a second language in their daily work or studies and who need to understand input in a language with which they have varying levels of competence.

Rubin's (1994) seminal paper distinguishes five major factors that influence L2 learners' LC, namely characteristics of the text, the interlocutor, the task, the listener, and the process. Although it is not possible to explore all these factors in the present study, it is important to try and explore the contribution of a number of important factors which explain the success in L2 learners' LC. The factors which have been investigated in previous studies include: linguistic knowledge and IQ (Andringa et al., 2012); vocabulary knowledge (Bonk, 2000; Matthews & Cheng, 2015; Mecartty, 2000; Oh, 2016; Staehr, 2008, 2009; Wang & Treffers-Daller, 2017); grammar knowledge (Mecartty, 2000); processing speed (Andringa et al., 2012; Oh, 2016); L1 listening ability and L2 proficiency (Vandergrift, 2006; Vandergrift & Baker, 2015); metacognitive awareness (Vandergrift, 2005; Vandergrift, Goh, Mareschal, & Tafaghodtari, 2006); strategies (Graham et al., 2008); working memory (Brunfaut & Révész, 2015; Kormos & Safar, 2008); speakers' accent (Ockey, Papageorgiou, & French, 2016); text characteristics and task difficulty (Révész & Brunfaut, 2013); and prior/background knowledge (see Macaro, Vanderplank, & Graham, 2005, for a review). These studies involve L2 learners with different L1s.

2.1.1 Word recognition in connected speech

As can be seen in Figure 1.3, L2 listeners have to apply phonological knowledge in order to recognise syllables and segment words from connected speech. Rost (1990) states that those listeners who expect to hear ideal pronunciations of words will have considerable difficulty decoding connected speech because phonemes change their perceptual features in different phonemic environments. This is supported by Buck (2001) who holds that, in normal speech, speakers modify sounds considerably and not all the phonemes are clearly and unambiguously encoded in spoken texts. Field (2008) mentions the fact that a listener does not work with raw material which comes in a neat standard format. Listening material comprises acoustic cues which listeners have to match to phonemes, syllables, and words. Therefore, if the listener has inadequate listening skills, they will fail to discriminate the phonemes accurately, or fail to gain the correct meaning of an utterance.

To recognise words from speech refers to the ability to process language input and transform speech signals into lexical units (Matthews & Cheng, 2015). Recognising spoken words in fluent speech is the basis of spoken-language comprehension (Cutler, 1997). Rost (2016) argues that there are two main tasks for listeners in word recognition: one is to identify words and lexical phrases and the other is to activate knowledge which is associated with those words and phrases. Since speech is processed primarily in a sequential fashion, word by word, it is necessary to locate the onset of the immediately following word (Rost, 2016). The model of lower-level processing in L2 LC (see Figure 1.3) shows that a listener's decoding processes reshape a piece of speech into ever-larger units of language. As can be seen in the model, when a listener hears a series of acoustic sensations, they have to match the sensations to sounds/phonemes of the target language. The phonemes are grouped into syllables and the syllables into words. Since words often differ greatly in connected speech from the forms they take when they are said in isolation, word recognition causes serious problems for second language listeners (Field, 2008). Examples seen in Section 1.2 highlight this issue, i.e., English speakers often adjust a sound to the following one in *light grey* to produce *like grey,* or in *white board* to produce *wipe board* (Field, 2008). The recognition and extraction of lexical forms, e.g., segmenting word boundaries to determine the end of a preceding word and the start of the next word in a speech stream, is one of the many challenges faced by L2 learners (Shoemaker, 2014).

Research on how L2 listeners employ phonological cues to segment the speech stream is limited. Based on Altenberg's (2005) study of L2 learners' use of acoustic-phonetic cues to segment the speech stream into words, Ito and Strange (2009) and Shoemaker (2014) carried out a partial replication cross-language study of L2 learners of English. The acoustic-phonetic cues investigated included the presence or absence of aspiration, glottal stop, and aspiration-glottal stop. For example, when a listener hears the sound stimuli /tʃiːfskuːl/, they have to decide which of two possible spellings, *chief's cool* or *chief school*, is right depending on whether or not aspiration of the voiceless stop is present in the sound. Ito and Strange (2009) and Shoemaker (2014) aimed to investigate whether L2 learners could use cues such as English aspiration and glottal stop for segmentation in the same way as native English listeners. Ito and Strange (2009) studied the perceptions of acoustic-phonetic cues to English word boundaries among 30 Japanese learners of English. Results indicated that Japanese learners' responses to the stimuli (83.8%) were significantly less accurate than those of English native speakers (96.8%). L2 learners' performance on aspiration pairs, such as in *lou stops/loose tops*, was significantly less accurate (at 73%) than their performance on glottal stop pairs, as in *a niche/an itch,* (at 91%). This was also true for their performance on double cue pairs (the absence of aspiration (*Asp−*) and the presence of a glottal stop (*Glot+*), as in *grape in,* or the absence of a glottal stop (*Glot−*) and the presence of aspiration (*Asp+*), as in *grey pin* (at 94% accuracy)). The authors found that L2 learners' performance on word segmentation in English word boundaries correlated with their length of residence in English speaking countries (r = .63) and aspiration cues took more immersion experience to learn than glottal stop cues. The findings were similar to those of Altenberg (2005), who investigated Spanish learners of English who were studying at universities in America. In her study Altenberg found L2 learners' percentage of correct responses on items with double cue pairs (92.5%) and those with glottal stop cues (88.4%) were much higher than their accuracy on items with aspiration cues (58.5%). However, there were also differences between the results of the two studies: the Spanish learners of English showed relatively higher accuracy to both positive aspiration stimuli and positive glottal stop stimuli. Shoemaker (2014) carried out a similar study among fifty French-speaking students of English at a Parisian university on their acquisition and perceptions of English word segmentation of potentially ambiguous phrases. The same stimuli were used with different recordings. Results showed that French L2 learners' mean accuracy was 74.6%, which was lower than in Ito and Strange's (2009) findings. The results also found a clear effect of language exposure between the two groups of participants; third-year students performed better than first-year students. In addition, participants in Shoemaker's (2014) study were more sensitive to the presence of glottal stops (e.g., *seen either*; *an ice man*) in English word boundaries than to the presence of aspiration stops (e.g., *loose tops*; *lace table*), which supports Ito and Strange's (2009) findings. Results from each of these studies indicate that the test designed by Altenberg (2005) is a valid test instrument for measuring L2 listeners' competence in segmenting English words from acoustic-phonologic cues. The test can be used to identify L2 listeners' perception of these cues to English word boundaries among Spanish L2 learners (Altenberg, 2005, with a mean accuracy 76.3%), Japanese L2 learners (Ito & Strange, 2009, a mean accuracy 83.8%), and French L2 learners of English (Shoemaker, 2014, a mean accuracy 74.6%). Results also indicate that the process of identifying acoustic-phonologic cues in English words by L2 listeners is not as efficient as it is for native speakers who perform near the ceiling in Altenberg's (2005) and Ito and Strange's (2009) studies. However, apart from the studies mentioned here, very few studies have been conducted which investigate the contribution of word segmentation and word recognition to L2 LC. It is not clear to what extent Chinese learners' ability to segment words/recognise words in the stream of English speech can contribute to an explanation for the variance in L2 LC. In addition, recognising words in isolation is one of the difficulties for L2 listeners. That includes failing to recognise a known word due to a phonetic variation, or knowing a word but attributing to that word a wrong sense (Field, 2003). Therefore, further studies among L2 learners of English are needed to fill this gap in the available literature.

Matthews and Cheng (2015) studied the relationship between word recognition from speech and LC among 167 Chinese university learners of English. In their study, a test of word recognition from speech (WRS) was designed to directly measure L2 listeners' ability to recognise words heard. In the test listeners completed a gap-fill task, so as they listened to individual sentences they had to fill in the gap when they heard the missing word. The target words were chosen from the first, second, or third thousand (henceforth 1K, 2K, 3K) most frequent words list. Participants' LC was measured with an IELTS test. Results showed that the WRS of high frequency word levels was strongly and positively correlated with the IELTS listening scores (r = .73, p < .01), and WRS frequency level of the 3K words alone predicted 52% of the variance in LC. Thus, recognition of high frequency words could be an important predictor of L2 learners' LC.

2.1.2 Vocabulary knowledge

Field (2008) and Wu (1998) highlight the importance of lower-order linguistic knowledge in L2 LC. As one of the types of lower-order linguistic knowledge, vocabulary knowledge is an important predictor of L2 LC (Mecartty, 2000; Staehr, 2008, 2009). Studies show that the contribution of vocabulary to explaining the success of L2 listening varies from 13% (Wang & Treffers-Daller, 2017) to 54% (Matthews & Cheng, 2015). The differences may be due to the use of different methods to measure L2 learners' vocabulary knowledge, as some researchers use written tests and others aural tests, or tests might focus on measuring either receptive knowledge or productive knowledge. In addition, different tests are used to measure L2 learners' LC (Wang & Treffers-Daller, 2017). In Liu's (1995) study of Chinese learners of English, participants' vocabulary size for reading comprehension and for LC was tested in the following way: In each test 200 words were randomly chosen from 3820 words which students were required to master in accordance with the College English Syllabus. Participants' vocabulary size was measured with multiple choice tests. Participants read or heard an English word followed by four options of the word's meaning in Chinese. They then had to choose the correct meaning of the word they had read or heard. Participants' vocabulary size could then be measured based on how many English words they could understand in the reading or listening test. Liu's (1995) study showed that both reading vocabulary size and listening vocabulary size correlated with LC, meanwhile the correlation between the former and LC (r = .57, p = .001) was weaker than the correlation between the latter and LC (r = .66, p = .001). Liu (1995) explored the correlation between L2 LC and vocabulary knowledge among Chinese learners of English, but he did not explore how much of the variance in L2 LC could be explained by learners' vocabulary size.

L2 learners' lexical knowledge can be measured through text coverage, i.e., the amount of familiar lexis in a listening text. Nation (2006) points out that it is important to study how much text coverage a learner needs for adequate reading and LC. Nation (2006) defines text coverage as the percentage of running words in a written or spoken text that a reader or a listener knows. According to Nation (2006) and van Zeeland and Schmitt (2013), a lexical coverage target of 95% is needed for adequate LC. Others assume the target is slightly lower, namely Bonk (2000). Bonk (2000) found the need for text coverage at 90+ percent levels for good performance in LC when he investigated the relationship between lexical knowledge and L2 LC among fifty-nine Japanese university students of low-intermediate to advanced English ability. Bonk used four short listening passages of increasing lexical difficulty as stimuli for L1 recall protocols and for dictation tests in which text-lexis familiarity was measured. The results showed that L2 LC correlated moderately, yet significantly, with text coverage (r = .45, p < .05), and good comprehension frequently occurred with text coverage levels at 90+ percent levels. However, the study also found that some participants who appeared to know only 75% or less of the words in a text were able to achieve quite good comprehension of the same text, whereas others who were familiar with up to 100% of the words in the text could not achieve good comprehension. This means that L2 learners' individual differences in vocabulary knowledge are not enough to explain the variation in listening scores and there must be other factors which contribute to the variance in LC (Macaro et al., 2016). The study suggests that vocabulary difficulty explained 23% of the variance in learners' L2 LC. However, van Zeeland and Schmitt (2013) queried the method by which Bonk (2000) measured text-lexis familiarity. Since lexical knowledge in the four listening passages was measured with dictation tests which were heard multiple times, it remains unclear to what extent learners' vocabulary knowledge can be measured by such an assessment method in which phonological knowledge and vocabulary knowledge were tapped integratively.

Although there are methodological issues with Bonk's (2000) study, it highlights the importance of vocabulary knowledge in L2 LC. The same conclusion is drawn by Staehr (2008), who explored the relationship between vocabulary size and the skills of listening, reading, and writing in English as a foreign language. In Staehr's empirical study, 88 Danish learners of English from lower secondary education participated. A paper-and-pencil test with multiple-choice questions was administered to measure participants' LC, and a test of receptive vocabulary was used to measure their vocabulary knowledge. The results showed that learners' receptive vocabulary size moderately correlated with their LC (r = .69, p < .001), which was lower than the correlations between vocabulary size and L2 reading (r = .83, p < .001). Staehr (2008) suggests three reasons that may explain these findings. Firstly, LC involves L2 learners' ability to recognise the aural form of the word, while the written vocabulary size test measures learners' knowledge of the written form of the words. Secondly, the accessibility of L2 learners' vocabulary affects learners' online processing of aural input and decides whether L2 learners can process aural input quickly and automatically. Thirdly, in addition to vocabulary knowledge, L2 learners' LC might be affected by a wide variety of other factors because the vocabulary size scores explained only 39% of the variance in L2 LC.

The mode in which vocabulary knowledge is tested is not the only variable that needs to be taken into account. Vocabulary knowledge encompasses a wide range of components which can be broadly grouped into form, meaning, and use (Nation, 2001). There are no tests that can assess all the different components of word knowledge, so choices need to be made as to the components that are most relevant for a particular study. Some researchers make a distinction between vocabulary size, i.e., the number of words for which language users know some aspects of their meanings (Anderson & Freebody, 1981), and vocabulary depth, i.e., the quality of the knowledge a learner has, or the number of different aspects of form, meaning, and use a learner knows. Liu's (2011) study indicates that both vocabulary size and vocabulary depth were significantly correlated with LC and that the correlation between vocabulary depth and LC (r = .60, p = .05) was stronger than that between vocabulary size and LC (r = .54, p = .05). Staehr's (2009) study differs from that of Liu (2011) who did not explore the contribution of vocabulary size and depth to L2 LC. Staehr (2009) investigated the contribution of vocabulary knowledge in LC among 115 advanced Danish learners of English and explored the depth and breadth of vocabulary knowledge and their respective impact on learners' LC. Learners' vocabulary breadth was measured using the Vocabulary Levels Test; their vocabulary depth was measured using the Word Associates Test (Read, 1993, 1998); and listening was measured using the Cambridge Certificate of Proficiency in English test. The researcher found a correlation of .7 between vocabulary size and L2 LC, and a correlation of .65 between vocabulary depth and L2 LC. Staehr's (2009) study indicated that vocabulary size and vocabulary depth together explain 51% of the variance in L2 LC, but vocabulary size was found to explain much more variance in LC than vocabulary depth. The former accounted for 49% of the 51% of variance and vocabulary depth accounted for 2% unique variance in listening. What Staehr (2009) found is in accordance with Zhang (2011). In the latter, 237 Chinese learners of English were investigated and the results showed that vocabulary size explained 27% of the variance in LC, and vocabulary depth explained 2% of the variance in LC. Although both Staehr (2009) and Zhang (2011) found that vocabulary size explained more variance in LC than vocabulary depth, but it does not mean that vocabulary depth is not important because measuring it is very difficult.

The studies mentioned above explored the role of vocabulary knowledge in L2 LC. However, only a limited number of studies have been conducted to identify the contribution of other variables to LC in native and non-native speakers. Andringa et al. (2012) investigated 121 native speakers and 113 non-native speakers of Dutch aiming to throw light on the determinants of success in LC. In their study, a computer administered receptive vocabulary test in written form was used to assess non-native speakers' vocabulary size. The results showed that linguistic knowledge, which was subdivided into vocabulary knowledge, grammar knowledge, and phonological knowledge, correlated significantly with both native listeners' LC (r = .88) and non-native listeners' LC (r = .96). In addition, the authors found that linguistic knowledge and processing speed together explained 91% of the variance in native speakers' LC and linguistic knowledge, and IQ explained 96% of the variance in non-native speakers' LC.

The correlations between LC and vocabulary knowledge reported in Andringa et al. (2012) (r = .68 for non-native speakers) are comparable to those of Staehr (2009). However, the correlation with vocabulary knowledge is much lower for native speakers (.35). The differences between native speakers and non-native speakers are also clearly visible in the correlations between LC and grammatical processing accuracy (.77 for non-native speakers, but not significant in native speakers), and segmentation accuracy (.64 for non-native speakers and .29 for native speakers). Vocabulary is also the variable with the strongest partial Eta squared value (.39) in the study. Thus, among non-native speakers, vocabulary knowledge is indeed likely to be a better predictor of LC than grammatical accuracy or segmentation skills, which could be expected on the basis of the findings of Mecartty (2000) and Clahsen and Felser (2006). Andringa et al. (2012) showed that the key factors which explain LC in native speakers are linguistic knowledge and processing speed, while non-native speakers rely mainly on linguistic knowledge and reasoning ability, and processing speed does not explain any variance in LC. Instead the results of the correlations between linguistic knowledge and LC in the two groups, and on the standard regression weights from linguistic knowledge onto LC (.95 for non-natives and .79 for natives), indicate that linguistic knowledge is more important to LC in non-natives than that in natives. If L2 linguistic knowledge is insufficient to understand the aural input, then non-linguistic knowledge is called upon to compensate for that gap, and vice versa. Although linguistic knowledge and reasoning ability together explain 96% of the variance in L2 LC, it is interesting to explore what proportion of the variance is explained by specific components of language ability, namely vocabulary knowledge, grammar knowledge, and phonological knowledge. For this reason, Wang and Treffers-Daller (2017) studied the contribution of different L2 learner variables in LC among 151 adult Chinese learners of English. Learner variables included in the study were general language proficiency, vocabulary knowledge, and metacognitive awareness. Participants' LC was measured through a College English Test Band 4 (the CET4) listening section, and vocabulary knowledge was measured through the Vocabulary Size Test (Nation & Beglar, 2007). Participants' language proficiency was measured using the Oxford Quick Placement Test (OQPT) and metacognitive awareness was measured with the Metacognitive Awareness Listening Questionnaire (Vandergrift et al., 2006). The results indicated that learners' vocabulary knowledge significantly correlated with their L2 LC ($r = .44$, $p < .01$), and that learners' vocabulary knowledge explained 13% of the variance in L2 LC, while learners' general language proficiency also explained 13% of the variance, and metacognitive awareness explained 4% of the variance. The study also found that vocabulary knowledge explained unique variance over and above general language proficiency. However, a limitation of Staehr (2008, 2009) and of Wang and Treffers-Daller (2017) is that vocabulary knowledge was measured through written tests, rather than aural tests. Written tests do not directly tap into L2 listeners' ability to apply vocabulary knowledge to their recognition of words in aural input in real time (Matthews & Cheng, 2015). Therefore, research which taps directly into L2 learners' ability to recognise vocabulary in aural form is needed.

Staehr (2008) suggests that if he had used a measure of phonological vocabulary size or aural vocabulary, the correlational patterns would have been different. Support for this view can be obtained from Milton and Hopkins (2006) who adopted an oral version of X_Lex, called A_Lex. Interestingly, Milton and Hopkins (2006) found that their respondents' aural vocabularies were generally relatively small by comparison with their written vocabularies, even though mid strength correlations existed between the two. Milton, Wade and Hopkins (2010) also found that A_Lex correlated more strongly with listening scores than X_Lex although the variance explained by X_Lex was larger, which was unexpected (see Wang & Treffers-Daller, 2017 for further discussion).

Others have explored the relationships between aural vocabulary knowledge and L2 LC. Vandergrift and Baker (2015) investigated the contribution of learner variables to L2 LC in which learners' L2 vocabulary knowledge was measured with Peabody Picture Vocabulary Test (PPVT). In Vandergrift and Baker (2015), participants were 157 learners from French immersion classrooms in a Canadian school. Learners' LC was measured through a multiple-choice listening test and their vocabulary knowledge was measured through aural vocabulary tests in their L1 and L2. The two authors found a substantial correlation of .51 between L2 vocabulary and LC in L2. In their study L2 vocabulary was a stronger predictor of LC than any other variable, although they also found working memory, auditory discrimination, metacognition (as measured with the MALQ, Vandergrift et al., 2006), and L1 vocabulary contributed indirectly to LC via L2 vocabulary knowledge. The latter, therefore, remains the key explanatory variable in the model.

In addition to the contribution of receptive vocabulary knowledge in aural form in L2 LC investigated in Milton et al. (2010) and Vandergrift and Baker (2015), the role of aural productive vocabulary knowledge in L2 LC was explored among adult Chinese learners in Matthews (2018). Learners had to perceive and produce 63 target words from three word frequency levels for an aural vocabulary knowledge test. Level one included words within the 0-2000 frequency range, level two included words within the 2001-3000 frequency range, and level three included words within the 3001-5000 frequency range. Learners' LC was measured with an IELTS listening section. Matthews (2018) found that each of the three word frequency levels of aural vocabulary knowledge contributed significantly to explaining the variance in LC. Among them, level three made the most important contribution, followed by level two, and then level one. The three levels together explained 52.8% of the variance in L2 LC. Matthews (2018) suggests that aural vocabulary knowledge throughout both the high frequency range (0-3000) and the mid-frequency range (3001-5000) is an important predictor of success in L2 LC.

2.1.3 Grammar knowledge

Although vocabulary knowledge correlates significantly with L2 LC, other components of linguistic knowledge also contribute to L2 LC. According to Bachman and Palmer (2010), grammar knowledge is an indispensable component of a learner's language ability. This view is widely shared in both the field of language testing and of applied linguistics. So far, very few studies have been conducted on the impact of grammar knowledge on L2 LC. Clahsen and Felser (2006) assume that non-native speakers are mainly guided by lexical and semantic cues, but not by syntactic cues. This could explain why, in a study of the relative contribution of grammar and vocabulary in explaining the variance in LC, Mecartty (2000) found that grammar knowledge correlated significantly with LC, but it did not contribute significantly to explaining any variance in LC. Mecartty investigated the relationship between L2 vocabulary knowledge and grammar knowledge in reading and LC activities. In her study, participants were 154 students of Spanish at university level. The results showed that L2 vocabulary knowledge and grammar knowledge both significantly correlated with reading, but only vocabulary knowledge emerged as a significant predictor, explaining about 25% of reading ability. It was found that vocabulary knowledge was also significantly related to LC, explaining about 14% of the variance, however, grammar knowledge did not explain any variance in LC. The results could mean that sampling and measuring syntactic knowledge is more complex than measuring vocabulary knowledge. In Mecartty (2000), L2 learners' vocabulary knowledge was measured through written tasks; they selected the meaning of a target word in Spanish with its equivalent meaning (word-association) or its opposite meaning (word-antonym) in English. Their grammar knowledge was measured through written tasks, a sentence completion multiple-choice task and a grammaticality judgement task. The correlation analyses showed that the correlation between lexical knowledge and LC (r = .38, p < .05) was weaker than the correlation between lexical knowledge and reading comprehension (r = .50, p < .01). It was also found that the correlation between grammar knowledge and LC (r = .26, p < .05) was weaker than the correlation between grammar knowledge and reading comprehension (r = .34, p < .05). The strength of the relationship between vocabulary knowledge and LC may have been reduced because of a weakness in the study, i.e., learners' ability to recognise spoken forms of the words is important to their LC, but what was measured in Mecartty's (2000) study was learners' ability to recognise the written forms of the words (Macaro, Graham, & Woore, 2016). In Mecartty (2000), participants' grammar knowledge was measured through written tests. The current study is going to use both an aural test and a written test to measure learners' grammar knowledge in LC.

Since very few studies have been conducted into the relationship between L2 learners' LC and grammar knowledge, such studies are needed. One such study, by Andringa et al. (2012), explored the correlation between grammar knowledge and LC using the measurement of native speakers and L2 learners' grammar knowledge via an aural task; participants were required to decide the distribution and combinatorial properties of a target language. Participants' reaction time and response accuracy to test items were recorded and results showed that L2 learners' grammar processing accuracy significantly correlated with LC ($r = .77$) (the correlation for native speakers was not provided).

2.1.4 Aural sentence processing speed

According to Roberts (2012), most researchers in linguistics and psychology do not tend to systematically explore the degree to which general cognitive variables affect L2 sentence processing. Roberts argues that although researchers try to control for any variability in their experiments, with the purpose of comparing and making generalizations across groups, it is difficult to control for all cognitive variables that might affect L2 processing.

L1 speakers and L2 learners comprehend heard ambiguous sentences with different efficiency. When processing language it is common for L1 adult native speakers to access different knowledge sources and integrate information from different sources efficiently and without difficulty (Gibson & Pearlmutter, 1998). When L1 speakers are confronted with ambiguity in language comprehension, their ambiguity resolution preferences are affected not only by syntax, but also by individual working memory capacity, and by lexical-semantic, prosodic, discourse level, and probabilistic information (see Gibson & Pearlmutter, 1998, for a review). Ambiguity resolution preferences refers to comprehenders' preferences to solve temporary ambiguities which might arise from syntactic or other factors. Examples can be found in Gibson and Pearlmutter (1998), the word *that* in sentences (1) and (3), below, is a demonstrative article which modifies the single word *hotel*, while the word *that* in sentences (2) and (4) is a complementizer which leads an embedded sentence (*cheap hotels were clean and comfortable*):

(1) That cheap hotel was clean and comfortable to our surprise.

(2) That cheap hotels were clean and comfortable surprised us.

(3) The lawyer insisted that cheap hotel was clean and comfortable

(4) The lawyer insisted that cheap hotels were clean and comfortable.

While processing sentences (1) and (2) with sentence-initial contexts, comprehenders prefer to solve the temporary ambiguity in favour of the article interpretation (1). However, when processing sentences (3) and (4) with post-verbal contexts, comprehenders prefer the complementizer interpretation (4). At present, only a limited number of studies have explored how adult L2 learners resolve structural ambiguities in real time. Felser, Roberts, Gross, and Marinis (2003b) and Papadopoulou and Clahsen (2003) found that when adult L2 learners parse ambiguous sentences they rely more on non-structural information, e.g., lexical-semantic information and contextual cues. Akker and Cutler (2003) showed that, compared with native speakers, L2 learners are less efficient in using prosodic cues during on-line processing. Clahsen and Felser (2006) indicated that adult L2 learners use syntactic representations which are less detailed than those of native speakers. They assume, therefore, that L2 learners are mainly guided by lexical-semantic information, but not by syntactic cues. According to Roberts (2012), only when L2 participants are required to perform a metalinguistic task will individual differences influence their on-line processing of a target language. In such a metalinguistic task, participants have to direct their attention to the manipulation at the same time they are asked to comprehend input. The individual differences included insufficient L2 proficiency and/or cognitive processing constraints, e.g., processing efficiency and working memory capacity.

In Andringa et al. (2012), native and non-native participants' processing speeds were measured through a series of language processing tasks and a word-by-word self-paced listening task with the assumption that efficient comprehension was reflected by fast pace. Participants' reaction times in a semantic processing task, a grammatical processing task, a word segmentation task, and a word monitoring task were recorded. In addition, a word-by-word self-paced listening task was designed to measure native and non-native speakers' sentence processing efficiency. During the task, participants listened to sentences one word at a time, at their own pace, by pressing the space bar (Andringa et al., 2012). In order to ascertain whether participants comprehended the items while doing the pacing, a simple yes-no question followed one sentence in each three. The items varied in lexical frequency (e.g., containing only high frequency words vs. containing at least three low frequency words), syntactic complexity (simple vs. complex), and length (short vs. long). The results suggest that, for native and non-native speakers, the speed of processing linguistic information is a separate construct (Andringa et al., 2012). For native speakers, processing speed was found to be correlated significantly negatively with their LC (r = -.64) and this explained a significant share of the variance in the success of LC. However, processing speed was not correlated with the knowledge factor. This is different for non-native speakers whose processing speed was negatively correlated with LC (r = -.67), but processing speed did not explain any variance in the success of LC. In addition, a strong correlation between non-native speakers' processing speed and knowledge factor was observed (r = -.68), which means that those who can process efficiently also have a high proficiency in the target language.

An explanation for the fact that processing speed was not a predictor in L2 LC in Andringa et al. (2012) is that linguistic knowledge or cognitive factors, or a combination of both, mediated the role of processing speed. More studies are needed on the role of processing speed in explaining the variance in L2 LC. This issue was taken up in Oh (2016) who investigated the relative contribution of linguistic knowledge and processing speed to L2 listening and reading comprehension. Participants were 75 Korean university students whose sentence processing speed was measured using two types of tasks, visual and auditory. An additional variable was sentence complexity with 40 simple sentences and 60 complex sentences. Participants' grammar knowledge (local grammar knowledge and syntactic knowledge) was measured using two types of tests, error correction and sentence completion. Vocabulary knowledge was measured using three types of vocabulary tests, an auditory vocabulary test, a receptive vocabulary test, and a productive test. L2 LC was measured using multiple-choice questions. Results indicated that written sentence processing speed correlated with LC (r = .45, p < .01, for simple sentences; r = .43, p < .01, for complex sentences). However, the correlation between aural sentence processing speed and LC showed a different picture (r = .48, p < .01, for simple sentences; r = .17 p > .05, for complex sentences). A hierarchical multiple regression revealed that vocabulary knowledge explained 39.1% of unique variance in LC, but grammar knowledge and sentence processing speed did not significantly explain unique variance in LC. Together, auditory vocabulary knowledge, grammar knowledge, and sentence processing speed explained 44.5% of the variance in L2 LC.

Another explanation for the fact that processing speed was not found to significantly explain unique variance in L2 LC in the above studies might be sought in the way learners' processing speed was measured. In Andringa et al. (2012) learners' aural sentence processing speed was measured using a word-by-word self-paced listening task. This method has its disadvantages. Certainly during real LC activities listeners cannot control the speed of input. Although the controlled slow-speed task can, to some extent, reflect sentence processing speed and accuracy while listening, it does not reflect listeners' performance under real listening conditions (personal communication, Andringa). In Oh (2016), the total number of accurate responses to 40 simple sentences and 60 complex sentences in a reading test and a listening test was taken to represent L2 learners' sentence processing speed. The two tasks were manipulated and participants were given time to decide whether the meaning of each sentence that was shown visually or aurally was semantically correct. In addition, the sentences in the reading processing speed task and the listening processing task were the same texts in different order and there was an interval of seven days between the reading task and the listening task. Therefore, Oh (2016) only recorded the accuracy of the two tasks which were completed under time constraints, but participants' sentence processing speeds were not recorded.

2.1.5 Working memory

Working memory matters in L2 LC because if a basic operation requires an effort of attention it puts demands on a listener's working memory and other operations can be interfered with as a result of the demand. By contrast, if a basic operation is highly automatic, it will leave working memory free to deal with higher-level processing (Field, 2013). According to Baddeley (2003), working memory plays a crucial role in language learning because working memory, as a temporary storage system, supports thinking capacity. Working memory has been regarded as a possible source of individual differences in comprehension ability (Kormos & Safar, 2008). While there are different views about working memory, Baddeley's (2007) model (see Figure 2.1) is one of the most widely used.

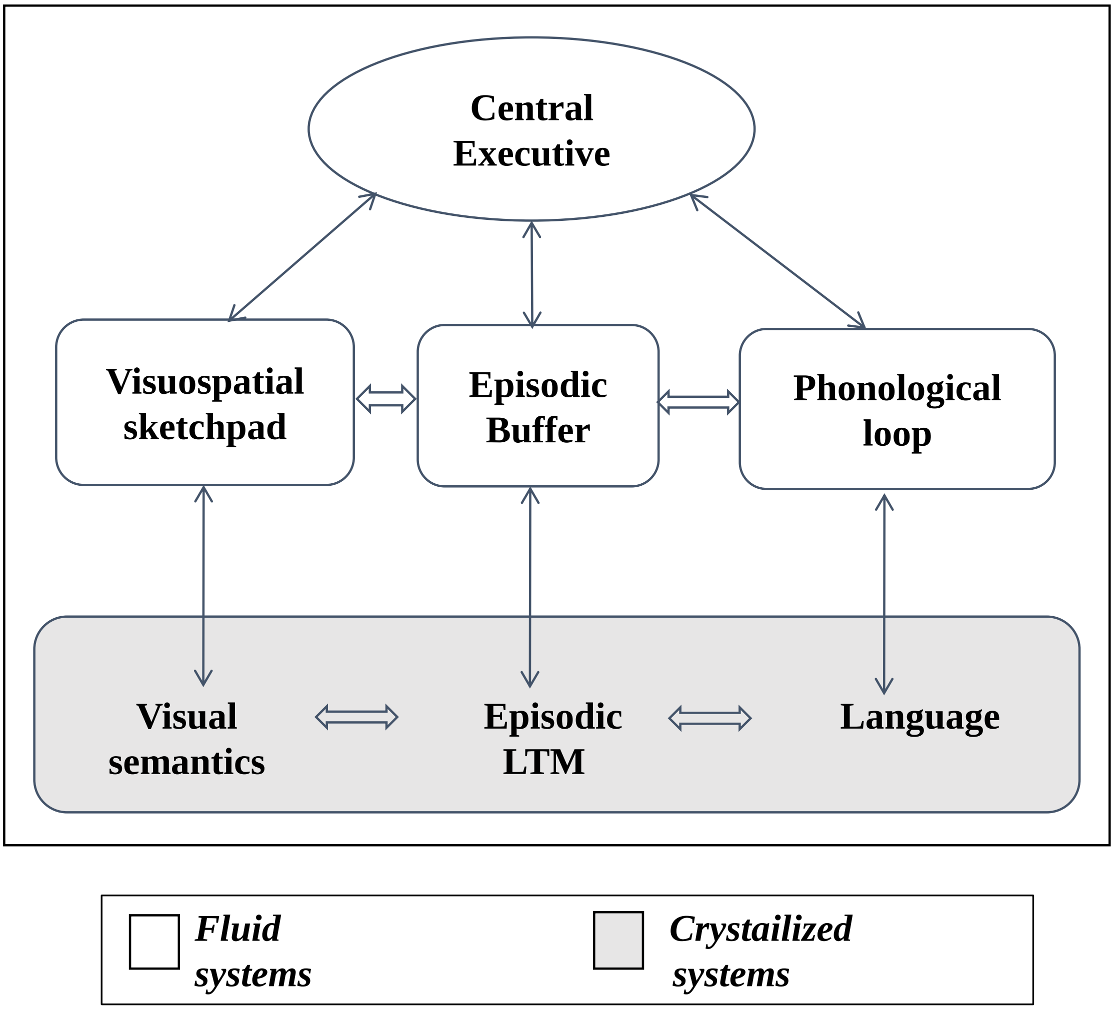


Figure 2.1 The multicomponent model of working memory (Baddeley, 2007, p. 147)

In this model, the conceptualization of working memory does not only include information storage, but also information processing and manipulation which are believed to play an important role in cognitive activities such as comprehension, reasoning, and learning. In the working memory model, four components are distinguished: a central executive; a phonological loop; a visual-spatial sketchpad; and an episodic buffer. Baddeley (2007) claims that the phonological loop consists of a phonological store and an articulatory rehearsal mechanism and that the loop is used to temporarily store speech-based and, possibly, purely acoustic information. The storage depends on a memory trace which is refreshed by rehearsal. According to Kormos and Safar (2008), the rehearsal process can be compared to subvocal speech and this process takes place in real-time, which leads to a limited span. Therefore, if the first item is not rehearsed it will fade in seconds after more items come into the loop. The rehearsal is dependent on either overt or covert vocalization. The phonological loop is the most widely researched component of working memory (Kormos & Safar, 2008). The second subsidiary storage system is the visuospatial sketchpad which performs a similar function for both visual and spatial information. The episodic buffer, the third storage system in the model, is directly linked to the other two subsystems and the central executive. Its role is to integrate information from different sources, such as the phonological loop, the visuo-spatial sketchpad, and long-term memory, into coherent episodes. It is a buffer because it includes a limited capacity storage system in which coded information from different sources can interact (Baddeley, 2007). The function of the central executive is to coordinate the three subsystems: the phonological system, the visual system, and the episodic buffer. According to Vandergrift and Baker (2015), the central executive plays an important role in two aspects: one is to control information flow between the four components of the working memory system and other cognitive processes; the other is to maintain focus and prevent other distracting information or behaviours from affecting the success of LC.

In LC, two components are of particular interest, the phonological loop and the central executive (Brunfaut & Révész, 2015; Vandergrift & Baker, 2015), both of which are limited in capacity. Baddeley (2003) proposed that working memory affects language processing as it is a temporary storage system which underpins thinking capacity. Findings in Kormos and Safar (2008) and Andringa et al. (2012) suggest a significant correlation between working memory and L2 LC. Kormos and Safar (2008) investigated the relationship between phonological short-term memory, working memory, and foreign language performance among 121 Hungarian secondary school learners of English. Participants' phonological short-term memory was measured using a non-word repetition task. Their working memory was measured using a backword digit span test in English, and English proficiency, including reading, writing, speaking, and listening, was measured using the Cambridge First Certificate Exam. The results indicated that backward digit span significantly correlated with listeners' LC (r = .37, p < .05). In Andringa et al. (2012), participants' verbal working memory capacity was measured using four digit span tasks (forward and backward tasks, in visual and auditory forms) and one non-word recognition task. For the auditory digit span tasks, participants listened to Dutch digits. The results indicated that working memory correlated with L2 LC (r = .32, p < .05), but the variable did not explain any variance in L2 LC. Similar correlations between working memory capacity and L2 LC were found in other studies. In Vandergrift and Baker (2015), participants' working memory capacity was measured using a backward digit recall task and a nonword recall task. The results indicated that working memory was related to L2 LC, but a significant correlation was found in only one out of three cohorts (r = .37, p < .05). For the combined cohorts, working memory was not found to be significantly correlated with LC. The findings differ from those of Brunfaut and Révész (2015). In their study, 93 adult non-English native students' listening proficiency was measured through a PTE Academic listening task. A forward digit span task and a backward digit span task, both in visual form, were used to measure participants' phonological short-term memory and working memory capacity. The results showed that both the forward digit span scores (r = .30, p < .05) and the backward digit span scores (r = .31, p < .05) significantly correlated with LC.

2.1.6 Reasoning ability

Reasoning is a process of drawing conclusions from evidence and principles (Sternberg, 2009). Reasoning ability is one of the learner variables in studies on individual differences in L2 acquisition beyond other variables (learners' L1, knowledge of L2, L2 proficiency, L2 experience, and age at onset of L2 acquisition) (Colantoni, Steele, & Escudero, 2015). Some studies have been conducted on reasoning ability in L2 reading comprehension or L1 LC, but few have been conducted on the impact of reasoning abilities on L2 learners' LC. Netten, Droop, and Verhoeven (2011), working with Grade 6 children, explored the predicting power of learner variables in reading literacy among 729 L1 learners and 93 L2 learners of Dutch. The variables included nonverbal reasoning skills, decoding, language, mathematics, reading motivation, academic self-confidence, reading literacy in Grade 4, and home reading resources. In the study, learners' nonverbal reasoning was measured with two subsets of a nonverbal intelligence test (ITS, 1994), i.e., composing figures and exclusion which required children to identify the missing part of a figure, or a deviant figure, out of four alternatives. Findings from the study suggested that nonverbal reasoning was a predictor of learners' reading literacy in both L1 and L2 groups. Zampini, Suttora, D'Odorico and Zanchi (2013) investigated the contribution of sequential reasoning to an explanation of the variance in LC ability among 110 preschool children, aged three to six years old, from monolingual Italian-speaking families. In their study, children's sequential reasoning ability was measured with a task specifically designed for the study. The task assessed children's ability to comprehend events and process complex semantic relationships between events. Children were required to arrange three to five picture cards in order to create a short story. The children's LC was measured using two forms of a LC test comprising two stories and ten multiple-choice questions with four alternatives each. The study found that children's sequential reasoning ability explained uniquely 23% of the variance in LC textual scores and 17% of the variance in LC inferential scores in the three-year age group. This factor was found to be one of only two variables which contributed significantly to explaining the variance in L2 LC in Andringa et al (2012). In their study, participants' reasoning ability was measured using a nonverbal IQ test, the Wechsler Adult Intelligence Scale (Wechsler, 1997). Results showed that participants' reasoning ability correlated with L2 LC (r = .51). Together with linguistic knowledge, the two variables explained 96% of the variance in LC. The factor of intelligence and psychological state was found to be the most influential factor in L2 learners' LC in Li (2012). The researcher investigated L2 learners' opinions on the impact of four factors on LC using a questionnaire among adult Chinese learners of English. Results from factor analyses indicated that the factor of intelligence and psychological state is the most important factor to impact learners' LC, followed by the other three factors in order, i.e., cultural background knowledge, listening strategies, and linguistic knowledge. Results from Li (2012) suggest that language learning aptitude plays an important role in affecting L2 learners' LC.

2.1.7 Personal background factors

In the present study, since L2 learners are investigated, personal background factors include: the age at which participants started learning their L1 and the L2; the frequency of their use of both L1 and L2 in their studies and daily activities; and their self-rated proficiency in both languages. Educators and administrators hold the view that language learners usually benefit from their study abroad experience because they have some level of immersion among native speakers during this time and certainly such immersion is not available at home where they use their L1 (Cubillos, Chieffo, & Fan, 2008). Cubillos et al. (2008) explored the impact of a short-term study abroad course on LC of L2 learners. They compared the study-abroad groups with learners on a similar course at a home campus. In general, the study-abroad groups did not show significantly higher gains in L2 LC than their peers studying on the home campus. However, when two subgroups with a higher level of listening competence were compared after the treatment, the study-abroad subgroup had made more significant gains in LC than their peers on the home campus. According to self-reports, study-abroad groups were more confident in communicating in the L2 than their peers on campus, however, for most students such increased confidence did not manifest itself in actual gains in L2 LC.

The researchers attributed this to the non-interactive nature of the assessment (Cubillos et al., 2008). Although Lapkin, Hart, and Swain (1995) also found that L2 learners from a three-month study-abroad experience did not show significant gains in L2 LC as measured through pre- and post-tests, other studies have reported different findings.

Other studies were conducted on the influence of L2 environment immersion on LC. Allen and Herron (2003) investigated 25 American undergraduates on a French summer study abroad programme in Paris which lasted for 40 days. During the programme, the participants enrolled on courses, received instruction in an L2, and were immersed in an L2 environment through activities, e.g., visits to museums, monuments, and parks. In addition, some participants lived with local French families. Allen and Herron (2003) explored participants' gains in French listening tasks using interviews and questionnaires and compared participants' scores between pre-test and post-test listening tasks. They found that participants' post-test scores were significantly higher than those gained in the pre-test, which means that the participants' listening proficiency significantly improved after the study abroad programme. They also found that over 75% of the participants self-reported their gains in one or more listening tasks. Similarly, Llanes and Botana (2015) investigated a group of North American university students who learned Spanish as a second language and enrolled in a six-week study abroad programme in Costa Rica. In this programme the participants received instruction in an L2 in courses on culture, writing, and language. They were also required to complete weekly course assignments for which they had to visit nearby facilities in order to communicate with local, monolingual, Spanish speakers. Llanes and Botana (2015) found that the participants' listening scores in the post-test were significantly improved from those gained than in the pre-test, which means that the six weeks of study in an L2 environment had a positive impact on the L2 learners' listening proficiency.

In conclusion, the available literature shows that a wide range of individual differences are correlated with L2 LC, but a limited number of studies have explored which of these are most important in explaining the variance in L2 LC, or provided further details about the unique contribution of each. Although Andringa et al. (2012) found that their model explained 96% of the variance in L2 LC, it is not clear whether the same model can account for LC among learners with other L1s, e.g., Chinese learners of English. In addition, adult Chinese learners of English are confronted with many LC problems which are discussed in the following section. Certainly, a study which focuses on the specific issues facing Chinese learners is urgently needed.

2.2 Issues of Chinese learners of English in listening comprehension

It is particularly important to study LC in the Chinese context because Chinese learners often find it hard to understand native speakers of English (Goh, 2000; Y. Wang, 2008). The problems are likely related to typological differences between Chinese and English, and also to intercultural differences (Jia, 2003; F. Wang, 2008). Presently, only a limited number of empirical studies have been conducted on the issues of Chinese learners of English in L2 LC.

Y. Wang (2008) explored Chinese learners' problems in English LC, and reasons which may have caused the problems, using questionnaires and interviews with university students. In the questionnaires, participants were asked to answer two questions:

1) Do you have any problems/difficulties in English LC?

2) If you have any problems/difficulties in English LC, please list them.

Based on Anderson's (1995) theory of three-stages of language comprehension, i.e., perception, parsing, and utilization, respondents' answers to the second question were catalogued. The study found that learners' English LC problems mainly occurred at the perception stage with few problems in the utilization stage reported. In the perception stage, learners faced problems related to sound recognition, limited vocabulary knowledge, and retention of attention. Further, three difficulties were very common for problems of sound recognition. Firstly, when input speech was fast, learners found it hard to segment the continuous speech stream into words or phrases for comprehension. Secondly, when learners came across assimilation or elision between two sounds, they thought that the sound came from a word with which they were not familiar, which led to incorrect comprehension. Assimilation, as seen in Section 1.2, means a speaker adjusts a sound to the one which follows (Field, 2008). For example, the sound /*hot bath*/ becomes /*hop bath*/, and the sound /*good play*/ becomes /*goob play*/. Elision means a speaker omits a sound (Field, 2008). For example, the sound */t/* in */'last'jɪə/*, i.e., *last year*, is omitted and listeners hear the phrase as */'las'jɪə/* (Brown, 1990). Other examples of elision can be found in Section 1.2. Thirdly, when learners heard unfamiliar intonations their comprehension was impacted.

For the problem of limited vocabulary, learners had two specific difficulties. One was their smaller vocabulary size, which affected LC; the other one was that they could not find a corresponding meaning for a word heard which affected their online comprehension, but later they found they knew the word in its written form. Learners' specific difficulty in retaining attention was that when they heard unfamiliar words or phrases, they were very soon distracted and could not concentrate on the next speech input. Y. Wang (2008) considered possible reasons which resulted in these problems on the basis of cognitive processes. The problems of sound recognition were caused by failing to segment the speech stream, lacking familiarity with changes of contexts, and having received incorrect sound input which remained in learners' long-term memories. In L2 LC, learners were not familiar with weak and strong syllables of words and even mixed up similar sounds. They did not realize that some sounds were changed in the speech stream as the result of assimilation and/or elision, and it was this that led to problems segmenting the speech stream. In addition, changes of context from different speakers, intonations, and occasions, might lead to changes in the sounds heard, but Chinese learners did not adapt to those changes. Y. Wang (2008) found that if sound input had been wrong and had been kept in a learner's long-term memory, hearing the correct sound did not lead to learners' correct comprehension. Y. Wang also analysed reasons why participants were unable to correspond sounds heard to a word meaning. Activation of word recognition depends on aural or visual input and there is a certain connection between word sound and word meaning. When a learner receives aural input, word sound is recognised and corresponds to a sound presentation of the word in the learner's mind, then word meaning is activated. Y. Wang (2008) estimated that when Chinese learners of English learned words, they stored properties of words coming from vision, sound, and semantics, into their long-term memory. However, sometimes the property of word sound was missing or vague, which led to the problem that word sound was recognised but word meaning was not activated. These learners commonly thought that it was enough to read a word aloud several times and to keep word spelling and word meaning in mind. In fact, they did not realize that to pronounce a word based on a phonetic symbol did not mean that sound was kept in a learner's mind, or even that the word sound was remembered, it did not necessarily mean a correct connection was built between a word sound and word meaning. As for the problem of attention, since attention was limited, it was likely that learners lost attention to other things when they concentrated their attention on one thing. In the processes of LC, Chinese learners of English had to increase their attention on processing unfamiliar words or phrases, and this led to their failure to assign enough, or any, attention to incoming speech input. As a result, some speech input was not processed. From the participants' answers to the questionnaires, Y. Wang (2008) estimated that the main reason for this problem was that these L2 learners' LC was not yet fully automatic. Indeed, the higher the degree of automaticity, the less attention is required to the processes in LC.

Y. Wang (2008) also found that learners' LC problems in the parsing stage were fewer than those in the perception stage. There were two main problems in this stage. Firstly, the Chinese learners missed incoming sentences while they were concentrating on a previous sentence heard. Learners found it hard to comprehend incoming sentences heard while they were still processing a previous sentence, or they missed other sentences as they recalled a certain word or sentence meaning from what had been heard. Secondly, learners' LC of incoming information was affected as a result of problems comprehending information which was heard before the incoming information. Chinese learners had such problems mainly because they did not make full use of their world knowledge, which had been stored in their minds (Y. Wang, 2008). Learners' problems at the utilization stage occurred because learners were not clear what the theme of a paragraph was after they heard it. According to Y. Wang, the main reason for this problem was that learners had no corresponding knowledge or experience of a related topic, which resulted in the phenomenon where learners comprehended the explicit meaning of a paragraph, but did not comprehend its implicit meaning. The listening problems found by Y. Wang were based on the questionnaires and the interviews, rather than on actual online listening processing.

The LC problems mentioned above were also found in Goh's (2000) study of adult Chinese learners of English. Analysis of data collected from learners' diaries, group interviews, and immediate retrospective verbalisations, allowed Goh (2008) to summarise ten most common self-reported problems which occurred during LC. Among the ten problems that learners perceived themselves to be faced with, five occurred in the perception stage, three in the parsing stage, and two in the utilisation stage. In the perception stage these difficulties concerned: not recognising known words; neglecting the following part when they were thinking about meaning of the current part; being unable to chunk streams of speech; missing the beginning of texts heard; and finding it hard to concentrate (Goh, 2008). The difficulties in the parsing stage included: quickly forgetting the information heard; being unable to form a mental representation; and being unable to understand subsequent parts of input because of previous problems. The problems found in the utilisation stage included: understanding the words, but not understanding the intended message; and not understanding the key ideas in the message. Goh (2008) concluded with a number of reasons for these learners' LC problems. Problems at the perception stage were likely because sound-to-script and word frequent relationships were not fully automatized and learners had limited-capacity short-term memory. Learners' failure to chunk the speech stream into recognised words or phrases also proved difficult. The problems which occurred during parsing were mainly because learners failed to form a mental representation from familiar words. Finally, problems at the utilisation stage were mainly caused by learners' lack of sufficient background knowledge and communicative competence. Similar to Y. Wang (2008), the learners' listening problems identified by Goh (2000) were based on learners' self-reports and interviews, not on actual processing.

The problems mentioned above were summarised from investigations among Chinese learners of English. Meanwhile, F. Wang (2008) explored obstacles to learners' LC among 100 adult learners and five Chinese teachers of English using interviews. Results indicated that the most common obstacle that both teachers and learners agreed on was learners' lack of both linguistic and background knowledge. Lack of vocabulary knowledge was found to be the biggest obstacle, followed by lack of background knowledge and grammar knowledge. For learners' obstacles which were caused by English sounds, teachers and learners had different opinions. The interviewed teachers did not regard "sound", "speech speed", "assimilation", "different sounds between American and British English" to be learners' difficulties. However, learners regarded the above-mentioned to be obstacles in their English LC, and the problem of "speech speed" was highlighted. It can be seen that the teachers were not aware of the learners' problems with phonetics. What teachers regarded as learners' obstacles were "heard but did not understand", "intonation", "understood words or phrases in written form but did not understand them in speech input". Teachers attributed these problems to learners not being familiar with English sounds, and not having sufficient phonological knowledge in mind. Differing from the teachers' opinions, learners regarded "understand individual words but not sentences" as their main obstacle which was related to features of English sounds. This obstacle indicated that the learners could recognise individual word sounds, but they had difficulties in LC when the continuous words formed sentences or discourses. According to F. Wang (2008), although the interviewed teachers and the students had some different opinions on the learners' obstacles, e.g., the "intonation" obstacle from the viewpoint of the teachers and the "sound", "speech speed" and "assimilation" obstacles from the viewpoint of the learners, the obstacles still meant that the learners had many LC problems when decoding sounds and needed instruction in basic English phonological knowledge for listening exercises.

Others have also investigated adult Chinese learners' LC problems using questionnaires. Su (2003) found that the problems included: "failing to comprehend word meaning"; "failing to keep in mind the information heard"; "failing to follow speech speed"; and, "being unable to process the next part heard". Some learners mentioned the psychological impact on them during LC, e.g., many felt very nervous when listening to English. Hu (2009) made similar findings when he investigated second-year university students, using introspection and self-reports, which indicated what learners perceived their listening problems to be. Hu (2009) found thirteen LC problems, of which eleven were categorised in Anderson's (1995) three stages, and two were caused by emotional and psychological factors. Sun and Li (2008) compared Chinese learners' LC problems between participants of higher-level and lower-level listening proficiency. They measured learners' listening proficiency using a self-designed test which comprised two longer monologues. They also used a retrospective questionnaire to investigate learners' listening problems and strategies. They found that participants at each level had similar LC problems; the most common occurred at the perception stage, followed by problems at the parsing stage; the fewest problems occurred at the utilisation stage. Sun and Li (2008) also found that the problems in the perception and the parsing stages for the lower-level group were more than for the higher-level group, and the problems in the utilisation stage for the lower-level group were fewer than those for the higher-level group.

Xu (2014) pointed out adult Chinese learners' LC problems from the perspective of Chinese teachers of English. The problems were mainly a result of a lack of linguistic knowledge (phonological and vocabulary knowledge), failure to follow speakers' speech speed, inability to adapt to sound differences between British and American English, lack of cultural background, and learners' unstable psychological state. Chen (2005), Li, Zhang, Wang, and Tang (2010), and Li (2013) found similar LC problems among Chinese learners of English. Zhang, Lv and Tan (2010) provided empirical evidence of Chinese learners' listening problems. They investigated learners' perceptions of their listening problems using a questionnaire, and measured learners' listening proficiency using a listening test. They found that Chinese learners' listening problems lay mainly at the parsing stage. Between the different levels of language proficiency, the lower-level learners' listening problems mainly occurred in the perception stage, and the higher-level learners had problems mainly at the parsing and the utilisation stages. They also found that learners' scores on their perceptions of listening problems explained 15.9% of the variance in LC.

From the studies discussed in this section, it can be concluded that the main LC problems for adult Chinese learners are mainly to do with the perception of speech input, including decoding sounds, segmenting syllables and words, and recognising words. Other common problems involve those which result from a lack of linguistic and cultural background knowledge, failure to follow speakers' speech rates, and failure to comprehend sentence meaning and to retain information heard in memory. Chinese learners of English have more LC problems in the perception stage than in the parsing and utilization stages. Learners' emotional and psychological state also impacts their LC and causes LC problems.

Literature reviewed in the Section 2.1 indicated that the variables which impact LC in L2 learners are mainly linguistic knowledge, sentence processing speed, working memory, reasoning ability, and frequency of the use of English. Although those learner variables are assumed to affect the LC of Chinese learners of English, Chinese learners have more decoding problems than other L2 learners, according to the review seen in Section 2.2. One of the reasons for this is that the two languages, Chinese and English, have different sound systems and Chinese learners try to use their L1 characteristic of segmenting speech to segment L2 speech. According to Cutler (2001), this is common among L2 listeners. Chinese sound occurs in single syllables, with distinguishing syllables, e.g., */da/, /li/, /ning/*. Most structures of syllables in Chinese sounds are simpler (CGVX, C is an onset, G is a glide, V is a vowel and X is a coda) and Chinese sounds have no consonant clusters (Lin & Wang, 2018). Polysyllabic sounds are rarely seen in Chinese. Each single syllable in Chinese sounds corresponds to one or more meanings. For example, a single syllable /i/ can be realized with four different tones and each possible realization corresponds to a different meaning, or sometimes to multiple meanings. With tone 1 /i/ means '衣 clothing' '医 doctor' or '一 one', with tone 2 it means '姨 aunt' or '疑 question', with tone 3 it means '椅 chair' and with tone 4 it means '意 meaning' or '异 difference'. Chinese learners segment Chinese sounds by recognizing different sounds of CV structure and taking a consonant as the start of a Chinese word in speech. However, English polysyllabic words are common and can have more than one CV structure in one sound. For example, in a CVCCVC sequence, possible segments are CV + CCVC for *reprise*, and CVC + CVC for *publish*. Chinese learners of English often segment the syllables as CVC+ CVC (Zhang & Lin, 2002). Using their L1 segmentation of speech in English speech leads to LC problems for Chinese learners of English. Another reason is that some sounds in English do not exist in Chinese, which makes Chinese learners struggle when they hear those sounds (Zhang, Li & Liu, 2005). Such sounds include /θ/ /ð/ /ʒ/

/v/ /tr/ /dr/. In English, some vowel sounds are similar to Chinese sounds, including /ɑː/ /ʌ/ /ɜː/ /ɜ/ /æ/ /ɔː/ /ɔ/ /e/ /əʊ/ /ʊ/ /i/ /iː/ /ʊ ː/, thus it is easier for Chinese learners to recognize them. However, some English sounds differ greatly from Chinese sounds and it is this which often causes LC problems when learners hear words which include them. These sounds include /ei/ /ɑi/ /ɑu/ /iɜ / /uɜ/ /eə/ /j/ /z/ /ŋ/.

The third reason is that Chinese learners of English do not realize that stress is an important symbol of segmentation of syllables or words. An English word might correspond to a different meaning when a different syllable is stressed within the word. For example, when the word *record* is produced as / rɪˈkɔːd/, it is used as a verb, and when it is produced as /ˈrekɔːd/, it is used as a noun. In the two sounds / rɪˈkɔːd/ and /ˈrekɔːd/ the first vowel in the first syllable is different; stressed syllables must have full vowels and reduced vowels are unstressed (Cutler & Clifton, 1999). Stress is expressed in the segmental structure of words as much as in the suprasegmental structure (Cutler & Clifton, 1999). The fourth reason is that there are many different grammar structures in each language. Chinese learners might have LC problems when they hear texts using English grammar structures with which they are not familiar, or when they comprehend texts by using the grammar of Chinese. For example, word order differs between the languages (R. Wang, 2014; Zhang, 2012). In the sentence, *Two boys quarrelled in the class yesterday*, 'yesterday' and 'in the class' are placed at beginning of the sentence in Chinese, but they are put at the end of the sentence in English. In Chinese, attributive modifiers (nouns and attributive clauses) are placed before a noun which is modified, but in English attributive modifiers can be placed either before or after the noun which is modified (R. Wang, 2014). Attributive clauses must be put after the noun modified or put at end of the sentence (R. Wang, 2014). For example, in the sentence *Distance learning courses are courses in which the instructor communicates with students using computer technology*, the attributive clause *in which the instructor communicates with students using computer technology* is placed before the second *courses* in *Distance learning courses are courses* in Chinese. But in English the clause is placed after the second *courses* which it modifies. Grammar differences are reflected in tenses as well. Tenses in Chinese are simpler than those in English (R.Wang, 2014). For example, the present perfect and past perfect tenses are used in English (have done and had done), but in Chinese no changes are made to a verb to express the two tenses. What is needed is the addition of the adverbial 'already' before the verb.

In summary, the differences in sound systems, tone systems, and grammars in Chinese and English are very complicated. Those differences result in many decoding problems in L2 LC among Chinese learners of English, which suggests that there is a big challenge for Chinese learners to understand English speech. Based on the variables which were found to affect LC in L2 learners in previous studies, and the LC problems which occur particularly in Chinese learners of English, a study on L2 LC is urgently needed to explore to what extent each of the learner variables impacts LC of L2 learners.

2.3 The current study

The review of the literature, and in particular that concerning LC issues among Chinese learners of English, showed individual differences in LC have not been explored in a large number of studies. Thus, the current study aims to explore the impact of learner variables on LC among Chinese learners of English. Participants include adult Chinese learners in China (N = 147) and in the UK (N = 40). The learner variables to be studied include: linguistic knowledge (phonological, vocabulary, and grammar); aural sentence processing speed; working memory; reasoning ability; and frequency of use of English in daily life. These learner variables are chosen because the literature (see Section 2.1) indicate that these variables play an important role in predicting L2 learners' LC. Theoretically, the current study is mainly based on the lower-level processes in LC developed by Field (2013) (see Figure 1.3). In Field (2013), L2 LC is affected by sources arising from linguistic knowledge, which is subdivided into the three types of knowledge mentioned above. The current study focused on exploring the lower-level processes in LC because Chinese learners of English have many decoding problems in processing aural English, as indicated in Section 2.2.

The study is a partial replication of Andringa et al. (2012) in which the variables of linguistic knowledge, processing speed, working memory, and IQ were explored among 121 native speakers and 113 non-native speakers of Dutch. The current study adds the variable of frequency of English use (EU) to the model they developed. Although frequency of L2 use was not considered in their study, the authors suggest that differences in opportunities to use the target language might explain variability in success in LC. To test this assumption, this study compares Chinese learners of English in China with those in the UK as learners in the UK have more opportunities to speak and listen to English in their daily activities than do learners in China. Andringa et al. (2012) built a LC model with the four variables mentioned above, but in which IQ and working memory were originally grouped into cognitive ability.

In their study, multi-sample structural equation modelling (SEM) was used to analyse data. The results showed that, for native speakers, linguistic knowledge and aural information processing speed together explained 91% of the variance in LC; while for non-native speakers, linguistic knowledge and IQ together explained 96% of the variance in LC. Although their SEM model of LC is exceptionally successful in explaining the variance in L2 LC, it is not clear whether the same model can account for LC among L2 learners with other L1s, such as adult Chinese learners of English. Therefore, the present study aims to investigate whether the model developed in Andringa et al. (2012), with the EU added, can also account for LC among Chinese learners. Data collected are analysed with SEM to test against two models of LC which are adapted from Andringa et al. (2012) for the entire cohort of participants. In the two models, LC was measured using two different listening tasks, and other variables were measured with individual tasks. Since participants in China and in the UK have different backgrounds of L2 use, the current study explores whether L2 LC proficiency between two subgroups (one in China and the other one in the UK) differs. It also investigates reasons that can explain the LC differences between the two groups. Since linguistic knowledge was found to have associations with LC for both native speakers and L2 learners in Andringa et al. (2012), an exploration of the contribution of the variable to L2 LC among Chinese learners of English is worthwhile. As has been seen in Section 2.2, decoding problems are the most common LC problems for Chinese learners, thus it is important to investigate the contribution of each type of linguistic knowledge source to explaining the variance in L2 LC. Therefore, hierarchical regression analyses are undertaken in the present study to explore the contribution of linguistic knowledge, phonological knowledge, vocabulary knowledge, and the grammar knowledge to explaining the variance in LC, as measured with two different listening tasks. Hierarchical regression analyses are also undertaken to explore the contribution of learner variables to explaining the variance in LC as measured with one listening task for each subgroup. Since the groups in the UK comprise a small number of participants, only some of the variables are chosen to build a regression model of LC for each subgroup.

The present project aims to take steps towards filling the gaps in our understanding of L2 learners' LC by providing empirical evidence about the role of individual differences in explaining the variance in L2 LC. Two structural equation models of LC, as measured with different listening tasks, are built to confirm associations between LC and the variables assumed to impact LC of L2 learners. The study investigates to what degree the assumed variables can contribute to explaining the variance in LC in L2 learners. The contribution of each learner variable to explaining the variance in L2 LC is also explored. SEM analyses and hierarchical regression analysis methods are used to analyse data collected. Results on the LC models among L2 learners will throw light on understanding what language sources learners use to process L2, whether frequency of L2 use impacts learners' listening proficiency, and how learners' cognitive ability associates with LC. Since different analysis methods are used, comparison of the results may give insights into which model can explain more variance than other models. Findings from the study on measuring L2 learners' LC proficiency are also discussed. Therefore, the current study contributes to pushing the development of testing in the L2 field. Finally, implications for different stakeholders and L2 learners are formulated.

Based on the literature review of variables which impact LC in L2 learners, and on the specific issues faced by Chinese learners of English in LC, the current study explores the following research questions (RQs):

1. Which individual differences can explain the variance in L2 LC among adult Chinese learners of English?

2. What is the contribution of the following variables in explaining the variance of L2 LC: linguistic knowledge; aural sentence processing speed; working memory; reasoning ability; and frequency of use of an L2?

3. To what extent do Chinese students in China and in the UK differ from each other in L2 LC ability? Which factors can explain these differences?

Chapter 3 Methodology

This chapter presents the methodology used in the current study. Section 3.1 introduces the research design. Sections 3.2 and 3.3 discuss the selection of participants and the instruments used to measure the learner variables. In Section 3.4 the procedure of data collection is presented and then the method of analysis is given in Section 3.5. Finally, ethical issues are considered in Section 3.6.

3.1 Research design

This study is a partial replication of Andringa et al. (2012) who investigated LC among L2 learners of Dutch and native speakers of the language. The current study aims to identify which factors explain the variance in L2 LC among adult Chinese learners of English. The tasks used by Andringa et al. (2012) and those used in the current study are presented in Table 3.1.

Table 3.1 Tasks used by Andringa et al. (2012) and those used in the present study

		Andringa et al. (2012)		**The present study**	
L1		35 languages		Chinese	
L2		Dutch		English	
Constructs		**Number of tasks**	**Tasks**	**Number of tasks**	**Tasks**
LC		1	A national examination	2	The listening section of Cambridge Preliminary English Test (The PET listening) The listening section of nation-wide College English Test Band 4 (The CET4 listening)
Linguistic Knowledge	**Phonological Knowledge**	1	Receptive, aural	3	Receptive & productive, aural
	Vocabulary Knowledge	1	Receptive, written	1	Receptive, aural
	Grammar Knowledge	1	Receptive, aural	2	Receptive, written & aural

Processing Speed	5	Semantic processing speed Grammatical processing speed Segmentation processing speed Word monitoring Sentence processing speed	1	Aural sentence processing speed
Working Memory	5	2x digit span visual tasks, forward & backward 2x digit span auditory tasks, forward & backward 1x non-word recognition task	4	2x digit span tasks in L1, forward & backward 2x digit span tasks in L2, forward & backward
Reasoning Ability	1	A nonverbal IQ test (the Wechsler Adult Intelligence Scale)	1	A nonverbal IQ test (Raven's Standard Progressive Matrices)
EU		n/a	1	A personal background questionnaire

Although the available literature indicates that personal background factors (e.g., L2 learners' study abroad experience and immersion in an authentic L2 environment) contribute to L2 learners' aural input comprehension, it is not clear which proportion of the variance can be explained by the personal background factors. Therefore, a personal background questionnaire (PBQ) was used in the present study to investigate 187 adult Chinese learners of English length of L2 study, dominance of L1 and L2 in their daily life and study, and self-rating of their L1 and L2 proficiency in the four skills of reading, writing, listening, and speaking. The participants were divided into two subgroups: one group of 147 university students in China and one group of 40 university students in the UK. The dependent variable (L2 LC) was measured using a listening section of the national College English Test Band 4 (CET4) in China and the Cambridge Preliminary English Test (PET) listening section. The CET4 is a nation-wide college English test used to measure adult Chinese learners' English proficiency in four skills: reading, listening, writing, and translating (see Section 3.3 for details). The PET is an international language test.

Structural equation modelling (SEM) analysis was applied to answer the first research question. SEM is a statistical methodology which uses a confirmatory approach to analyse phenomenon based on a certain structural theory (Byrne, 2016). SEM includes two important aspects: firstly, there are a series of structural (e.g., regression) equations which represent the causal processes under study; secondly, a model is built to picture structural relations in order to provide a clearer conceptualization of a theory under study (Byrne, 2016). A hypothesized model is tested through analyses of variables to determine to what extent the model is consistent with data. If its goodness-of-fit is adequate, it means that the model is plausible, but if the goodness-of-fit is inadequate, it means that the model is not plausible (Byrne, 2016). SEM shows relationships between latent variables and observed variables. Latent variables are constructs which cannot be measured or observed directly, such as self-concept, motivation, verbal ability, and teacher expectancy. Latent variables must be operationalised through direct measurement of the observed variable(s) which will be used to represent latent variables (Wu, 2010). For example, observed measuring instruments include scores on an achievement test, self-reported responses to an attitudinal scale, and coded responses to interview questions. In a structural equation model, a series of equations show relations between latent variables and observed variables, or between latent variables. Testing the plausibility of a particular model involves finding to what extent hypothesized latent variables contribute to explaining the variance in a dependent variable.

In this project, two structural equation models of LC were built based on theoretical knowledge and available empirical studies of individual differences which explain the variance in L2 LC. Data collected were analysed to test against the SEM models. There may be other factors, e.g., metacognitive strategies, language learning anxiety, motivation, and cultural background knowledge, which can explain the variance in L2 LC, but not all can be included in the models.

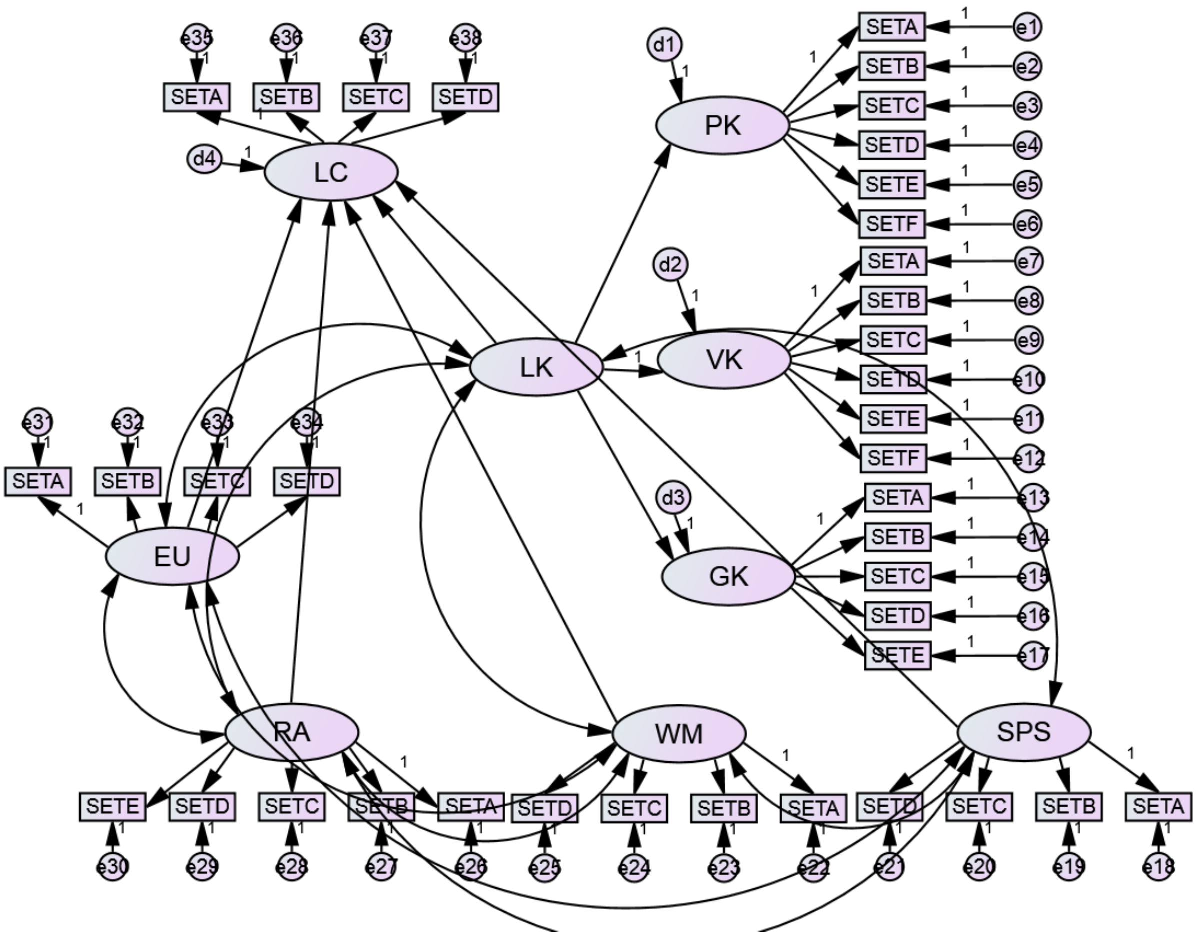


Figure 3.1 The first hypothesized key structural equation model based on Andringa et al. (2012)

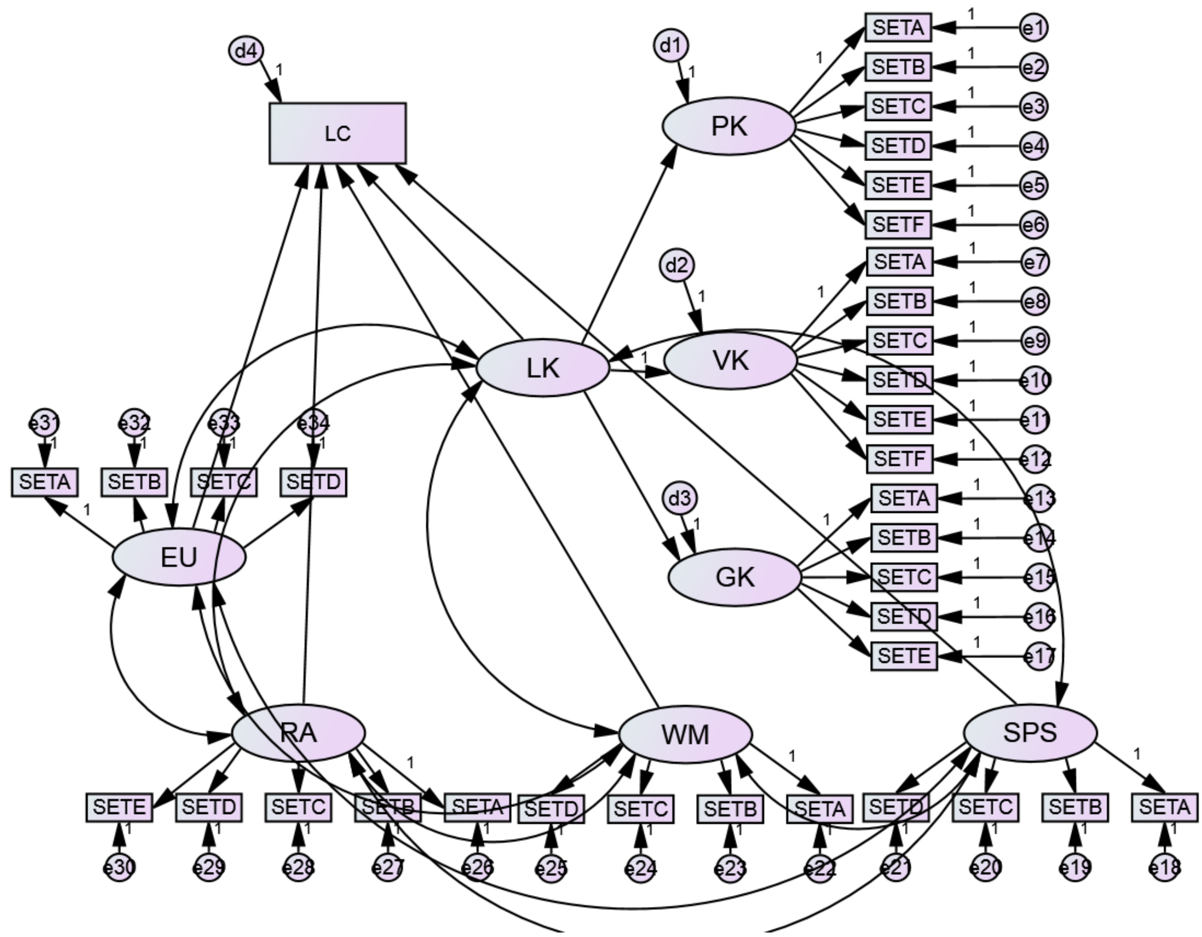


Figure 3.2 The second hypothesized key structural equation model based on Andringa et al. (2012)

Figures 3.1 and 3.2 are based on Andringa et al. (2012). In the two models, square boxes represent observed variables e.g., L2 learners' performance on grammar tests and their performance on an aural vocabulary size test. Circles in the model represent unobservable latent variables e.g., L2 learners' Phonological Knowledge (PK), Linguistic Knowledge (LK) and aural Sentence Processing Speed (SPS). Single-headed arrows represent the impact of one variable on another, and double-headed arrows represent covariances or correlations between pairs of variables, "e" refers to measurement error and "d" refers to unexplained variance in latent variables.

In this project, SEM and hierarchical multiple regression analyses are adopted to answer the first research question so that it can be found from the study which model can explain more variance in LC. The aim of the first research question is to explore individual differences which can explain variance in L2 LC among adult Chinese learners of English.

To answer this question, based on the available literature (see Section 2.1), five latent variables are assumed. The process of testing the SEM models involves finding which latent variables contribute to explaining the variance in L2 LC, and the associations between each latent variable or between observed variables and the dependent variable. In a simple or multiple regression analysis model, there are only one dependent variable and one or more independent variables, but in a SEM model there are both latent and observed variables. According to Schoonen (2015), one of the advantages of SEM is that when L2 researchers propose hypotheses about relationships between observed scores and underlying latent variables, they can test the tenability of their hypotheses about latent variables. SEM is able to deal with multiple dependent and independent variables and can be used to uncover very complicated relationships between variables (Schoonen, 2015). It can not only uncover the bivariate relations as addressed in a simple regression, but can also uncover the multivariate relationships addressed in a multiple regression analysis (Schoonen, 2015). Another advantage of SEM is that substantive analyses can be undertaken in the structural part of a model (Schoonen, 2015). If collected data are modelled in a measurement model and the fit of the model is good, then substantive hypotheses with latent variables can be tested error-free (Schoonen, 2015). Hierarchical multiple regression analysis is used to analyse the data collected and answer the three research questions. The second research question explores the contribution of each learner variable to explaining the variance in LC. Hierarchical regression is more suitable because by adding variables step by step in analyses, changes of contribution of each added variable can be found by partialing out the impact from other added variables. Hierarchical multiple regression analysis is also used to answer the third research question on variables which can explain differences in LC between subgroups in China and in the UK.

3.2 Participants

In total 220 students participated in the study, but the data of 33 participants could not be used in the analysis because they were incomplete on at least one of the tasks for the main study. Complete data sets were obtained from 147 students in China and 40 students in the UK (Male = 90, Female = 97). All participants were undergraduate university students for whom English was a second language. They had learned English for 10.5 years on average. The group of participants in China (from now on Group 1) were first-year non-English major students and the group of participants in the UK (from now on Group 2) were non-English students who had studied at a UK university for at least 12 months.

For Group 1 (Male = 84, Female = 63), the age range was from 17 to 22 (*M* = 19, *SD* = 0.8). Only students who had registered for the CET4 to be held in June 2017 took part, as it is only possible to obtain test results for students officially registered for the test, and only registered university undergraduates can take the test. The CET4 test is a national English test developed by the Ministry of Education of China with the purpose of measuring Chinese undergraduate students' English proficiency and promoting the implementation of *College English Curriculum Requirements* (Ren, 2011). The test is held twice each year in China, in June and in December. About two months after taking the CET4 test, test-takers receive an official report which gives a total score along with individual scores for each subsection (reading, writing, listening, and translating). In the present study, participants' scores on the subsection of listening in the CET4 held in June 2017 were used as one of the two measurements for their LC. Group 1 had studied English for an average of ten years at the time of data collection, ranging from six years to seventeen years. The minimum age at which they started learning English was three years. None had studied in the UK or in any other English-speaking country. Most learned Mandarin Chinese from the age of one, but they spoke local dialects at home since birth. One spoke Korean in the home from one year old, and still does, although Group 1 predominantly use Chinese at home. The data from the questionnaire showed that only five participants felt comfortable speaking English in every day conversation from the age of twelve, or later. The results of data analysis indicate that the rate at which Group 1 use Chinese (Mandarin and local dialect) in everyday different activities/topics (work, study, cleaning, leisure, trips, education, sports, politics, among others) is 95.67%, and the rate they use English in such activities and topics is 4.33% (see Appendix 1).

For Group 2 (Male = 6, Female = 34), the age range was from 18 to 23 (*M* = 20.5, *SD* = 1.1). In order to explore whether learners' English language learning background impacts LC, Group 2 were recruited on the condition that they had at least 12 month's experiences of studying in the UK as a university undergraduate student prior to taking part in the study. Group 2 had studied English for an average 12.6 years, ranging from nine to eighteen years at the time of data collection. The minimum age at which they started learning English was four years. The duration of their time studying in the UK ranged from 12 to 36 months (*M* = 16.1, *SD* = 6.6). More than half of this group learned Mandarin Chinese from the age of one, but spoke it from birth. Although most of them responded that they predominantly use Chinese at home, two claimed that they used both Chinese and English at home. Fourteen out of 40 felt comfortable speaking English in every day conversation from the age of twelve, or later. As with Group 1, Group 2 carried out number computations in Chinese when doing mental maths (e.g., computing 243 x 5). Three participants in Group 2 felt that they had lost fluency in Chinese at the age of 16, 18, and 21 respectively. The results indicate that Group 2 participants used English in different everyday activities/topics (work, study, cleaning, leisure, trips, education, sports, politics, among others) 51.12% of the time (see Appendix 1). They used Chinese in everyday activities and topics 48.88% of the time. On average, Group 1 used English in daily activities/topics (M = 4.33, SE = 0.33) less than Group 2 (M = 51.12, SE = 2.61) (see Appendix 2). This difference, -46.78, BCa 95% CI [52.10, -41.46], was significant, t (40.25) = 17.76, p < .001, and represented an effect of d = 11.73. (see Appendix 3). Five participants from Group 2 had taken the CET4 test in China before they came to the UK to study. Their total scores for the CET4 test ranged from 454 to 568 out of 710.

The English proficiency levels of the two groups were compared. Based on the results of data analyses from the questionnaire, the self-rating of English language proficiency (see Appendix 4), comprised of four subskills (speaking, writing, listening, and reading), of the two groups is significantly different. On average, the proficiency of Group 1 (M = 17.10, SE = 0.59) was lower than that of Group 2 (M = 22.10, SE = 0.90) (see Appendix 5). This difference, -5.00, BCa 95% CI [7.14, -2.85], was significant, t (76) = -4.64, p < .05, and represented an effect of d = 0.7 (see Appendix 6). All participants' total scores on the Oxford Placement Test (OPT) listening and grammar sections (the OPT, Allan, 2004) were calculated. According to guidelines for the OPT (Allan, 2004), test-takers' scores can be mapped onto the Common European Framework of Languages (CEFR) levels A1 to C2. For example, according to Allan (2004), if a learner's total score on the OPT is between 120 and 134, the learner's English language proficiency corresponds to level B1 on the CEFR scale; if a learner's score on the OPT is between 135 and 149, their proficiency level is mapped to level B2 on the CEFR scale. The results of analyses of the OPT original scores for Group 1 indicated that the mean score was 135.59 (SD = 12.78) (see Appendix 7), ranging from 100 to 172, which sits just on the bottom line of the B2 level on the CEFR scale. However, according to my own experience of teaching Chinese university undergraduates who share many characteristics with Group 1, and based on the descriptors of language proficiency levels on the CEFR, the output from university undergraduates in China on writing and speaking tasks can be mapped to approximately B1 level, rather than B2 level. For example, it was common to find that first-year undergraduates used a double object construction NP-V-NP-NP in their writing tasks, as in *He gave her a pen*. Such construction is considered a critical feature which appears from level B1 onwards, whilst object control sentences (NP-VNP-AdjP), as in *He painted the car red*, would appear from level B2 onwards (Salamoura & Saville, 2010). Therefore, Group 1 is defined as at approximately B1 level on the CEFR.

The results of the analysis of the original OPT scores for Group 2 indicated that the mean score was 146.3 (*SD* = 15.02) (see Appendix 7) ranging from 113 to 177, which nearly corresponds with the top line of B2 level on the CEFR scale (135 – 149, as mentioned above). Therefore, Group 2 is defined as at approximately B2 level on the CEFR. Participants' English proficiency was also compared with an independent *t*-test on their total scores in the OPT test (the listening and grammar sections). On average, the proficiency of Group 1 (*M* = 135.59, *SD* = 12.78) was lower than that of Group 2 (*M* = 146.3, *SD* = 15.02). This difference, -10.71, BCa 95% CI [-15.91, -5.50], was significant, t (55) = -4.12, $p < .05$ (see Appendix 8). Since Group 1 did not take an IELTS test, it was not possible to evaluate the English proficiency level of Group 1 from IELTS. According to the responses from Group 2 to the questionnaire on the English proficiency level from IELTS scores on the CEFR scale, three were at the C1 level, and the others were at B2 level. According to the IELTS Official Test Centre, the IELTS 9-band scale corresponds to six levels of language proficiency on the CEFR, which is helpful for institutions to refer to for the appropriate level of overall language ability required for their institutions or courses, and also for test-takers to better understand their own level of language ability (IETLS official website). For example: IELTS scores between 4 and 5 correspond, approximately, to B1 level on the CEFR; IELTS scores between 5.5 and 6.5 correspond, approximately, to B2 level on the CEFR; and IELTS scores between 7 and 8 correspond to C1 level on the CEFR. I am aware that although the CEFR provides valuable information on language proficiency, it has some limitations in terms of context validity, theory-based validity, and scoring validity (Weir, 2005). It is not easy to define a group of learners as having language proficiency of the same level, as learners may be at B1 for listening but at B2 for reading, for example. In addition, to enrol for undergraduate study in the university where the data was collected, international students must meet the minimum requirement for English language, which is an IELTS score of 6.5.

Each participant in Group 1 received a present worth 25RMB for their participation, and each participant in Group 2 was paid £35 for their participation in the study. Although their rewards differ, each is in line with what is regularly paid to students for participation in similar projects in each country. For Group 1, the reward of 25RMB is roughly the value of a restaurant meal in China. For Group 2, the reward of £35 is in line with the university's rules on payments for such participation. Participants were offered a chance to obtain feedback on their performance in the study and for advice on how to improve their English LC. The study was sponsored by a Chinese university and the University of Reading, but the sponsorship was not enough to pay more than 40 participants. Although a disadvantage of the approach chosen for this study was that the rewards were different between the two groups, the participants were generally motivated. Some evidence for this can be seen from the overall acceptable, or very high, internal reliability of the tests used in the study (Cronbach's Alpha ranging between .66 and .97). Further evidence is that almost all participants completed all tests and there was a very low level of attrition.

In conclusion, Group 1 were at approximately B1 level and Group 2 were at approximately B2 level on the CEFR scale of English proficiency. The two groups were statistically significantly different in their daily English use, their self-rating English proficiency of four subskills, and their English proficiency as measured by the OPT.

3.3 Instruments

3.3.1 The CET4 listening test

Participants' L2 LC was measured with the listening section of the CET4 and the Cambridge Preliminary English Test (PET). As seen in Section 3.2, the CET4 test is a standardised comprehensive English test that is used widely in Chinese colleges and universities (Garner & Huang, 2014; Zhu & Zhu, 2007). Test-takers' scores on the CET4 are expected to show their English skills in reading, writing, listening, and translating. Students must register for the test three months beforehand and take the test at their college or university. For the listening section, the recordings are played once by each organising college or university and test-takers wear headphones. The listening section lasts 30 minutes.

The CET-4/6 Testing Committee provides testing specifications as guidelines to test-takers and university teachers of English. In terms of LC, the aims of testing include measuring test-takers' abilities to comprehend various contexts, including: a slow-paced short English broadcast; multi-turns in simple English dialogues on familiar topics; slow-paced long speeches and reports on familiar topics; and also their ability to make use of basic listening strategies to help with comprehension. The speed of aural texts is 120-140 words per minute. According to the CET-4/6 Testing Committee, generally, the listening section measures test-takers' abilities to obtain aural information, including: comprehending themes, main ideas, important facts, and specific details; and inferencing implicit information and communicative functions of speech; as well as speakers' attitudes and opinions. Specifically, the listening section aims to measure test-takers' abilities:

A. to comprehend explicit information (themes, key or specific information, speakers' viewpoints or attitudes which are conveyed explicitly);

B. to comprehend implicit information (inferencing implicit information, deciding communicative functions of speech, inferencing speakers' opinions and attitudes);

C. to comprehend aural input based on language features (distinguishing phonological characters, e.g., distinguishing similar phonemes, noticing syllables stressed and different intonations from continuous speech, comprehending relations between sentences, e.g., cause and effect, compare and contrast); and,

D. to adopt listening strategies (suitable strategies to help with comprehension) (CET-4/6 Testing Committee, 2016).

The maximum score for the CET4 listening section is 249 out of 710, which makes up 35% of the total CET4 score. The listening part includes three sections: the first includes three news reports, each of which is followed by two to three questions; the second includes two long conversations, followed by eight questions; the third section includes three short monologues, followed by ten questions. All 25 questions are in multiple choice format.

For example, test-takers will hear a piece of news:

Thousands of bees left a town after landing on the back of a car when their queen got stuck in its boot. Tom Moses who works at a nearby national park, noticed a "brown patch" on the back of the car after the owner parked it to do some shopping. When he looked closer he realized it was a huge group of bees.

Moses said: "I have never seen that many bees in one spot. It was very unusual. They were very close together and there was a lot of noise and movements, it was interesting to see *such a strange sight. But there were a lot of people around and I was a bit worried about the bees and the people stopping to look. I thought that someone might do something stupid. Moses called two local bees specialists who helped removed the bees by attracting them into a box.*

Moses spent three hours looking after the bees and was stung five times, he said "my stings are a bit painful but I am pleased that all worked out and I could help, people need to realize that bees are valuable and they should be looked after".

After listening to this news report, test-takers hear 2 - 3 questions based on what they have heard. One of the questions is *What do we learn about Tom Moses?.* Test-takers read four options on their test paper and are expected to choose the correct answer from four alternatives

He is a queen bee specialist.

B) He works at a national park.

C) He removed the bees from the boot.

D) He drove the bees away from his car.

According to Zhu and Zhu (2007), the CET4 listening has high content validity because items in this section involve characteristics of authenticity and communication, and item types are designed to measure test-takers' LC and communication abilities. The main reason for choosing the CET4 listening to measure participants' L2 LC is that Chinese learners of English in universities are very familiar with the types of tasks in the listening section of the CET4 and are familiar with the topics which are closely related to university students' life and study (Wang & Treffers-Daller, 2017). In addition, one of the purposes of the study is to formulate implications for teachers of English, learners of English, and test developers in China. These stakeholders are likely to be more interested in results from a study which focuses on the CET4 than from a study which uses a different (non-standardised) test, or a test developed for a different context.

3.3.2 The Cambridge Preliminary English Test (PET) listening section

PET is a comprehensive exam developed by Cambridge English Language Assessment. The exam is designed for B1 level of CEFR and is accredited in the UK as an Entry Level 3 Cambridge English Language Assessment certificate (Cambridge English Language Assessment, 2014). According to Cambridge English Language Assessment (2014) the PET is ideal for language learners who need to communicate in English in a practical, everyday way. For example, learners might need to use English to handle most situations they meet when travelling in an English-speaking country. The PET measures test-takers' skills in reading, listening, writing, and speaking. The components and the test focus of each section differ. The listening section includes 25 items (25 marks in total) and represents 25% of the total marks for the exam. The listening section aims to assess test-takers' abilities to comprehend dialogues and monologues in both informal and neutral settings on a range of everyday topics. Those include daily life, the environment, hobbies and leisure, transport, and personal identification, among others, and all are based on authentic situations. There are four parts to the listening section, ranging from short exchanges to longer dialogues and monologues. Part 1 is comprised of seven short neutral or informal monologues or dialogues, each of which is followed by a three-option multiple-choice item with pictures. The focus of this part is to measure test-takers' ability to identify key information from short exchanges. Part 2 is a longer monologue or interview (with one main speaker), followed by six three-option multiple-choice items. This part focuses on measuring test-takers' abilities to identify specific information and detailed meaning. Part 3 is a gap-filling task. It is a longer monologue followed by six items in each of which there is a gap to be filled, so test-takers have to write one or more words in the gap. This part measures learners' ability to identify, understand, and interpret information. Part 4 is a longer informal dialogue followed by six True/False items. Test-takers have to decide whether the six statements in the test paper are correct or incorrect. This part focuses on measuring test-takers' ability to listen for detailed meaning and to identify attitudes and opinions of speakers. Each text is heard twice. The listening section lasts 35 minutes which includes six minutes given to transfer answers to an answer sheet.

For example, test-takers will hear a question and a short monologue:

Five. What is the woman phoning about?

Woman: Hello, this is Sarah Wright. I arranged to collect a guitar I ordered. I was meant to come this afternoon, but there is a problem. I had to get a book from the library and I've just missed the bus – so I won't be able to get out to the shop before you close. I've got an appointment in the town centre to choose some new glasses tomorrow, so I could come in and pick it up then. I hope that's ok.

Test-takers hear the recording twice. On the test paper, test-takers will read the question What is the woman phoning about? and they are supposed to choose one picture from three picture options:

A) a guitar

B) a pair of glasses

C) a book

The main reason why the PET listening was chosen as the second tool to measure participants' LC is that it is an international, standardized exam which is recognised for purposes of business, study, and immigration by employers, further education institutions, and government departments (Cambridge English Language Assessment, 2014). The second reason is that the PET is designed to correspond to approximately B1 level of the CEFR, i.e., the proficiency level of the Group 1 participants. The test was piloted before it was used in the main study and evidence from that pilot study showed that the test was not too difficult for adult Chinese learners of English (see Section 4.3 for details). The third reason for using the PET is that by comparing its listening section with the CET4 listening section, recommendations can be formulated for test developers in HE in China.

3.3.3 Tests of phonological knowledge

The focus of testing phonological knowledge in this research project is on measuring L2 learners' ability to recognise spoken words in context because evidence in the literature indicates that Chinese learners experience many decoding problems in English LC. Y. Wang (2008) found that learners had problems segmenting a continuous speech stream into words or phrases. Goh (2000) found that learners do not recognise words they actually knew, or could not chunk streams of speech. Similar decoding problems were found by F. Wang (2008). Sun and Li (2008) and Zhang et al. (2010) confirmed that learners had more listening problems in the perception stage of comprehension than in the parsing and utilization stages. Therefore, it is important to measure learners' word recognition abilities in the current study.

Test of auditory discrimination in English

Participants' auditory discrimination competence in English was measured using the Oxford Placement Test 1 Listening Test (OPT listening section, Allan, 2004). OPT is comprised of two sections (a listening section and a grammar section) to test language skills and knowledge of English as a language system. The OPT listening section is primarily a test of reading and listening skills, but also of vocabulary size (Allan, 2004). The test measures listeners' discrimination ability in hearing a target word which is inserted in a sentence and has a minor difference with another word in spelling. This test consists of 100 items derived from a corpus of several hundred examples of "slips of the ear" recorded over a number of years in native and non-native English speakers' conversations. Listeners' performance depends on how they apply knowledge of the sound and writing systems of English, and also on their ability to make use of this knowledge at a task-speed well within the competence of a native English speaker (Allan, 2004). According to Tauroza and Allison (1990) an average speech rate for conversations in British English is about 210 wpm. Although this is faster than the average speed of the Gaokao English listening test in China, which is about 120 wpm, the OPT listening test does not aim to measure listeners' comprehension ability. In addition, listeners hear sentences which are also given in written format with a target word and an alternative word from which listeners choose. This helps listeners adapt to the speed rate. According to the designer of the OPTs, the listening section helps to identify test-takers' phonological problems. All items represent whether test-takers can adapt from a test format to an actual situation. If a test-taker has inadequate listening skills, they are likely to fail in communication, or transmit the wrong meaning (Allan, 2004).

During the test, participants were given the Listening Test paper and then they listened to the recordings. In each item on the test paper, a minimal pair of easily confused words were included in a short carrier sentence. Participants were required to choose one option from the two answer options after they heard the sentence. Each item was said only once at normal speaking speed and the recording (on CD) was not stopped during the test, which lasted about ten minutes.

For example, participants read the carrier sentences on the test paper:

(1) *I doubt if he's very comfortable in his present/prison bed.* (Item 4)

(2) *Why/Where are you going to live in London?* (Item 21)

Only one of the two words (e.g., p*resent/prison*) was pronounced. Since the target sentences were played one after the other at normal speaking speed and there was almost no pause between two sentences, participants had no time to read the next item before hearing it. As shown in examples (1) and (2), the two options in the carrier sentence were both grammatically and semantically acceptable in English. Therefore, it was not possible to choose the correct answer without listening to the item (Allan, 2004).

According to Allan (2004), the OPT listening test has high reliability across test populations. Only items that were consistently answered correctly by the trial groups of native-speakers were included in the bank of items from which the final 100 were drawn. The designer claims that the discrimination indices of particular items show a high level of consistency from one large multilingual sample to another. All items were tested over a five-year period on multilevel samples of students over 40 different nationalities. Zoghlami (2015) used the OPT listening test in her study on EFL learners' LC among first-year university students. Her results indicated that participants' scores on the OPT listening test significantly correlated with their performance on a L2 listening test, although rather weakly ($r = .39$, $p < .01$). Since the OPT listening test is a standardised test, it was not included in the pilot study.

We have to acknowledge that while completing the task, learners could predict more or less what was likely to come up later on the basis of what they had read. Since learners had the opportunity to read each item before they heard the recording, they had a model of each whole sentence in their mind and they were likely to use that information to predict what was about to come up. Therefore, they did not just listen to the bottom-up sound, but they had the top-down sentence information which made it easier to process the sentence.

Test of word segmentation

A word segmentation test was used to measure participants' ability to distinguish segments in a stream of speech. While processing running speech, L2 listeners have to segment continuous streams of sound into recognisable words or phrases (Ito & Strange, 2009). English speakers try to transfer syllables that begin with a vowel through resyllabification (Field, 2008), as in (3) where *it* begins with a vowel and listeners interpret the final consonant of *help* as being the first sound of *it*:

(3) *Can't help it → (carn) tell pit*

(4) *Great ape → grey tape*

Consonants are also reattached to form a cluster:

(5) *Need rain → knee drain*

(6) *Let's leave → let sleeve*

In order to identify how Chinese learners of English employ phonetic-acoustic cues e.g., presence of aspiration and presence of glottal stop, in English speech segmentation, a segmentation test was used. The test includes the same stimuli as used in Altenberg (2005).

Included are 84 stimuli (42 pairs), in addition to four pairs of practice stimuli used before the experimental stimuli begin. In studies conducted by Ito and Strange (2009) and Shoemaker (2014), the same stimuli were used, except that Ito and Strange (2009) added six extra stimulus pairs (12 stimuli) for familiarisation trials.

In the present study, the 42 pairs of stimuli are subdivided into three types: 18 aspiration stimulus pairs in which boundary cues depend on the presence or absence of aspiration of a voiceless-stop at a word boundary (e.g., + aspiration: *lace peach* vs. – aspiration: *lay speech*); 18 glottal stop stimulus pairs in which the boundary cue depends on the presence or absence of glottal stop or creaky voice (e.g.,+ glottal stop: *team at* vs. – glottal stop: *tea mat*); and six double cue stimulus pairs in which segmentation cue depends on either aspiration and no glottal stop (e.g., *my toe*) or a glottal stop and no aspiration (e.g., *might owe*).

The aspiration stimulus pairs and the glottal stop stimulus pairs were further divided into three groups. The division of the former was based on what segments surround the word boundary: a vowel-s-consonant (VsC) group (e.g., *lace car* vs. *lay scar*); a consonant-s-consonant (CsC) group (e.g., *chief's cool* vs. *chief school*); and a consonant-s-consonant-consonant (CsCC) group (e.g., *tops pry* vs. *top spry*). The latter were divided on the basis of the class of the pivotal consonant: a nasal group (e.g., *claim annual* vs. *clay manual*); a fricative group (e.g., *loaf ate* vs. *low fate*); and a liquid group (e.g., *beer ice* vs. *be rice*). There were no subcategories for double cue stimuli.

A two-alternative forced-choice identification task was employed in which participants listened to recordings of the carrier phrases one by one and were required to choose from two answer options on the answer sheet. The order of response was kept the same for a minimal pair on the answer sheet the two times they appeared, but the 84 stimuli were randomized with the help of a Random Numbering Generator. The stimuli of each pair were separated by at least fifteen other stimuli to avoid influence of participants' memory while choosing the correct answer for the other member of a pair. Correct and incorrect answers were counterbalanced on the answer sheet, i.e., the correct answer appeared on the left in half of the sentences and on the right in the other half. After choosing an answer option, participants were required to indicate how confident they were in their answers by choosing the number corresponding to their level of confidence from one of the seven boxes (1 = not sure at all and 7 = very sure).

All stimuli were produced in a carrier phrase *Say* **______** *again* in order to keep intonation constant. As the original recordings were not available, a native speaker of English was asked to record the stimuli for the current project. As in Altenberg (2005), in order to minimize the possibility of listeners' performance errors, the native speaker was instructed to speak naturally, but at a slightly slower rate than usual. The interval between two aural stimulus phrases was six seconds. The test lasted 13 minutes and the suitability of the test was realised in a pilot study.

For example, participants heard two items, as in the following sentences (7) and (8), then read the answer options and chose one number to indicate how confident they were in their answer:

(7) *say lie told again* (Item 1)

(8) *say might owe again* (Item 2)

Items	Answer options		How sure are you of your answer?						
1	A. light old	B. lie told	1	2	3	4	5	6	7
2	A. might owe	B. my toe	1	2	3	4	5	6	7

The test was piloted among English speakers and Chinese learners of English before it was used in the main study. Results of the pilot indicated that mean accuracy of target items was 86.3% for English speakers and 72.9% for Chinese participants in the UK, which means that the test is suitable for Chinese learners from the perspective of its difficulty. Mean confidence ratings of target items were 5.7 for English speakers and 5.0 for Chinese participants in the UK. The internal reliability of the test (Cronbach alpha) was .80, which means the test has high internal reliability.

We have to acknowledge that while completing the task, the learners did not only use the information from bottom-up processing, but they could also have tried to retrieve in their memory chunks that were stored in memory, which means that top-down processing could have played in role in this task.

Test of word recognition from speech

Matthews and Cheng's (2015) test of word recognition from connected speech (WRS) was used to measure listeners' ability to recognise the aural form of a word in a stream of speech, comprehend the word meaning, and produce the word in written form, with one more item added by the designers in the version used in the current study.

The WRS test (Matthews & Cheng, 2015) consists of 89 items. Each item was selected by the designers from two test instruments on a measurement of L2 WRS (Matthews, Cheng, & O'Toole, 2015). The test focuses on high frequency words which belong to the 1000 to 3000 word frequency levels in the BNC/COCA word family lists. Previous studies indicate that the most frequent 3000 word families are sufficient to reach a 95% coverage level in spoken language (Adolphs & Schmitt, 2003; Nation, 2006; Webb & Rodgers, 2009). The WRS test consists of 23 target words from the 1K word frequency level, 37 target words from the 2K word frequency level, and 29 target words from the 3K word frequency level; 89 target words were selected in order to ensure that there were enough target words from each of the three word frequency levels. The frequency levels of content words in each written stimulus sentence were also checked with the constraint, mainly in the 1K word frequency level, with a few from the 2K. Each item was a single short sentence in which the target word was embedded. Participants read the stimulus sentence with a blank for the target word on the answer sheet. They then listened to each stimulus sentence and were required to write the target word in the blank space after it was heard. The recording was played only once. The test lasted 17 minutes. Sentences (9) to (11) are three examples from the test:

(9) *The number of people in this country is* ____________ *growing*. (Item 47)

(10) *Milk is an important* ____________ *product of this country*. (Item 58)

(11) *The protection of Vietnam's forests is an important* ____________. (Item 86)

Before using the test in their experiment, the designers had shown the written stimuli sentences to native speakers in order to ensure that the target word could not be selected immediately, but only by understanding its written context. The test was also piloted further to ensure that native speakers could recognise the target word readily after hearing the aural stimuli sentences. Matthews and Cheng (2015) provided some evidence for the validity of the test with the groups they studied. In their study participants' mean score on the test was 63.82 out of the maximum score of 89 (accuracy = 71.71%). The mean accuracy on the 1K word level was 81.96%, followed by that on the 2K level (73.98%), and that on the 3K level (60.71%). While using the test to measure participants' levels of WRS, the authors found that knowledge of words from the 1K and 3K frequency ranges predicted 54% of the variance in LC, with measures of words from 3K frequency range alone predicting 52% of this variance. The participants were adult Chinese learners of English at a university in China, with an average age of 19.42 years. The authors also found the test to be a reliable instrument with high internal consistency (Cronbach's alpha = .91).

Since Matthews and Cheng (2015) have found the WRS test to be appropriate for Chinese students with a similar level of ability, the test was not piloted in the present study. I decided to use the test in the present study because Chinese learners have many LC problems in recognising words from speech (see details in Section 2.2) and this may provide additional information about the validity of this test in the Chinese context.

3.3.4 Vocabulary size tests

In this project, an aural vocabulary size test and a written vocabulary test were piloted to measure learners' L2 vocabulary knowledge before data collection for the main study began. However, out of consideration of time limitations, I decided not to use the written test in the study. Both tests are introduced in this section.

An aural vocabulary test

The Listening Vocabulary Levels Test (LVLT, McLean, Kramer, & Beglar, 2015) was used to measure participants' aural vocabulary knowledge. According to McLean et al. (2015) the LVLT was designed to measure Japanese learners' aural vocabulary knowledge of English words from the first five 1K frequency levels and the Academic Word List (AWL). An aural, rather than a written, vocabulary test was used in the present study because L2 learners' aural and written vocabulary knowledge differs (Field, 2008; Milton & Hopkins, 2005). In an aural vocabulary test, L2 learners have to make use of their knowledge of English phonology, rhythm, and stress patterns (McLean et al., 2015). The LVLT is comprised of six parts with 150 items in total. In each of the first five parts, 24 items are included measuring 1K frequency on one level. In the sixth part, 30 items are included which measure the AWL. The LVLT items were chosen from the British National Corpus (BNC) / Corpus of Contemporary American English (COCA) list (Nation, 2012), the first five 1K frequency levels of which provided adequate coverage for listening across a wide range of genres (McLean et al., 2015) and provided nearly 96%-97% coverage of conversations (Nation, 2006). The AWL was included in the LVLT because it covers 10% of tokens in academic texts and 4.41% of academic spoken English (McLean et al., 2015). According to the designers, having knowledge of AWL vocabulary is a prerequisite for students taking academic English programmes throughout the world.

In this study, as in the LVLT, participants heard a word in a simple carrier sentence; this provided context in the event that the target word had more than one possible meaning or use. The designers claim that the carrier sentence will not give clues about the meaning of the target word. Each item has four answer options. After hearing the target word and the carrier sentence, participants chose one option in the Chinese version with the closest meaning to the target word. The four answer options are provided in Chinese in order to avoid the possible confusion of measuring listeners' aural vocabulary knowledge and L2 reading ability (McLean et al., 2015). There was a five-second pause between the reading of each item during which participants could process the aural input and choose one answer option. Participants were given a 15-second pause to allow enough time to turn test pages and prepare for the next section. The recordings were played only once and the whole test lasted 35 minutes.

For example:

(12) *basis: This was used as the <basis>*. (Item 4 in LVLT Part 2)

A 答案　　B 休息的地方　C 下一步　D 基础，依据

(13) *rove: He is <roving>*. (Item 9 in LVLT Part 5)

A 喝醉　　B 漫游，徘回　C 哼歌　　D 努力

McLean et al. (2015) claim that the LVLT has high validity. The designers provide evidence for the content aspect of construct validity, including content relevance and representative and technical quality. They also provide construct validity evidence for the substantive aspect involving the rationale for the item difficulty and the personal ability hierarchies, the structural aspect, and the generalizability aspect as well. Qualitative investigations into the LVLT, provided by the designers, support its high face validity and that the test is easily understood. The test was piloted in the current study before it was put into use in the main study. Results of the pilot study indicated that the test had high internal reliability (Cronbach's alpha = .89) (see Section 4.1). The mean score found in the pilot was 97 out of 150 (*SD* = 17.57), and the mean accuracy score was 64.7%, which means the test was not too difficult for Chinese learners. Since the LVLT is a relatively new test, it still needs to be applied in a variety of contexts; it is expected that its use in this PhD project will help with its validation.

A written vocabulary test

The written vocabulary size test piloted was the Vocabulary Size Test (the VST, Nation & Beglar, 2007). The VST is a receptive vocabulary test which is widely used in studies of both L1 and L2 acquisition. The vocabulary was drawn from different frequency layers in English (Nation & Beglar, 2007). The test consists of 140 multiple-choice items, with 10 items from each 1K family level. A reduced version was piloted which consisted of 50 items (from 1K family level to 5K family level). Participants in the pilot were first-year university students in China who knew about 3000 English word families on average. In the test, each target word is presented in a simple carrier sentence which reveals a word class, but does not provide any clues about the meaning of the word. Participants had to choose one answer from four alternatives. The four alternatives consisted of one correct and three incorrect definitions of the word meaning in Chinese because in this way participants' ability to understand the target items is confounded with their ability to read answer options in the L2 (Nation & Beglar, 2007). Participants' total receptive vocabulary size was calculated through the number of correct responses multiplied by 100. The test lasted 15 minutes.

For example:

(14) *drawer: The drawer was empty.* (Item 4 in Second 1000)

a. 抽屉 b. 车库 c. 冰箱 d. 鸟笼

(15) *deficit: The company had a large deficit.* (Item1 in Fifth 1000)

a. 出现赤字 b. 贬值 c. 有这笔大开销的计划 d. 在银行里有很多存款

Pilot results indicate that the test had internal reliability (Cronbach's alpha = .71) and the mean accuracy rate was 60%, which provides some evidence that VST was not too difficult for the participants in the study, but the task was not used in the main study.

3.3.5 Tests of grammar knowledge

Participants' grammar knowledge was measured using two offline grammar tests, one in written and the other one in aural form.

A written offline grammar test

The written offline test was the Oxford Placement Test Grammar test (the OPT grammar, Allan, 2004). The test is comprised of two parts. In total 100 multiple-choice grammatical and lexical items are contextualized or thematically linked. In Part 1, items from 1 to 20 are short individual sentences in which grammar knowledge is embedded, as in (16) and (17), and participants made decisions based on their grammar knowledge.

(16) *Mohammed Ali has won/won/is winning his first world title fight in 1960*. (Item 11)

(17) *If he has/would have/had lost his first fight with Sonny Liston, no one would have been surprised.* (Item 14)

Participants had to choose one option from the three answer options provided. Items from 21 to 50 are included in a short passage. In order to choose correct options, participants must have a good understanding of the context. Therefore, according to Allan (2004), participants' grammar knowledge, vocabulary knowledge, and reading skills were all tested. Part 2 is very similar to Part 1, except that items from 91 to 100 are sentences, with a different question tag for each, and participants were required to choose the correct tag from three answer options. Four practice sentences were given at the start of the test. Participants read the following testing sentences:

(18) *John's coming to see you, hasn't he/ wasn't he/ isn't he?* (Item 91)

(19) *I think I'm expected to pick him up, aren't I/ don't I/ are you*? (Item 96)

Participants answered the questions by simply ticking the correct answer option. Each item was worth one point and the maximum score was 100. Allan (2004) claims that the test items were analysed across a range of contexts and populations that had a consistent track record of providing discrimination between and within the Common European Framework (CEF) levels. The items reflected extensive research into syllabus and coursebook content knowledge. According to Allan (2004), all the items were tested over a five-year period on multilevel samples of students involving over 40 different nationalities, and the test item reliability across test populations is very high.

The participants adapted to the formats of the OPT grammar test very quickly because they had done a large number of such English grammar exercises before taking the University Entrance English Exam. In addition, they were familiar with the multiple-choice format. Sentences (20) and (21) were two items in the 2016 University Entrance English Exam (for Beijing city) in which students' grammar knowledge was measured.

(20) *Jack ______ in the lab when the power cut occurred.* (Item 21)

A. works *B. has worked* *C. was working* *D. would work*

(21) *Why didn't you tell me about your trouble last week? If you ______ me, I could have helped*. (Item 34)

A. told *B. had told* *C. were to tell* *D. would tell*

The task was piloted with a small number of participants (N = 11) and results showed that the mean score was 67.8 out of 100 and that there was no ceiling or floor effect. Therefore, the task was considered to be of an appropriate level for the participants. Detailed results of the pilot are presented in Section 4.2.

An aural offline grammar test

The Test for Reception of Grammar version 2 (TROG-2, Bishop, 2003) was used to measure participants' grammar knowledge in aural form. This comprehensive grammar test was designed to assess understanding of English grammatical contrasts which are marked by inflections, function words, and word order. According to Bishop (2003), by using this test researchers can not only discover how a participant's grammar comprehension compares with that of other participants, but they can also find the participant's specific area of difficulty.

The test is comprised of 80 items. Participants' understanding of each item is assessed through their answers. They have to choose an answer from four options which consist of four pictures, one depicting the target sentence and the other three depicting distractor sentences. Distractor sentences have been altered by substituting a grammar or lexical element with an alternative form. Target grammar knowledge covers a wide range of constructions, including negation, reversible structure of Subject + Verb + Object (SVO), comparative/absolute, relative clause in object, and centre-embedded sentences. There are 20 blocks in total and there are four items in each block. The blocks are arranged in order of increasing difficulty.

Sentences from (22) to (25) are examples of the test in one block:

(22) J1 *The duck is bigger than the ball*

(23) J2 *The tree is taller than the house*

(24) J3 *The pencil is longer than the knife*

(25) J4 *The flower is longer than the comb*

The construction tested in the block is comparative/absolute. For each item, participants heard a sentence and saw the four choices consisting of four pictures on a computer screen. Participants were required to choose one picture as their answer on the answer sheet. For example, for sentence (22), participants saw the four pictures which represented different scenes corresponding to the meanings expressed by the following sentences:

A. *The duck is bigger than the ball*

B. *The duck is the same size as the ball*

C. *The duck is smaller than the ball*

D. *The duck is bigger than the shoe*

In order to choose the correct answer, the participants must have relevant grammar knowledge, namely, in this case, knowledge of comparatives and, in particular, knowledge of the use of the suffix *–er*, attached to the adjective *big*, and the comparative marker *than*, used to compare two items.

In the process of completing the task, four options (pictures) for each item were projected on a big screen and I read out loud each target item to the participants. After hearing a target item, the participants were given five seconds to choose one answer from four choices. The task lasted 12 minutes.

3.3.6 On-line sentence processing tasks

To measure participants' speed in processing aural sentences which consist of certain grammar knowledge, two on-line sentence processing tasks were piloted, but only one (a grammaticality judgment task) was applied in the study due to time limitations.

According to Roberts (2012), processing speed refers to efficiency in many different processes undertaken during reading and LC, e.g., decoding of written words or spoken input, vocabulary access and selection, integration with grammatical knowledge, and the prediction of incoming input. In the present study, processing speed is taken to be an indicator of participants' efficiency in processing complex spoken sentences.

The two tasks piloted were a listening grammaticality judgement task and an aural questionnaire. One of the aims of the former task was to measure whether, in listening tasks, participants were sensitive to number agreement violations between a potential antecedent and the auxiliary verb in a relative clause (RC) which modified the potential antecedents. The other aim was to measure whether participants' reaction time (RT) when making decisions during such tasks correlated with listeners' LC in other listening tasks. The aims of the aural questionnaire were to explore whether or not participants' RC attachment preferences were determined by a universal Recency Preference strategy (Fodor, 1998), whether L2 learners' preferences were more similar to those of native speakers or to those of non-native speakers, and whether participants' RT in making decisions correlated with their LC.

The focus of the two tasks was on RCs because RC is a difficult grammar point for Chinese learners of English. According to Biber, Johansson, Leech, Conrad, and Finegan (1999), in standard English, RCs can be formed using eight different relativizers (relative pronouns or relative adverbs): *which*, *who*, *whom*, *whose*, *that*, *where*, *when*, and *why*. Juffs (1998) found that L2 learners whose L1 was typologically different from English (e.g., Chinese, Japanese, and Korean) had more difficulty processing reduced RC ambiguities than did learners whose L1 was similar to English typologically. In Juffs (1998), a reduced RC is led by a passive participle, as in (26) *watched almost every day*, which is compared with an unreduced RC, *who were watched almost every day*. Reduced RC ambiguity might be caused temporarily by *watched every day* because the passive principle *watched* could be followed by a direct object.

(26) *The bad boys watched almost every day were playing in the park.*

In the present study, experimental written sentences used in Felser, Marinis and Clahsen (2003a) were adopted. Felser et al. (2003a) used those sentences among native English-speaking children and adults with the aim of comparing how the two groups processed ambiguous sentences with a RC attachment. The experimental sentences for the two tasks all consisted of a main clause which contained a transitive verb in a past tense. The verb was followed by a complex noun phrase (NP) object which contained a prepositional phrase (e.g., *the princess with the maid* or *the soldiers of the colonels*), which formed the antecedent of a subject RC (e.g., *who was eating chocolates*). The ambiguity arises from the fact that a RC can be interpreted as modifying either the first or the second NP. Thus, in (27) either *the princess* or *the maid* can be interpreted to be the antecedent of the RC *who was eating chocolates*. In other words, either *the princess* or *the maid* eats chocolates. Similarly, in (28) either *the pupils* or *the teachers* can be interpreted to be the antecedent of the RC *who were standing in the hall*. In other words, either *the pupils* or *the teachers* were standing in the hall. In the experimental sentences, a subject of a main clause cannot be an antecedent of a RC.

Sentences (27) and (28) are examples of experimental items:

(27) *The little girl envied the princess with the maid who was eating chocolates*.

(28) *The headmaster smiled at the pupils of the teachers who were standing in the hall*.

We created filler sentences where a relative pronoun *who* was replaced with *and*, or a RC was replaced with an adverbial clause. In these co-ordinated sentences a subject of a main clause had to be the same as a subject of the second clause. So, in that case *the little girl* was eating chocolates in (29) and *the headmaster* was standing in the hall in (30).

(29) *The little girl envied the princess with the maid and was eating chocolates*.

(30) *The headmaster smiled at the pupils of the teachers when he was standing in the hall*.

The task was used in a pilot study and results showed that the internal reliability was not high (Cronbach's alpha = .69) because a small number of participants (N = 16) took part in the pilot (see Section 4.1 for details). The problem was solved in the main study. Results also indicated that the mean accuracy of the task was 63.3% for Chinese participants, which was lower than that for English speakers (accuracy = 83.8%). Pilot results provided some evidence that instructions for how to complete the task were easy to understand and the software used to run the items and record participants' scores and RTs worked well on a computer.

Listening grammaticality judgement task

The listening grammaticality judgement task comprised 52 items, including four practice items, 24 experimental items (12 grammatical, 12 ungrammatical), and 24 filler sentences (12 grammatical, 12 ungrammatical). In half of the experimental items, two potential antecedents were in singular form, and in the other half, two potential antecedents were in plural form. Therefore, the following four sentence types were produced:

(31) Single ---- Grammatical:

The reporter phoned the boss of the secretary who was reading a book. (Item 4)

(32) Plural ---- Grammatical:

The coach looked at the football players with the fans who were very happy. (Item 29)

(33) Single ---- Ungrammatical:

The doctor recognised the nurse of the patient who were feeling very tired. (Item 9)

(34) Plural ---- Ungrammatical:

The woman blamed the hairdressers with the apprentices who was smiling all the time. (Item7)

As mentioned above, the two potential antecedents were connected either by *of* or by *with*. Therefore, in half of the grammatical experimental items and half of the ungrammatical experimental items, the two potential antecedents were connected by the preposition *of*, while in the other half, the potential antecedents were connected by the preposition *with*.

The 24 filler sentences were produced with the distinctions of grammatical and ungrammatical items, either *of* or *with* as a connector between two potential antecedents. Since the main purpose of using the filler sentences was to avoid participants noticing that the focus of the task was on RCs, the filler sentences did not contain RCs. Instead, the second half of the sentence consisted of an adverbial clause, as in (35), or another main clause which was attached to the first half with the help of a co-ordinating conjunction (e.g., *and*) which replaced the relative pronoun, as in (36). In all these cases the subject of the second clause was the same as the subject of the first clause. The filler sentences were kept at the same length as the experimental sentences. The filler-to-target ratio is 1:1. (as in Felser et al., 2003a; Traxler, 2002). Two examples of filler sentences are given in (35) and (36):

Grammatical:

(35) *The headmaster smiled at the pupils of the teachers when he was standing in the hall*. (Item 10)

Ungrammatical:

(36) *A strange woman called to the travellers with the guides and were about to cross the dangerous river*. (Item 18)

During the task, participants were instructed how to complete the test on a computer. At the beginning of the test, participants heard the recording of an item, when this ended, two answer options, *grammatical* and *ungrammatical*, appeared immediately. Participants then had to decide whether the item they had heard was grammatical or ungrammatical. After participants had made their choice, the next item appeared. Participants' answers and the RT needed to make a decision were recorded from the moment the sound of the spoken sentence ended and the answer options appeared on the screen, to the moment they pressed a button to choose one of the two options. The items were randomized with the help of a Random Number Generator. The option of *grammatical* was always on the left of the screen and *ungrammatical* was always on the right. After each 16 sentences, there was a break.

The task lasted eight minutes and was tested in a pilot study. Pilot results indicated that English speakers' mean accuracy score on target items in this task was 83.8%, much higher than that of Chinese learners (63.3%) (see Section 4.1 for details). In addition, English speakers' processing speed (RT = 1467 ms) appeared to be much faster than that of Chinese participants (RT = 2359 ms). The accuracy scores and RTs show that the task was not too difficult for Chinese participants, therefore, the task was used in the main study.

Auditory questionnaire task

The task was piloted, but was not used in the main study. The auditory questionnaire task comprised 24 items, including four practice items, ten target items, and ten filler sentences.

All target items were ambiguous and two potential antecedents and an auxiliary in a RC were in singular form. In half of the target items, two potential antecedents were connected by a preposition *of*, and in the other half by a preposition *with*. After hearing each sentence, participants heard a content question which was always headed with *Who*. Since ambiguity arises from the fact that a RC can be interpreted as modifying either the first or the second NP, in (37) either *the teacher* or *the doctor* can be interpreted to be an antecedent of the RC *who was preparing to go home*. In other words, either *the teacher was preparing to go home* or *the doctor was preparing to go home*. In (38) either *the fan was very happy* or *the actress was very happy*.

Sentences (37) and (38) are two examples of the target items:

(37) T*he nurse trusted the teacher with the doctor who was preparing to go home.*

Question: Who was preparing to go home?

A.the teacher　　　*B. the doctor*

(Item 1)

(38) *The student photographed the fan of the actress who was very happy.*

Question: Who was very happy?

A.the actress　　*B. the fan*

(Item 17)

Filler sentences were created to avoid participants noticing that the focus of the task was on RCs. Two NPs which followed the first transitive verb were connected by the preposition *of* or *with* (half *of* and the other half *with*). Filler sentences and experimental sentences were of equal length, but the filler sentences were unambiguous and did not contain a RC. In filler sentences, the relative pronoun *who* was replaced with *and* or *but*. In these co-ordinated sentences the subject of the main clause had to be the same as the subject of the second clause. Thus, *the journalist* allowed him to sit down in (39) and *the woman* found the singer reading a book in (40).

Sentences (39) and (40) are examples of filler sentences:

(39) *The journalist hated the colonel of the soldier but allowed him to sit down.*

Question: Who allowed him to sit down?

A.the journalist　　　*B. the colonel*

(Item 7)

(40) *The woman knew the photographer with the singer and found the singer reading a book.*

Question: Who found the singer reading a book?

A.the photographer *B. the woman*

(Item 9)

Participants had to choose from two answer options, which appeared on the screen, to show their preference for to NP1 or NP2. The two NPs in the target items were both possible answers. The NPs were always preceded by the definite article. One NP appeared on the left side of the screen and the other appeared on the right. In order to avoid participants always choosing NP2 (e.g., due to Recency Preference strategy, Fodor, 1998), the order of two possible answer options was counterbalanced. For example, in the target items in which *with* was a connector to the two NPs, if the order of two answer options was the same as the order of two NPs in one item, then in next item having *with* as a connector the order of the two answer options was opposite to the order of the two NPs in the item. In the target items with the connector *of*, the order of the two answer options was also counterbalanced in the same way. The order of the two answer options in filler sentences were counterbalanced respectively in *of* conditions and *with* conditions. All items were randomized using a Random Numbering Generator. Participants' answer and RTs in making a decision were recorded from the moment that the recording ended and the answer options appeared on the screen, to the moment they completed their choice. The task lasted 12 minutes and was tested in a pilot study. Pilot results are presented in Section 4.1.

We have to acknowledge that there are issues with the tasks which measured sentence processing speed. As we used very long sentences with complex grammar in the tasks, we measured not just learners' processing speed, but also learners' grammar knowledge. In order to comprehend the sentences, learners had to use their grammar knowledge and their world knowledge; in order to memorize the entire sentence while processing, or at least keep part of the sentences in memory all the time, they had to use working memory. Therefore, grammar knowledge, working memory and top-down processing might have impacted on the learners' performance. But it was very difficult to create tasks that measure only one construct and it is probably true that most tasks measure more than one construct. A vocabulary task, for example, measures non-verbal intelligence in many cases as well.

3.3.7 Working memory tasks

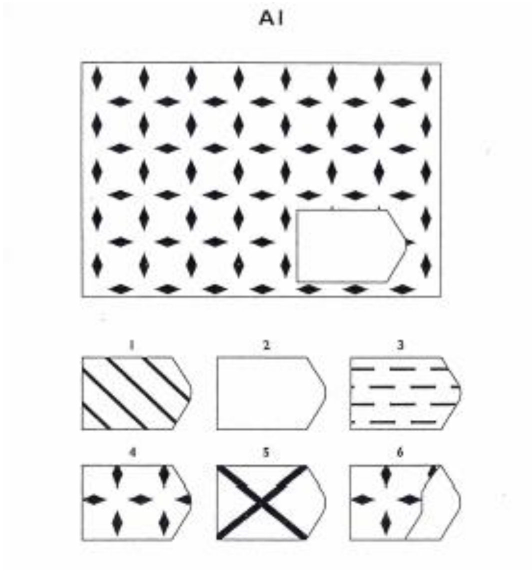
In order to investigate to what extent learners' working memory capacity influences their L2 LC, learners' working memory span has to be measured. In this research project, four digit-span tasks (forward and backward, in a Chinese version and an English version) were used as measures of verbal working memory capacity based on Wechsler (1997). Working memory tasks in both languages were used in order to explore which tasks (working memory span measured in L1 or measured in L2) contributed to explaining the variance in L2 LC. Gu (2007) studied relationships between working memory capacity and L2 LC among 59 adult Chinese learners of English. The researcher used the two versions' tasks to measure participants working memory capacity. She found that participants' working memory capacity, as measured in both languages, significantly correlated with listening proficiency. The correlation between LC and working memory measured in L1 (r = .54, p < .01) was lower than that between LC and working memory measured in L2 (r = .65, p < .01). This means that the tasks in different languages to measure working memory matter when exploring associations between the working memory and LC of L2 learners. In the current study, for the digit-span tasks, participants were asked to listen to a series of digits and to reproduce the series by repeating the digits orally in the same (forward) order, or in reverse (backward) order. The minimum series length was two digits. This was increased by one digit every two trials until the maximum length of nine (forward) digits or eight (backward) digits was reached, or until participants failed to correctly respond to both trials of a particular length. The digits were read at a rate of one per second. The sum of items that participants gave answers correctly represented a participant's digit span capacity, forward or backward, measured in L1 or L2. For example, I gave instructions in L1 for tasks to be taken in L1. In the item 2-6-9-3-5, after I read the trial participants had to repeat it (forward) in L1. When the tasks in L1 ended, I gave instructions in L2 to indicate that next task would be in English and they had to produce digits that they heard forward in English, e.g., 6-1-5-7. After each forward task ended, I reminded participants that for the next task they had to produce the digits they heard backward, e.g., for 4-7-2-9, they had to produce 9-2-7-4 in English.

The task lasted five minutes at most and was tested in a pilot study, results of which showed that participants' scores on tasks in L1 were greater than those in L2, which means tasks in L1 were easier to undertake than in L2. Pilot results also provided some evidence that instructions were easy to understand and that the working memory tasks were not difficult for the participants. Details of the pilot study results are presented in Section 4.1.

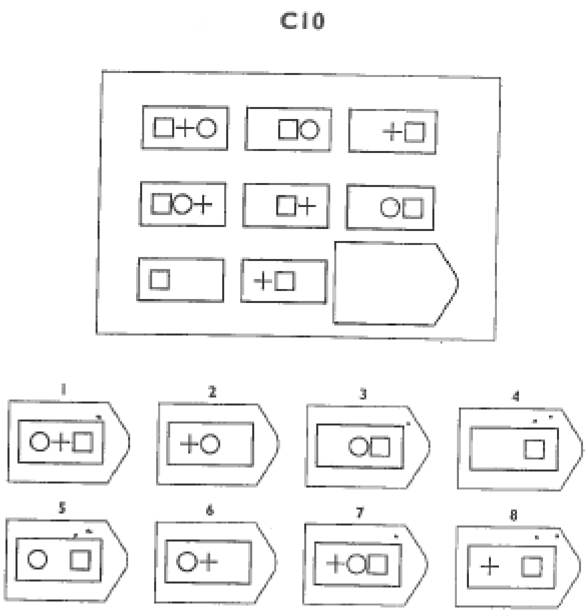
3.3.8 Raven's Standard Progressive Matrices

Non-verbal reasoning ability was measured through the complex matrices component of Raven's Standard Progressive Matrices (Raven's SPM, Raven, Court, & Raven, 1992), which is widely used to assess general non-verbal cognitive abilities. The test measures test-takers' reasoning ability to understand the relationship among abstract items. Test-takers are required to select which of six or eight pattern pieces fit best into an overall matrix by using their non-verbal abstract reasoning ability (Mills, Ablard, & Brody, 1993). The test comprises 60 items in five sets of 12 figures. Within each set, the difficulty of items increases progressively, so the easiest items serve as a learning experience for more difficult items (Mills et al., 1993). According to the designers, in this way, test-takers' ability to solve problems quickly by learning from their immediate past experience is assessed. Each set of the test has a different logic (Raven et al., 1992). Although I have made efforts to find whether or not the test has been previously used in China, so far I have found no evidence to suggest it has. However, according to Mills et al. (1993), the entire test measures very general reasoning ability which is not affected by test-takers' educational or cultural background. Participants provided answers on a test paper for the current study. Examples of the test are given in (41) and (42), (41) was chosen from Set A. Test-takers look at a big picture under *A1* with one part missing and decide which one of six alternative pictures, numbered from 1 to 6, is the missing part which best fits into the big picture. Example (42) was chosen from Set C. This item requires test-takers to find the logic underlying a group of eight smaller pictures under *C10*, and to select one from eight alternatives, numbered 1 to 8, which depicts the same logic and best fits the group. The task lasted 45 minutes.

(41)



(42)



The task was piloted with a small number of participants (*N* = 11). Pilot results showed that the mean score was 46.6 out of the maximum 60 (see Section 4.2 for details), which means the task was of an appropriate level for the participants.

3.3.9 A personal background questionnaire

In order to explore whether aspects of L2 learners' personal background impact on their performance in L2 LC, a personal background questionnaire was used to investigate L2 learners' language proficiency and language use experiences. The questionnaire was designed on the basis of the Quick Gradient Language Dominance Questionnaire (Dunn & Tree, 2009) and Grosjean (2015). In the questionnaire, participants were asked to provide information about the age at which they started learning English, which language was used dominantly at home, and which language they used to calculate numbers in mathematics. Participants provided self-rating scores for their proficiency in speaking, listening, reading, and writing in Mandarin Chinese, English, and other languages/dialects. They were asked the length of time they had been studying in the UK, and reported what proportion of time they spent using English and Chinese in daily activities, e.g., studying, playing sport, shopping, and communicating with both immediate and distant families. For example, participants read an item and wrote down a number or a word as an answer in a blank with a line for (43) and (44), or they read instructions first and filled in forms, as seen in (45), by providing numbers as answers.

(43) At what age did you first learn English ________?

(44) When doing math in your head (such as multiplying 243 × 5), which language do you calculate the numbers in? ________

(45) Please indicate below how often you use Chinese and English for the different activities/topics below. For example, if you use mainly English at work and very little Chinese, write 80% English and 20% Chinese. The totals have to add up to 100%

	English (%)	Chinese (%)	Other language namely…
work			
studies (in general)			
immediate family (with whom you live)			

It took participants five minutes to complete the questionnaire in a pilot study. Participants were instructed how to answer questions and did not find any items difficult to understand or difficult to answer. Therefore, the questionnaire was used in the main study.

Table 3.2 gives a brief description of the instruments used in the main study.

Table 3.2 A brief description of the instruments used in the main study

Instruments	Variables to measure	Time duration	Form	Notes
CET4 listening section (CET4/6 Testing Committee)	LC	30 minutes	Aural, receptive	
Cambridge Preliminary English Test listening section (PET listening, Cambridge English Language Assessment, 2014)	LC	30 minutes	Aural, receptive	Used in a pilot study
Oxford Placement Test 1 Listening Test (OPT listening, Allan, 2004)	Phonological knowledge	10 minutes	Aural, receptive	
Word segmentation test (WST, Altenberg, 2005)		13 minutes	Aural, receptive	Used in a pilot study
Test of word recognition from speech (WRS, Matthews & Cheng, 2015)		17 minutes	Aura, productive	
Listening Vocabulary Levels Test (LVLT, McLean et al., 2015)	Vocabulary knowledge	35 minutes	Aural, receptive	Used in a pilot study
Vocabulary Size Test (VST, Nation & Beglar, 2007)	Vocabulary knowledge	15 minutes	Written, receptive	Used in a pilot study, but not in the main study

Oxford Placement Test 1 – Grammar Test (OPT grammar, Allan, 2004)	Grammar knowledge	25 minutes	Written, receptive	Used in a pilot study
Test for Reception of Grammar version 2 (TROG-2, Bishop, 2003)		12 minutes	Aural, receptive	
Online grammaticality judgement task (OGJ, Felser et al., 2003a)	Aural sentence processing efficiency	8 minutes	Aural, receptive	Used in a pilot study
Auditory questionnaire (Felser et al., 2003a)	Aural sentence processing efficiency	12 minutes	Aural, receptive	Used in a pilot study, but not in the main study
Working memory digit-span tasks (Wechsler, 1997)	Working memory	5 minutes	Aural, productive	Used in a pilot study
Raven's Standard Progressive Matrices (Raven's SPM, Raven et al., 1992)	Non-verbal reasoning abilities	45 minutes	Written, receptive	Used in a pilot study
A personal background questionnaire (PBQ, Dunn& Tree, 2009; Grosjean, 2015)	Contact with the target language and culture	5 minutes	Written	Used in a pilot study

3.4 Procedure

Data collection in China

Data collection among participants of Group 1 began at the end of March 2017 and ended in mid-June 2017. Participants took four hours and twenty minutes over four sessions to complete all tasks.

In the first data collection session, participants completed most tasks in a language lab equipped with computers and headphones. Before the start of the data collection, all the equipment in the computer lab had been checked to verify that it was working properly. The equipment included a computer with a teaching system for teachers to control the computers on students' desks, 30 computers running Windows 2007 Microsoft system which could be used with or without control of the teaching system from the teacher's desk, and a large screen at the front of the room. Each computer came with a monitor screen and was equipped with headphones and a keyboard. Since one of the tasks had to be run using the software Psychopy-2, this had been installed on each computer prior to the data collection starting. The lab could accommodate 30 participants at most, therefore, participants were divided into twelve groups according to their majors and the number of participants in each group ranged from 7 to 22. The twelve groups completed the same tasks in twelve sessions, but six groups completed the tasks in one order and the other six groups in the counterbalanced order, so as to reduce the impact of the order of the tasks on the results (see Table 3.3). There was a three to five minute break after each task in the first session. The administration of all tasks took place under controlled conditions (i.e., under supervision by me and student helpers). Answers were given to participants' questions about how to fill in Table 4 in the questionnaire which was about the percentage of daily use of English language, Mandarin, and/or local dialects. In each session, an English-major student from the university assisted with the data collection. There were four student helpers in total. They were paid 50RMB each time for their help during the first sessions of data collection, which was sponsored by a university in China. I am aware that the payment for the student helpers was more than the value of the present given to the participants, however these students gave more time to the project than was required of the participants. The student helpers had to arrive at the language lab 40 minutes earlier than the participants in order to help re-check that the computers and headphones all worked properly, to move all presents to the language lab, and to put the correct number of test papers in order. They left the language lab 30 minutes after each session ended. They also helped to copy each participant's performance record for the Online Grammaticality Judgment (OGJ) task to a U-drive and tidied the language lab at the end of each session. The tasks for the first session lasted three hours and fifteen minutes, including the breaks.

Table 3.3 The order of tasks in the first session for six groups of participants in Group 1

The order of tasks	Instruments	Variables to measure	Time duration	Form
1	OPT listening	Phonological knowledge	10 minutes	Aural, receptive
2	WRS	Phonological knowledge	17 minutes	Aural, productive
3	Raven's SPM	Non-verbal reasoning abilities	45 minutes	Written, receptive
4	LVLT	Vocabulary knowledge	35 minutes	Aural, receptive
5	OPT grammar	Grammar knowledge	25 minutes	Written, receptive
6	OGJ	Aural sentence processing efficiency	8 minutes	Aural receptive
7	TROG-2	Grammar knowledge	12 minutes	Aural receptive

8	WST	Phonological knowledge	13 minutes	Aural, receptive
9	PBQ	Contact with the target language and culture	5 minutes	Written

The order of tasks was counterbalanced across groups to avoid order effects. Table 3.3 provides an overview of the tasks participants in Group 1 completed. If one group started from task 1 and ended with task 8, then the other group started from task 8 and ended with task 1. The tasks were arranged in this order out of the following considerations:

1) the tasks which were considered easier for Chinese learners of English were put at the start of the session in both orders. These included the OPT listening and the WST;

2) the tasks which involved listening and those which did not, such as the WRS and the Raven's SPM, were alternated. However, since there were fewer tasks which did not involve listening than tasks which did, it could not be avoided that two tasks which involved listening were arranged continuously;

3) the tasks which were assumed would take participants a long time were not arranged in a continuous order;

4) the tasks which were assumed to measure participants' L2 LC (e.g., the PET listening and the CET4 listening) were taken after all other tasks were completed. The PET listening was not administered in the first session to take into consideration the relatively long duration of the first session;

5) the paper-based questionnaire was always administered as the last task of the first section in order to avoid priming effects (Steele & Ambady, 2006);

6) according to the participants' performance in the first session, the time duration of the OPT grammar task was changed to 25 minutes, two minutes less than in the pilot study, and the Raven's SPM was changed to 45 minutes, five minutes more than in the pilot study.

Before the start of each task, participants were given oral instructions in Chinese. Most tasks were paper-based, except for the OGJ. For the paper-based tasks, a student helper handed out the test papers to each participant before a task started. The test papers were collected when each participant finished a task. For the OGJ, participants were instructed to operate the keyboard properly and to adjust the volume of the sound from their desktop. Participants were told that for this task they were going to hear 52 sentences (four practice sentences and 48 experimental sentences). Participants had to decide whether the sentences they heard were grammatically correct or not. They were told that they would first read instructions for the practice sentences at the beginning of the task on the screen. Then they should press the spacebar of the keyboard to start listening to the first practice sentence. After the recording of a sentence stopped, participants were asked whether or not they thought the sentence was grammatical. If they thought the sentence they had heard was grammatically correct, they had to press the left arrow on the keyboard as quickly as possible; if they thought the sentence was grammatically incorrect, they had to press the right arrow on the keyboard as quickly as possible. They could see that the word *grammatical* would always appear on the left-hand side of the screen, and the word *ungrammatical* would always appear on the right-hand side. The RTs were measured from the end of the recording of the sentence to the moment they pressed the left arrow or the right arrow. After they had pressed one arrow for their decision, they heard the next sentence in the same section. Participants were told that they had to answer as quickly as possible and that their RTs to each item and the accuracy of their answers would be recorded automatically by the software. They knew that there were breaks after the practice section and after each of the 16 experimental sentences. They could start the next section after a short break by pressing the spacebar. If participants listened to the practice sentences and found no problem with the operation and the equipment, they continued to complete the task. If any problem occurred in the process of task completion involving the equipment, a participant moved to another seat without disturbing other participants. The data collection for the first session went well.

In the second session, participants completed the Working Memory tasks (WM) individually. Data collection for this session started one or two weeks after participants had completed the tasks in the first session. Four English major students who helped with the data collection had received training on how to lead a participant through the completion of the WM tasks which they did by explaining the requirements, the speech speed, tones, and pause duration. In the process of conducting the tasks, the student helpers read the digits at a rate of one per second and participants repeated the digits they had heard in forward or backward order. The student helpers made a note on the answer sheet by scoring participants' performance on each trial; one point was scored for each correct trial. The tasks were continued if one trial was passed, but stopped if two different trials at the same level were failed. The tasks lasted five minutes on average for each participant. Before the start of the WM tasks, the student helpers agreed with participants on a meeting time and location for them to complete the tasks based on participants' majors. Each student helper conducted data collection of the tasks among participants from the same major. The WM tasks were conducted in eight sessions.

In the third session of the data collection, participants completed the PET listening. In the present study, participants' scores on the PET listening was one of the two measured variables of their L2 LC. The task was completed under the supervision of the participants' teachers of English and was conducted in the university's language labs over seven sessions, based on participants' majors and available times. Each lab was equipped with PCs and headphones and could accommodate 50 users. Before starting the task, the teachers in charge explained to participants that they would complete four parts in a listening task. For Part 1, participants had to choose a correct answer from three options of pictures after they heard a dialogue; for Part 2, they had to choose a correct answer from three options of statements after they heard a passage; for Part 3, they had to fill in six blanks in statements based on what they heard from a short passage; for Part 4, they had to decide whether the statements on the test paper were true or false based on what they heard. The teachers handed out the test papers to each participant before playing the CD of the recordings. Participants were instructed to trial the headphones and adjust the volume. The teachers played the CD only once. After the recordings stopped, the test papers were collected. The task took 30 minutes, five minutes less than the time originally set for this task because participants did not have to transfer their answers to an answer sheet.

In the fourth session of the data collection, participants took the College English Band 4 Test (CET4). In the study, participants' scores on the CET4 listening task was the other measured variable of their L2 LC (in addition to the PET listening task). In 2017, the first CET4 test was held on 17th June. The test is administered only by the universities and colleges authorised to do so, of which the participants' university is one. The issues relating to the CET4 test were under the administration of the university, including: the starting and ending of the test; playing recordings; arranging of test-takers' seats; supervising staff; and handing out and collecting test papers. In the present study, participants were recruited in March on the condition that they had registered for the CET4 test on 17th June 2017. The scores for the CET4 listening task received in September, 2017, from the Department of Teaching Affairs of the university revealed that all participants took the test in June.

Data collection in the UK

Data in the UK were collected from 40 participants between 10th September and 18th October, 2017. It took participants four hours and 20 minutes to complete all tasks (including breaks) in two sessions, one of two hours and 40 minutes and the other of one hour and 40 minutes. The intervals between the two sessions for each participant varied from two to seven days. The tasks were counterbalanced to avoid order effects. Table 3.4 shows one of the two possible orders for the tasks. If one participant started from task 1 in the Table and ended with task 7 in session one, then the next participant started from task 7 and ended with task 1 in this session. If a participant took the order task 1 to task 7 in session one, then they took the order task 1 to task 4 in session two, the next participant took the order from task 4 to task 1 in session two. The tasks were arranged in this order due to the following considerations:

1) the PET listening task and the CET4 listening task were arranged in two sessions as both of the tasks were measurements for L2 LC;

2) the tasks which involved listening and the tasks which did not involve listening were alternated. However, since there were fewer tasks which did not involve listening than tasks which did, it could not be avoided that two of these tasks were arranged continuously;

3) the tasks which were assumed to take participants a long time were not arranged in a continuous order;

4) the receptive tasks and the productive tasks were alternated. However, since there were fewer productive tasks than receptive tasks, it could not be avoided that two receptive tasks were arranged continuously;

5) the paper-based questionnaire was always administered before the end of the data collection.

In each session, one or two participants took the tasks listed in session one or session two in Table 3.4. When the participants were recruited for the study, they indicated their availability for taking the tasks. Therefore, according to the information participants provided, suitable teaching or research rooms with a PC were booked in advance from the Booking Centre of the university where the data were collected. The date and the location were confirmed with each participant before a session started. At the beginning of the first session, participants were informed about the purpose of the study and signed consent forms. Before each task, participants received oral task instructions. Most tasks were paper-based, with the exception of the OGJ task. The test papers were handed out to each participant before starting each task and were collected after each participant completed a task. As the OGJ task had been set up to run with the Psychopy-2 software, participants had to complete this task on my laptop because I was not allowed to load the software on the university computer. Participants were given the same instructions as Group 1 on the formation of the task, the operation of the keyboard, and the speed requirement. Headphones were provided in each session. Participants' scores and RTs for each item were recorded on my laptop.

Table 3.4 The order of tasks in the two sessions for the participants of Group 2 (N = 40)

Session	The order of tasks	Instruments	Variables to measure	Time duration	Form
One	1	The OPT listening	Phonological knowledge	10 minutes	Aural, receptive
	2	The PET listening	Listening proficiency	30 minutes	Aural, receptive
	3	The Raven's SPM	Non-verbal reasoning abilities	45 minutes	Written, receptive
	4	The WST	Phonological knowledge	13 minutes	Aural, receptive
	5	The WM tasks	WM	5 minutes	Aural, productive
	6	The OGJ task	Aural sentence processing efficiency	8 minutes	Aural receptive
	7	The LVLT	Vocabulary knowledge	35 minutes	Aural, receptive
Two	1	The TROG-2	Grammar knowledge	12 minutes	Aural receptive
	2	The WRS	Phonological knowledge	17 minutes	Aural, productive

3	The OPT grammar	Grammar knowledge	25 minutes	Written, receptive
4	The CET4 listening	Listening proficiency	30 minutes	Aural receptive
5	PBQ	Contact with the target language and culture	5 minutes	Written

The questionnaire was filled in after all other tasks were completed. Answers were given to participants' questions on how to fill in Table 4 in the questionnaire about the percentage of daily use of English, Mandarin, and local dialects.

Scoring

All participants' performance on the tasks was scored. For the paper-based receptive tasks, namely the WST, the OPT listening task, the LVLT, the OPT grammar task, the TROG-2, the Raven's SPM and the PET listening task, each test paper was scored based on the standard keys provided. One point was given for each correct option chosen, and 0 was given for each item where the wrong option was chosen.

For the CET4 listening, the test papers of participants in Group 1 were scored under the administration of the CET4 Testing Committee and the scores were reported officially. The standards for scoring the CET4 test papers have never been publicised by the CET4 Testing Committee. In order to have access to the exact weighting of the different parts of the listening task, I contacted one of the directors of the CET4/6 Testing Committee. I promised that if I had access to the scoring system, it would be used only for my PhD study, and for academic purposes, and would not be disclosed to any profitable organisation. However, the request was refused for the reason that the scoring system of the CET4/6 is highly confidential at national level, and therefore cannot be disclosed to anyone or any organisation for any purpose whatsoever. The test papers from the CET4 listening completed by the participants in Group 2 were scored based on the standard keys provided. One point was given to each item with the correct answer, and 0 to each item with a wrong answer. The total scores were the sum of the number of items with correct answers with a maximum score of 25. The maximum was 249 for the official version, but I could not compute a new total score for Group 2 because I did not know the weighting of different components.

The OGJ was computer-based and participants' answers were recorded automatically on computers. Participants' RTs and accuracy were recorded by the software Psychopy-2; one point was given to each item with the correct answer.

Differing from the receptive tasks in the present study, the WRS was a productive task for which participants were required to fill in the blank in each item when they heard the word on the recording. The designers of the test suggested scoring these tests very carefully in order to minimise the potential threat to its validity which could be caused by the requirement to provide words in the written form. The test papers were scored on the basis of the structured rubric provided by the designers (Matthews, O'Toole, & Chen, 2016):

1) 1 point, i.e., full credit for each item, was given to a response that was written in the correct orthographic form;

2) 1 point was given to a response which included minor spelling errors which did not prevent recognition of the target word;

3) 1 point was given to a response which showed that the root word had been recognised except for the incorrect form of the verb, such as ending with "ed" or "ing", or incorrect use of pluralisation of "s" or "es";

4) 0 was given to a response which revealed that a single or multiple incorrect vowel or consonant had basically changed the phonological form of the word when it was pronounced;

5) 0 was given if no answer was provided.

Each item was worth 1 point and the highest possible score was 90.

The WM tasks were also productive tasks. Student helpers read out each digit trial and participants repeated what they heard. One point was given to each correct production, either forward digit production in Chinese or in English, or backward digit production in either language; 0 was given for incorrect production. The total scores for each WM task were the sum of correct answers. The total possible score for the forward task in each language was 16, and for the backward task in each language it was 14.

In the questionnaire, responses to items from 1 to 18 (except items 11, 13, & 17) were analysed and results are reported in Section 3.2. Items 11, 13, 17, and 19 were scored by referring to Dunn and Tree (2009). For item 11, *At what age did you first learn English______?*, 5 points were given to answers between 0 and 5 years old, 3 points were given to answers between 6 and 9 years old, 1 point was given to answers between 10 and 15 years old, and 0 was given to answers 16 years old and over. For item 13, *At what age did you feel comfortable speaking English and Mandarin in every day conversation? (If you still do not feel comfortable, please write "not yet.") Mandarin________ English _______* , only answers to *English________* were scored based on the following rule: 5 points were given to the responses between 0 to 5 years old, 3 points were given to responses between 6 and 9 years old, 1 point was given to responses between 10 and 15 years old, and 0 was given to responses of 16 and over and "not yet". For item 17, *How many years of schooling (primary school through university) did you have in: Mandarin Chinese ________ English ________*, only answers concerning English use were scored, based on the following rule: 1 point was given to responses between 1 and 6 years, and 2 points were given to responses 7 and more years. For item 19, *What country/region do you currently live in?* ________, 4 points were given to those living in the UK.

In the three sections of the self-rating of language knowledge task, participants rated their own language knowledge on a ten-point scale ranging from 0 (minimum) to 10 (maximum) for each of the four language skills, *speaking*, *writing*, *listening,* and *reading*. The maximum score for all four sections was 40.

For the section about daily English, Chinese, or dialect use, each participant's average percentage of daily use for each language was calculated based on Grosjean (2015). Each participant's scores for all items for one language use were summed to get a total score, then the total score was divided by the total number of items. The answers to the last item *other topic, namely…*were not calculated because most participants did not provide an answer.

3.5 Data analysis

Treatment of outliers

Prior to data analysis, outliers in each task were checked. According to Kline (2016), scores that are very different from the rest are outliers. Data points that are above or below two standard deviations (SDs) are usually regarded as outliers (Marinis, 2010). Two SDs, rather than 2.5 or 3 SDs, is adopted in the current study in order to keep the remaining scores representative and reduce the impact of unrepresentative scores. Field (2017) suggests winsorizing data by replacing outliers with a score that is not an outlier, but is just next to the smallest or the largest outlier. Therefore, in the present study, data that had absolute z-score values larger than 2 SDs from the mean in each variable were winsorized by replacement with data that were just below the absolute z-score values of 2 SDs from the mean in that variable. For example, in the OPT listening task, after the total scores were transformed into z-scores, two z-score values were larger than 2 and seven z-scores were smaller than -2. Since the original scores which had a z-score value just below 2 and a z-score value just above -2 were 90 and 63 respectively, the original scores which had z-score values greater than 2 were changed to 90, and the original scores which had z-score values smaller than -2 were changed to 63.

The numbers of outliers for the total scores in the paper-based tasks are given in Appendix 9, with the exception of the CET4 listening and the questionnaires. Participants' performance on the CET4 listening was scored within different scales between Group 1 and

Group 2 because I did not have access to the weighting of different questions in the Chinese scoring procedure. For Group 1, the maximum score was 249, and for Group 2 it was 25. In order to put the scores of the two groups together as one variable for later analysis, the scores of the two groups were transformed into z-scores separately first, and then the z-scores of the two groups were put together. The results showed nine outliers in the scores for this task; these were changed into the z-score values that were just smaller than 2 or just greater than -2. The outliers in each section of LVLT and Raven's SPM were checked. The numbers of outliers in each section are given in Appendices 10 and 11.

For the SPS task, only the data for the 24 experimental sentences were kept to calculate the RTs, thus the data for the 24 filler sentences were not used to calculate the RTs. According to Wall (2012), RTs from error trials should not be used in analyses because of an additional component process which might operate on error trials, e.g., any operation that produced the error. Therefore, the data for the experimental items with incorrect answers were deleted and the data for the experimental items with correct answers were kept for analysis of the RTs for this study. The data for each variable were transformed into z-scores for checking outliers. Those with absolute z-score values larger than 2 SDs from the distance of the mean in each variable were replaced by the original scores which had z-scores just below 2, or just greater than -2. For example, for item 25, after the original scores were saved as standardised values, five z-score values appeared to be larger than 2, and the nearest z-score value to 2 was 1.96185. The original score in this variable for the z-score value of 1.96185 was 4833. Therefore, the original scores of the five participants with standardised values larger than 2 were replaced by the value 4833. The outliers in the other 23 variables in the task were changed the same way. The numbers of outliers in each experimental item in the SPS task are shown in Appendix 12.

After the outliers in each of the 24 experimental items had been replaced, each participant's total RTs for the experimental items with correct answers were calculated, and then the average RTs were calculated. The results showed that there were ten outliers in the scores of the total RTs and six outliers in the scores of the average RTs. The outliers in the total RTs and the average RTs were changed accordingly.

Preliminary descriptive analyses

Before answering the research questions a number of preliminary descriptive analyses were carried out. First, the reliability of each task was analysed. Among the tasks, the reliability of the CET4 listening was analysed based on the total scores from the participants in Group 2. This was done because the scores for individual items are not released to students as this is a high stakes test used throughout China; second, the scores for the PET listening and the CET4 listening were analysed in order to investigate whether the L2 LC proficiency of Group 1 and Group 2 was significantly different; third, the means, SDs, the minimum and maximum values for each task for all participants, irrespective of group membership, were established; fourth, for each group separately, the means and SDs for each task and the partial eta squares were computed; fifth, the correlations for all observed variables for all participants and per group were carried out.

Checking assumptions for testing hypothesized structural equation models of LC

The assumptions for testing against the hypothesized structural equation models of LC were checked; results are reported in Chapter 5.

The maximum likelihood (ML) estimation method was used in the structural equation modelling in the study. According to Kline (2016), ML assumes multivariate normality for continuous variables, which means that variables should meet the following standards:

1) all the individual variables are distributed normally;

2) any pair of variables are bivariate normally distributed, which means that “each variable is normally distributed for each value of every other variable” (Kline, 2016, p.74);

3) The bivariate scatterplots show linear relationships with homoscedastic residuals.

The violation of distribution normality of an individual variable can be detected through analysis of the indicators of skewness and kurtosis of distribution (Byrne, 2016). The violation of multivariate normality could be detected through analysis of the indicator of Mardia’s coefficient (Byrne, 2016). The test results of distribution normality of individual variables and multivariate were revealed by running data analysis in Amos. The results were satisfactory.

Collinearity among the measurement variables was tested in order to check whether the measurement variables in the study were highly correlated. According to Kline (2016), if two or more variables actually measure the same thing, it can happen that the variables have extreme collinearity. Extreme multivariate collinearity should be avoided in a regression analysis.

In order to ensure statistical precision and reasonable power for significance tests in SEM, Kline (2016) and Byrne (2016) suggest that the minimum sample size for SEM should be 10 times as many as the number of estimated parameters. Positive definiteness of the data matrix was checked because nonpositive definiteness would lead to failure of analyses to a data matrix. A positive definite data matrix is required for most estimation methods (Kline, 2016). It should have three properties: having an inverse, all eigenvalues being positive, and having no out-of-bounds correlations or covariances (Kline, 2016).

SEM is different from traditional multivariate procedures because it takes a confirmatory approach to the data analysis, rather than an exploratory one. CFA is used when a link between a latent construct (i.e., a factor) and the observed variables which measure this construct, assumed on the basis of theoretical or empirical research and the hypothesized structure, is tested (Byrne, 2016). Traditional multivariate procedures are only based on observed measurements (Byrne, 2016). The same variables that were included in Andringa et al. (2012) were also analysed in the current project, namely: LC (LC); Linguistic Knowledge (LK); Phonological Knowledge (PK); Vocabulary Knowledge (VK); Grammar Knowledge (GK); Sentence Processing Speed (SPS); Working Memory (WM); and Reasoning Ability (RA). English Use (EU) was added in the current study. These variables were the latent constructs/factors which were assumed to be represented by the measured variables. Only the latent construct of EU was added to the module to capture individual differences in learners' backgrounds. The hypothesized LC models are five-factor structures comprising LK, SPS, WM, RA, and EU. LC was measured with two different listening tasks and was the dependent variable in each model. The construct of LK is assumed to be represented by the three latent constructs of PK, VK, and GK, as in Andringa et al. (2012). Originally WM and RA were grouped into the factor of cognitive ability, but they were subdivided in Andringa et al. (2012) because the researchers found the two subfactors represented different constructs. First-order CFA and second-order CFA were conducted in this study. The purpose of carrying out second-order CFA was to confirm that a main construct loads onto a number of underlying sub-constructs or components (Awang, 2012). In this study, PK, VK, GK, SPS, WM, RA, and EU were assumed to be first-order constructs, and LK was assumed to be a second-order construct which comprises the sub-constructs PK, VK, and GK. Before the two hypothesized SEM models for the entire group were tested, each measurement model was checked in order to confirm that observed variables used to measure each construct were valid for that purpose. Observed variables which had lower factor loadings onto a construct were cut off from a construct. After each measurement model was checked, they were put together to form structural equation models of LC. Goodness-of-fit of each measurement model, and the two SEM models of LC, were tested separately by a χ^2- test in order to evaluate whether the measurement models and the hypothesized SEM models satisfactorily fit the data. Results of SEM analyses provide a χ^2 statistic with a corresponding *p*-value and degrees of freedom (*df*) (Schoonen, 2015). Although in traditional null hypothesis testing, null hypotheses, e.g., $p < .05$, are usually rejected, in SEM analyses, most of the time, researchers do not want to reject a model just because the *p*-value is significant (Schoonen, 2015). χ^2 in SEM analyses is sensitive to both sample size and the number of parameters to be estimated, therefore, it is preferred that the χ^2 statistic is used as a more descriptive indicator of model fit than a statistical significance test (Schoonen, 2015). In addition to a χ^2 value with degrees of freedom and *p* value, Kline (2016) and Schoonen (2015) suggest reporting values of SRMR, RMSEA, and CFI as descriptive fit indices. SRMR stands for standardised root square residual, RMSEA stands for the root mean square error of approximation; and CFI stands for the comparative fit index (Schoonen, 2015). It is preferable that χ^2/df ratio should be lower than 2, SMRM lower than .08, RMSEA lower than .06, and CFI higher than .95 (Hu & Bentler, 1999). These values are reported as indicators of goodness-of-fit for each measurement model and two SEM models of LC in Chapter 5.

3.6 Ethical issues

Prior to the data collection for the pilot study and the main study, ethics approval was obtained from the Research Ethics Committee of the Institute of Education, the University of Reading, on 9th September, 2016. Since a part of the main study was carried out in a Chinese university the Vice Chancellor of that university and some English teaching staff were informed of the purpose of the research and its process. The English teaching staff offered to help to organise the participants. Signed consent forms from the vice chancellor and the English teaching staff were received. Before the start of the first task, each participant was given an information sheet detailing the purpose of the study, the tasks they would be asked to do, and the time these would take. Participants signed consent forms were collected before the tasks were started.

Chapter 4 Pilot studies

In order to ascertain whether the selected measurements were suitable for use in the main study, three pilot studies were conducted between April 2016 and March 2017. This chapter presents how each pilot study was conducted in turn, and then draws a conclusion from all three pilot studies.

4.1 Pilot study 1

Pilot study 1 was conducted from April 2016 to June 2016. It focused on the following tasks: online grammaticality judgement; online auditory questionnaire; word segmentation; working memory; aural and written vocabulary; and a personal background questionnaire. The original plan was to use an online auditory questionnaire task in the main study as one of the two online tasks to measure participants' sentence processing speed, and a written vocabulary task to measure participants' written vocabulary size. However, consideration of time limitations meant the two tasks were not used in the main study. Nevertheless, as the two tasks were piloted, the process and the data analyses concerning them are reported here as part of the study.

The participants in this first pilot study were ten English speakers in the UK, nine Chinese learners in the UK, and 56 Chinese learners at a university in China. The ages of those Chinese learners of English ranged between 18 – 20. A group of English speakers was needed in the pilot study to check the difficulty of the following three tasks: the online grammaticality judgement task; the online auditory questionnaire task; and the word segmentation task. If the tasks were difficult for English speakers, they would be unsuitable for non-native speakers. Two groups of Chinese learners were involved: the nine Chinese learners of English in the UK completed the tasks on a computer under my instruction; the 56 participants in China completed the tasks in a language lab, or in an ordinary classroom as they did not require access to technology. This was easier to realise for the English teaching staff who assisted with the data collection.

The three tasks were completed by each of the ten English speakers individually in one session, in a quiet room. The tasks were administered in a fixed order: the grammaticality judgement task, the auditory questionnaire task, and the word segmentation task. Before the start of each task, participants were given time to read the instructions and were shown how to carry these out on the computer. Participants wore headphones for all three tasks, which took 35 minutes to complete.

The grammaticality judgement task and the auditory questionnaire task were online tasks which were set up and run using the Psychopy software (version 1.83), and participants' responses were recorded automatically on the computer. The utterances presented in the three tasks were read at a normal pace (145-150 words per minute), in a natural way by a female English speaker who spoke Standard British English, in a quiet room using an MP3 digital voice IC recorder (Sony ICDPX333.CE7 4GB PX Series). In order to record reaction times accurately, the sound editing software Audacity was used to cut off the extra blank at the end of each sentence recording. The recording of participants' reaction time started immediately at the moment the sentence recording ended and the answer options appeared on the screen.

After collecting the data among English speakers, participants' error rates for each task were analysed and it was found that the mean error rate was lower than 40%. This was deemed acceptable because if a participant's error rate is higher than 40%, it can be assumed that either the participant did not understand the task, or the difficulty of the task was not suitable to the participant's proficiency (Andringa et al., 2012). Therefore, it was decided that the tasks were not too difficult for native speakers and could be tried with Chinese participants in the UK.

The Chinese participants in the UK completed the following five tasks in a fixed order: the grammaticality judgement task; the auditory questionnaire task; the word segmentation task; the working memory task; and the personal background questionnaire. All tasks were completed individually in one session, in a quiet room. Before the start of each task participants were given time to read the instructions and were shown how to carry them out on the computer. Participants wore headphones for the first three tasks and took 45 minutes to complete them all.

The Chinese participants in China completed three tasks: two vocabulary size measurement tasks and a personal background questionnaire. One of my colleagues helped to organize the data collection. The aural vocabulary size test was completed first in a language lab with my colleague playing the recordings. It took participants 35 minutes to complete the test. The written vocabulary size test was completed ten days later in a classroom and was again administered by my colleague, which was followed by the personal background questionnaire. It took participants 15 minutes to complete this test and five minutes to complete the questionnaire.

In the grammaticality judgement task, participants wore headphones and read instructions on how to carry out the task on the computer. After hearing a sentence, they decided as quickly as possible whether the sentence that they had just heard was grammatical or ungrammatical by pressing the left arrow key (*grammatical*) or right arrow key (*ungrammatical*) on the keyboard. For example, hearing *The reporter phoned the boss of the secretary who was reading a book* should evoke a grammatical-response, whereas hearing *The woman blamed the hairdressers with the apprentices who was smiling all the time* should evoke an ungrammatical-response. Breaks were included after completion of 16 and 32 sentences. After the first 16 sentences participants chose when to start listening to the next 16. The task took approximately 8 minutes.

For the auditory questionnaire task, participants wore headphones and read instructions on how to carry out the task on the computer. After hearing a sentence and a question based on the sentence they had just heard, they chose one answer that they preferred from two answer options, as quickly as possible, by pressing the left or the right arrow keys. For example, after they heard the sentence, *The nurse trusted the teacher with the doctor who was preparing to go home* and the question, *Who was preparing to go home?* participants had to choose from the two answer options on the screen *A. the teacher* or *B. the doctor*. Both preference response and reaction time were recorded automatically. The task took approximately 12 minutes.

The word segmentation task measured participants' word segmentation ability in the stream of speech as well as how confident they were with their answers. Participants wore headphones and read instructions on the test paper about how to choose an answer from two answer options and how to choose one number from the numbers 1-7 (where 1 represented "not sure", and 7 "very sure") to indicate how confident they were in their answer. Participants heard short phrases comprising two to three words in a carrier sentence *Say____ again* and chose one answer from two options. For example, after hearing the sentence *Say tops pry again*, participants chose between *A. tops pry* and *B. top spry*, and chose from 1-7 to indicate their confidence in their answer. The task took approximately 13 minutes.

Four digit span tasks (forward and backward in both a Chinese version and an English version) were used to measure participants' working memory capacities. Accordingly, I read each digit span trial in Chinese or in English, forward or backward, and the participants reproduced it in the same way. Each task ended when the participants failed to reproduce both trials of a particular length. Participants' highest number of one or two correct items in each of the four tasks stood for their auditory digit span in the corresponding working memory task. For example, if a participant succeeded in reproducing one or two items in the serial of eight digits in the Chinese forward task after hearing me read “52193748” and/or “29614857”, but failed to reproduce two items in the serials of nine digits in this task, their Chinese forward auditory digit span was eight. The same applied to other tasks. This task took about five minutes.

Participants in the UK filled in the personal background questionnaire after they had completed the other tasks and participants in China filled in the questionnaire after they had completed the written vocabulary size test.

Preliminary analyses were made. In order to decide whether the tasks were suitable for English speakers, the first step was to check English speakers’ error rates for the filler sentences in the grammaticality judgment task, error rates for the filler sentences in the auditory questionnaire task, and error rates for the word segmentation task.

The analyses showed that English speakers’ mean error rate for the filler sentences in the grammaticality judgement task was 25%, with scores ranging from 37.5% to 4.17%. The mean error rate for the filler sentences in the auditory questionnaire task was 6%, with scores ranging from 20% to 0%. The mean error rate for the word segmentation task was 13.7%, with scores ranging from 23.8% to 8.33%. Since no participants’ error rate was higher than 40%, responses from all English speakers were retained.

Next, in order to check the difficulty of the target items in the grammaticality judgement task and the word segmentation task, the accuracy of target items for the English-speakers in the two tasks were analysed. Participants’ accuracy in choosing the target items in the auditory questionnaire task could not be computed because the experimental sentences in this task were ambiguous and both answer options were correct. The data analyses (see Table 4.1) showed that the mean accuracy for the target items in the grammaticality task was 83.8%, and the mean accuracy for the target items in the word segmentation task was 86.3%. As the accuracy for both tasks was higher than 80% it was assumed that the three auditory tasks in the pilot study were not too difficult for English speakers. Then the tasks were conducted with the group of Chinese learners of English in the UK.

In order to ascertain whether the nine Chinese participants in the UK understood the tasks, or that the participants’ proficiency was not too low, firstly these participants’ error rates for the filler sentences in the two online tasks were analysed. Then their error rates for the word segmentation task and the accuracy of target items in the grammaticality task were analysed.

As can be seen in Table 4.1 two of the nine participants' error rates for the filler sentences in the grammaticality judgement task were higher than 40% (both were 58.3%). In addition, one of the two participants' error rates for the filler sentences in the auditory questionnaire task was higher than 40% (it was 60%) and all participants' error rates in the word segmentation task were lower than 40%. Therefore, the results from two participants were omitted in the data analyses and data from only seven participants were included in the pilot study. The results showed that the accuracy for the target items of one of the seven participants in the grammaticality task was lower than the chance level (41.7%), therefore, this participant was excluded and their responses to all the tasks were deleted. Finally, only six participants' responses were retained for the data analysis.

In the process of data analyses, one English participant's reaction time on one item in the grammaticality judgement task was taken as a missing value because the participant's answer was postponed due to a processing problem caused by the computer.

Table 4.1 Results of the first pilot study

Instruments	Computing Items	English Speakers (*N* = 10)	Chinese participants
Written vocabulary test (Cronbach's α = .71)	Vocabulary size	Not applicable	M = 3000 (60%) (*N* = 56)
Aural vocabulary test (Cronbach's α = .89)	Vocabulary size	Not applicable	M = 97 (64.7%) (*N* = 56)
Grammaticality judgment test (accuracy of RCs) (Cronbach's α = .69)	Accuracy of target items	83.8%	63.3% (*N* = 6)
	Mean RTs of target items answered correctly	1467 (ms)	2359 (ms)
Auditory questionnaire (processing of RCs) (Cronbach's α = .69)	Preferences of NP2	of = 42%; with = 60%	of = 28%; with = 48% (*N* = 6)
	Mean RTs of target items answered correctly	of = 1990 (ms); with = 2198 (ms)	of = 2432 (ms); with = 4452 (ms)
Word segmentation test	Accuracy of target items answered	86.3%	72.9%

(Cronbach's α = .80)	correctly		(*N* = 6)
	Mean confidence rates	5.7	5.0
Working memory test	Four digit span tests	Not applicable	CF = 7.8; CB = 6.2 EF = 5.2; EB = 4.6 (*N* = 6)

For the segmentation test, the mean accuracy and the mean confidence ratings were analysed. The mean accuracy of the target items was 86.3% for English speakers and 72.9% for Chinese participants in the UK. The mean confidence ratings for all items were 5.7 for English speakers and 5.0 for Chinese participants in the UK. The internal reliability of the word segmentation test (Cronbach alpha) was .80.

The scores on the two vocabulary size tests for the 56 Chinese participants in China are shown in Table 4.1. The internal reliability (Cronbach's alpha) of the written vocabulary size test (the VST, Nation & Beglar, 2007) was .71, the mean score of the VST was 3000 (*SD* = 508.57), and the mean accuracy rate was 60%. The internal reliability of the aural vocabulary size test was .89, the mean score was 97 (*SD* = 17.57), and the mean accuracy score was 64.7%.

For the online grammar tasks, responses from the English speakers and the Chinese participants in the UK to the grammaticality judgement task and the auditory questionnaire task were analysed (see Table 4.1). The scores for the two groups in the former task showed that the mean accuracy of target items was 83.8% for the English speakers, and 63.3% for the Chinese participants. The internal reliability (Cronbach's alpha) for the grammaticality judgement task was .69. In the auditory questionnaire task, participants were required to choose which potential antecedent (NP1 or NP2) they would prefer for *of* conditions and *with* conditions. The English speakers showed 42% preference for *of* conditions and 60% for *with* conditions; for Chinese participants in the UK, 28% preferred *of* conditions and 48% preferred *with* conditions. A comparison of results with those of Felser et al. (2003a) is made below.

The RTs of the English speakers and the Chinese participants in the UK were analysed. For the grammaticality judgement task, the mean RTs for target items which were answered correctly was 1467 milliseconds for the former, and 2359 milliseconds for the latter. For the auditory questionnaire task, the mean RTs of target items were as follows: for English speakers, 1990 milliseconds (*of* conditions) and 2198 milliseconds (*with* conditions); for Chinese participants in the UK, 2432 milliseconds (*of* conditions) and 4452 milliseconds (*with* conditions).

In the working memory tests, Chinese participants in the UK took the four digit-span tasks (backward and forward, in the Chinese version and the English version). The results showed the means of participants' digit-spans: 7.8 for the Chinese forward test, 6.2 for the Chinese backward test, 5.2 for the English forward test, and 4.6 for the English backward test.

Participants' Chinese digit-span memory capacities turned out to be larger than their English digit-span memory capacities, and participants' forward digit-span memory capacities were larger than their backward digit-span memory capacities.

The two groups of Chinese participants filled in the personal background questionnaire task. Since some participants in China did not provide full information about all items, it was difficult to analyse the data collected from this group. Three of the six participants in the UK did not provide full information for items about their daily use of Chinese and L2. Therefore, only a part of the data from the participants in the UK is shown (see Table 4.2 and Table 4.3), and no comparisons between the two groups were made. As a result, in the main study the presence of all information requested was checked when the papers were collected.

Table 4.2 Results of the personal background questionnaire (N = 6)

Questions	Chinese participants in the UK
The age range	18 - 24
Years of English study	M = 13
Undergraduate or graduate	4 undergraduates, 2 graduates
The length of studying in the UK	M = 10.4 months
English proficiency level from IELTS scores	5 = B2, 1 = C1
Age of starting learning English	7 years
Which language is used at home, predominantly?	5 Mandarin, 1 Mandarin and English
Which language is used to calculate numbers?	5 Mandarin, 1 Cantonese

Table 4.3 Results of self-rating of language knowledge (*N* = 6)

Skills	Chinese language	English language
Speaking	8.4	5
Writing	7.6	4.2
Listening	8.2	4.4
Reading	8.2	5

The data analyses showed that in the pilot study for the word segmentation test, L2 learners' accuracy of their responses to target items (72.9%) was lower than the accuracy of the native speakers (86.3%). Although in Altenberg (2005) and Ito and Strange (2009) native speakers' performance on all stimulus groups is at or near ceiling, this was not the case in this pilot study. Ito and Strange (2009) found that the accuracy of their Japanese L2 learners' responses to the stimuli was 83.8%, higher than that of Chinese participants in general (72.9%), and of the French learners of English (74.6%) in Shoemaker (2014). Further analyses of the Chinese participants' performance in the three stimulus groups showed that Chinese participants' performance on double cue pairs (63% accuracy) was less accurate than their performance on glottal stop pairs (71% accuracy) and aspiration pairs (78% accuracy). The findings differ from Ito and Strange's (2009) Japanese L2 learners who showed much better performance on double cue pairs (94% accuracy); these learners' performance on double cue pairs was found to be better than their performance on glottal stop pairs (91% accuracy) and aspiration pairs (73% accuracy). In Altenberg (2005), Spanish learners of English showed relatively higher accuracy on both positive aspiration stimuli and positive glottal stop stimuli. Shoemaker (2014) and Ito and Strange (2009) indicated that French learners of English and Japanese L2 learners of English were more sensitive to the presence of glottal stops in English word boundaries than to aspiration stops. However, Chinese learners of English showed higher scores only on positive aspiration stimuli.

The items most frequently not recognised correctly by Chinese learners of English were *seem able* (100% error), *wife ill* (80% error), *loaf ate* (80% error), and *weep at* (100% error);

this differed from the most error-inducing items for native speakers, which were: *seen either* (80% error), *claim annual* (80% error), and from the items that Japanese L2 learners made errors with, such as *Lou skis* (56.7% error), *keep stalking* (56.7% error) and *tops crawled* (60% error).

One of the aims of the vocabulary tests in the pilot study was to ascertain whether the aural vocabulary test (the LVLT, McLean et al., 2015) was suitable for Chinese learners of English. As the Cronbach's Alpha results were satisfactory, there is some evidence from the pilot results that the LVLT is indeed a reliable aural vocabulary test. Although the internal reliability of the VST (Nation & Beglar, 2007) was not very high (Cronbach's alpha = .71), this might be due to the small number of participants. In the main study there was a larger number of participants, and thus reliability could be expected to improve. In addition, the pilot results showed that the 56 participants' mean vocabulary size was 3000 (60% accuracy), which is similar to the results from Y. Wang (2014). In Y. Wang (2014), first-year university students in China were found to have an average vocabulary size of 3000 word families. Since there was some overlap of items in the two vocabulary tasks, repetition of those items was analysed. Results indicated that 33 out of 50 items in the VST are covered in the LVLT, which means 66% of items in the written vocabulary task were repeated in the aural vocabulary task. Results also showed that among the 33 repeated items, 18 items were spread over different levels between the two tests. For example: *standard* in the first 1K level of the VST is grouped into LVLT *Part 2*; and *haunt* in the fifth 1K in the VST is grouped into LVLT *Part 4*. This result means that the standard for choosing words for word frequency levels might be different in the two vocabulary size tasks. No words in LVLT *Part 6* (Academic Word List) were found repeated in the VST.

The results from the online grammar tasks show that the internal reliability (Cronbach's alpha) of both the grammaticality judgement task and the auditory questionnaire task was .69. The reliability is not high because of the small number of participants in the pilot study. The problem is expected to be resolved when the two tests are carried out by a larger number of participants in the main study.

Results indicated that English speakers' mean accuracy score on target items in the grammaticality judgement task was 83.8%, which was very close to the adult English speakers' mean accuracy (88%) in Felser et al. (2003a). As could be expected, the mean accuracy score of Chinese participants was much lower at 63.3%.

The pilot test results for the auditory questionnaire task showed the mean percentages of NP2 responses by the English speakers and the Chinese participants to be: English speakers, *of* = 42%, *with* = 60%; Chinese, *of* = 28%, *with* = 48%. The English speakers' responses were very close to the findings in Felser et al.'s (2003a) study, which indicated that in adult English speakers' responses, *of* sentences NP2 preferences were 41% and *with* sentences 67%. One of the aims of the aural questionnaire task was to explore whether participants' RC attachment preferences are determined by a universal Recency Preference strategy. The pilot results indicated that Chinese participants do not use a universal Recency Preference strategy because less than half the participants preferred to choose NP2 as an answer for both *of* conditions and *with* conditions. Meanwhile, less than half of the English speakers preferred to choose NP2 as an answer for *of* conditions. Therefore, the results indicated that English speakers and Chinese learners are similar in their preference for NP1 in *of* conditions, but they differ in their preference for NP2 in *with* conditions.

Results show that for the grammaticality judgement task, English speakers' aural sentence processing speed (RT = 1467 ms) was much faster than the speed of Chinese participants (RT = 2359 ms), although the RT differences between the two groups were not statistically significant (*p* = .517).[1] For the auditory questionnaire task, in *of* conditions the RT differences between the two groups were not statistically significant (*p* = .06), while in *with* conditions the differences were statistically significant (*p* = .001). The RTs to *of* conditions (RT = 2432 ms) and *with* conditions (RT = 4452 ms) among Chinese participants were statistically different from each other (*p* = .02).

4.2 Pilot study 2

Pilot study 2 was organized in December, 2016. The focus of the piloted tasks was on the grammar section of the Oxford Placement Test (OPT, Allan, 2004) and the Raven's SPM. The purpose was to establish whether the grammar section of the OPT was suitable for Chinese learners of English in the current study, and how long it would take participants to complete this task and the Raven's SPM.

Eleven first-year university students (four males and seven females) participated in the pilot. They were chosen because they shared the same characteristics with the target participants in the main study, i.e., they were non-English majors ranging in age from 18 to

[1] Since the number of samples in the pilot study was smaller than 50, the Shapiro-Wilk test was used (Li, Zhang, & Shu, 2008).

22 who were preparing to take the CET4 in June, 2017 and had no background of living or studying in an English-speaking country. Before the start of this pilot, permission was obtained from the Ethics Committee of the IoE and participants were informed of the purposes of the study.

The pilot was conducted in a language lab in a Chinese university. The OPT grammar section was taken first, followed by the Raven's progressive test, with a five-minute break between the two tasks. The average time spent on the tasks was 27 minutes for the OPT grammar section and 40 minutes for the Raven's SPM.

For the grammar task, the highest score was 81 out of 100, and the lowest was 52. The mean score was 67.8. Two scores were under 50, three were between 60-69, and five were between 70-79. For the Raven's SPM, the highest score was 54 out of 60, and the lowest was 40. The mean score was 46.6.

It is not possible to say whether the scores were normally distributed because of the small number of participants, but there was clearly no ceiling effect, nor a floor effect. Therefore, the two tasks were considered to be of an appropriate level for the participants.

4.3 Pilot study 3

In order to ascertain whether the difficulty of the PET listening section was suitable for the target participants, a pilot study was conducted before the data collection for the main study. Thirteen university students participated in this pilot in March, 2017. The participants were chosen on the same basis as for pilot study 2. The pilot was conducted in a language lab in a Chinese university. Before the start of the pilot, participants were informed of the study purposes and ethics consent forms were signed.

At the start of the pilot, the participants were required to put on headphones. The PET listening test papers were handed out and the audio CD was played. The participants wrote down their answers on the test paper while they listened to the audio recordings. The test papers were collected when the audio CD ended. The task lasted 30 minutes.

As can be seen in Appendix 13, the mean score for the PET listening section was 12.85 out of 25, and the minimum and the maximum scores were eight and 22. Five scores were above the mean score. Results suggest that the PET listening section was not too difficult for the participants. The Cronbach's alpha of the PET listening section was .79 (see Appendix 14 for further details). Field (2017) provides general guidelines on acceptable levels of Cronbach's alpha and .79 is within the acceptable range. Therefore, I decided to adopt the

PET listening as one of the instruments to measure participants' LC proficiency in the main study.

4.4 Conclusion to the pilot studies

The main purpose of the pilot studies was to ascertain whether the tests/tasks that were self-designed or adapted from other researchers, and those that had not yet been used among Chinese L2 learners, were suitable for Chinese learners of English. There is some evidence from the pilot results that these tests/tasks are reliable. It is expected that the reliability of the written vocabulary test and the two online speed tasks might be improved when they are used in the main study with a larger number of participants. The accuracy scores of the two vocabulary tasks, the grammaticality judgement task, the word segmentation task, the OPT grammar section, the Raven's SPM, and the PET listening section indicated that these tasks were not too difficult for the Chinese participants. Therefore, the pilot tasks were used in the main study to help with validation, with the exception that the auditory questionnaire task and the written vocabulary task were omitted in the main study due to time limitations.

Chapter 5 Results

This chapter includes five sections. First, in Section 5.1, the descriptive results for all informants from both groups are presented, after which the results from each group are given separately. In Section 5.2, regression analyses are offered, first with the PET listening as the dependent variable, and then with the CET4 listening as the dependent variable. Prior to carrying out any further analyses on the structural equation models, assumptions for testing against measurement models and hypothesized structural equation models are checked in Section 5.3. Then the hypothesized structural equation models with both dependent variables are tested in Sections 5.4 and 5.5. In the final section of this chapter, answers to the three research questions of the study are concluded.

5.1 Analyses of reliability and means of each task and LC proficiency

In order to check whether each task consistently measured the construct that it was expected to measure, the reliability of each task was analyzed. As can be seen in Table 5.1, nine out of the ten tasks in the study were found to have a reliability coefficient higher than .70, and only one task had a lower reliability, i.e., .66. The results indicate that the internal reliability of the tasks was acceptable or high (Field, 2017). Therefore, it was assumed that the data from all the tasks were sufficiently reliable and could be used in the analyses.

Table 5.1 Reliability coefficients (Cronbach's alpha) of each task

Task	N of items	Cronbach's Alpha	Notes
PET listening	25	.77	
CET4 listening	25	.77	Number of participants = 40
WST	61	.66	23 items were deleted from the original task in order to improve the internal reliability of the task (See Appendices 15[2]-17)
OPT listening	90	.73	10 items were deleted from the original task in order to improve the internal reliability of the task (See Appendices 18-19)
WRS	90	.97	
LVLT	150	.86	Six subsets in total
TROG-2	80	.88	20 blocks in total
OPT grammar	90	.75	Two subsets in total. The third subset was deleted in order to improve the internal reliability of the task
OGJ	48	.95	
Raven's SPM	60	.78	

An independent *t*-test was used to test whether the LC proficiency of Group 1 and Group 2 was significantly different. Inspections of the two group means indicated that the mean LC score of Group 1 (M = 12.06, SE = 0.26) was lower than the score of Group 2 (M = 17.93, SE = 0.66) (see Table 5.2). This difference, -5.86, 95%CI [-7.24, -4.78], was significant, t (49.55) = -8.54, p < .001. The scores of the two groups on the CET4 listening have been transformed

[2] In the SPSS output, I used different names for the variables in the SPSS computations. PETTL = the PET listening; ZCET4TL = the CET4 listening; WSEGTTL = the word segmentation test; WRSTL = the word recognition from speech test; OPTLSNTL = the OPT listening; LVLTTL = LVLT; OPTGTL = the OPT grammar; TROG2TL = the TROG-2; RAVENTL = the Raven's SPM; Enguse = PBQ; SPSAVER = OGJ.

into z-scores, i.e., they have the same mean, equal to zero, and standard deviation (SD), equal to one, and therefore there is no point in comparing them. Thus, the scores on the PET listening indicated that LC of Group 1 was lower than that of Group 2 and that the difference was significant.

Table 5.2 Mean scores on the PET listening between Group 1 (*N*=147) and Group 2 (*N*=40)

Group	Mean	SD
1	12.06	3.14
2	17.93	4.02

Table 5.3 presents means for all participants' scores on each task used to measure the dependent variables and the observed variables.

Table 5.3 Summary of the minimum and the maximum scores, the means and SDs of each task (for all informants from both groups) (*N*=187)

Task	*Mean*	*SD*	Minimum score	Maximum score
PET listening	13.32	4.11	5	22
CET4 listening	0	1	-1.84	1.87
WST	44.03	5.22	34	54
OPT listening	70.19	6.43	57	83
WRS	44.40	21.76	6	88
LVLT	100.47	16.98	66	135
TROG-2	60.80	10.67	39	79
OPT grammar	55.20	8.61	38	72
OGJ	1787.77	821.89	422.47	4130.64
WMCF	15.08	1.31	12	16
WMCB	8.45	3.02	3	14
WMEF	7.87	1.86	4	12
WMEB	5.33	1.59	2	8
Raven's SPM	44.95	4.74	33	56
PBQ	14.34	20.97	0	89.47

The results per group provided a different picture. Table 5.4 presents the statistics for the scores from each group on each task used to measure LC and on the observed variables. Table 5.4 shows that, generally, the mean scores for Group 1 were lower than those for Group 2 for the PET listening, WST, OPT listening task, WRS, LVLT, TROG-2, OPT grammar task, WM, and Raven's SPM. For the OGJ, the RTs for Group 1 were less than those for Group 2 which means the aural sentence processing speed of Group 1 was faster than that of Group 2. The strongest partial eta squared was found for PBQ, followed by WRS and LVLT, and the weakest partial eta squared was found for OGJ. The results mean that the difference in the frequency of use of English in daily life between the two groups was greater than any other difference, and the difference in sentence processing speed between two groups was the smallest of all other differences. The mean LC scores from the CET4 for the two groups were taken from the raw scores which were calculated with different scoring systems.

Table 5.4 Means, SDs, and Partial eta squared for the differences in Group 1 (*N*=147) and Group 2 (*N*=40) mean scores

Measure	Group 1 Mean (s)	Group 2 Mean (s)	Partial eta squared
Dependent variables			
LC from PET (Max = 25)	12.06 (3.14)	17.93 (4.02)	0.34
LC from CET4 (Max N/A)	140.08 (23.96)	15.13 (4.51)	N/A
Language knowledge measures			
Phonological knowledge measures			
WST (Max = 61)	43.4 (5.2)	46.3 (4.7)	0.05
OPT listening (Max = 90)	68.7 (5.7)	75.6 (6.1)	0.19
WRS (Max = 90)	36.2 (15.7)	74.5 (12.4)	0.52
Vocabulary knowledge measure			
LVLT (Max = 150)	95.1 (13.9)	120.2 (12.3)	0.37
Grammar knowledge measures			
TROG-2 (Max = 80)	58.3 (10.3)	70.0 (6.3)	0.20
OPT grammar (Max = 90)	54.4 (8.4)	58.0 (9.0)	0.03

Processing speed measure			
OGJ (ms)	1740 (785)	1963 (937)	0.01
Working memory measures			
Forward digit span in Chinese	15.0 (1.4)	15.5 (0.8)	0.03
Backward digit span in Chinese	8.1 (3.1)	9.9 (2.4)	0.06
Forward digit span in English	7.5 (1.9)	9.1 (1.3)	0.11
Backward digit span in English	5.0 (1.5)	6.5 (1.2)	0.16
Reasoning ability measure			
Raven's SPM (Max = 60)	44.6 (4.7)	46.1 (4.9)	0.02
Social factor measure			
PBQ (Max = 100)	4.33 (3.99)	51.12 (16.53)	.78

Table 5.5 Mean percent correct for each stimulus type (positive vs negative) in WST (N = 187)

Stimulus type	Mean	SD[a]	SE[b]
Aspiration (n = 36)	74.14	10.57	0.77
Positive aspiration (n = 18)	77.63	14.56	1.06
Negative aspiration (n = 18)	70.65	16.72	1.22
Glottal stop (n = 36)	67.28	6.91	0.51
Positive glottal stop (n = 18)	58.85	11.00	0.80
Negative glottal stop (n = 18)	75.70	11.61	0.85
Double cue (n = 12)	61.54	12.79	0.94
Double cue (Asp - / Glot +) (n = 6)	51.69	23.67	1.73
Double cue (Asp + / Glot -) (n = 6)	71.39	22.20	1.62
All stimuli (n = 84)	69.40	6.23	0.46

SD[a] = standard deviation.

SE[b] = standard error.

Results of the mean percentage correct responses for each stimulus type from original scores of WST are presented in Table 5.5. Results indicated that participants' mean accuracy on all items was 69.4%. Among three types of stimulus, participants had less accuracy on Double cue stimuli (61.54%) than on Glottal stimuli (67.28%), or on Aspiration cue stimuli (74.14%). Results from the positive and negative stimuli showed that participants had the lowest accuracy on stimuli with a positive glottal stop without aspiration cue (51.69%), and had the highest accuracy on stimuli with a positive aspiration cue (77.63%). The results suggested that for Chinese learners it was easier to segment words which have aspiration cues than to segment words which have glottal stop cues. Accuracy on stimuli with a positive glottal stop cue (58.85%) provided more evidence for this. Appendix 20 shows that in the subcategories of positive Aspiration, negative Glottal stop, and Double cue with an aspiration cue but without a glottal stop cue, the maximum accuracy was 100%. However, in the Double cue with a glottal stop cue but without an aspiration cue subcategory the maximum accuracy was 100% and the minimum accuracy was 0. The lowest accuracy was for item 18 (***an ice man***/*a nice man*) (22.5%), and the highest accuracy was for item 71 (*be rice*/***beer ice***).

Correlation analyses were conducted for all participants. Table 5.6 provides a correlation matrix with all correlations between variables for all participants. The correlation analyses were undertaken to prepare for the hierarchical regression analyses of LC models. In later hierarchical regression analyses, the order of entry of independent variables into LC models depends on values of correlation between the dependent variables and each independent variable. The key results in this Table can be summarised as follows: the PET listening and the CET4 listening correlated significantly, though this correlation was not very strong ($r = .31$, $p < .01$). This might be a result of the two listening tasks measuring different skills. The current study found that the CET4 listening measured learners' ability to comprehend explicit information, but the PET listening measured learners' abilities to apply linguistic knowledge to processing speech input and to apply world knowledge and strategies. This is discussed in Section 6.7. When the PET listening was used to measure LC, all variables, except OGJ, significantly correlated with LC. The amount of PBQ and WRS had the highest correlation with the PET listening ($r = .59$, $p < .01$). The other significant correlations varied from .47 ($p < .01$) to .16 ($p < .05$). When the CET4 listening was used to measure LC, seven out of 13 predictors correlated significantly with LC. The seven predictors were WST, OPT listening, WRS, LVLT, TROG-2, OPT grammar, and WMEF. WRS and LVLT had the highest correlation with LC ($r = .36$, $p < .01$). The other significant correlations varied from .33 ($p < .01$) to .19 ($p < .05$). The highest significant correlation between each pair of predictors was between WRS and LVLT ($r = .82$, $p < .01$), followed by the correlation between WRS and PBQ ($r = .73$, $p < .01$), and the correlation between WRS and TROG-2 ($r = .71$, $p < .01$). OGJ was found to have no significant correlation with the other predictors and LC.

Table 5.6 Pearson correlations for all observed variables for all participants (*N* = 187)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. PET listening	1	.31**	.22**	.42**	.59**	.47**	.45**	.36**	.10	.16*	.21**	.25**	.37**	.20**	.59**
2. CET4 listening		1	.19*	.22*	.36**	.36**	.31**	.33**	.12	.02	.10	.20**	.14	.03	.08
3. WST			1	.29**	.35**	.37**	.32**	.33**	.02	.16*	.20**	.30**	.22**	.11	.21**
4. OPT listening				1	.61**	.47**	.52**	.35**	-.05	.08	.16*	.31**	.29**	.15*	.43**
5. WRS					1	.82**	.71**	.46**	.02	.18*	.30**	.36**	.43**	.5*	.73**
6. LVLT						1	.61**	.37**	-.06	.20**	.21**	.34**	.29**	.09	.61**
7. TROG-2							1	.44**	-.07	.16*	.14	.30**	.35**	.22**	.45**
8. OPT grammar								1	.02	.21**	.19**	.19**	.31**	.21**	.22**
9. OGJ									1	-.00	.09	-.03	.12	.12	.14
10. WMCF										1	.13	.17*	.11	.02	.18*
11. WMCB											1	.17*	.52**	.10	.28**
12. WMEF												1	.41**	.11	.31**
13. WMEB													1	.13	.41**
14. Raven's SPM														1	.17*
15. PBQ															1

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

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

Notes:1. PET listening = the listening section of the Preliminary English Test; 2. CET4 listening = the listening section of the College English Test Band 4; 3. WST = word segmentation test; 4. OPT listening = the listening test of the Oxford Placement Test; 5. WRS = the test of word recognition from speech; 6. LVLT = the Listening Vocabulary Levels Test; 7. TROG-2 = the Test for Reception of Grammar version 2; 8. OPT grammar = the Oxford Placement Test grammar test; 9. SPS = sentence processing speed test; 10. WMCF = working memory forward task in Chinese; 11. WMCB = working memory backward task in Chinese; 12. WMFE = working memory forward task in English; 13. WMEB = working memory backward task in English; 14. Raven’s SPM = Raven’s Standard Progressive Matrices; 15. PBQ = personal background questionnaire.

An overview of the correlations between all observed variables for Groups 1 and 2 can be found in Table 5.7. Correlation analysis was conducted to prepare for the LC models of hierarchical regression analyses for each subgroup. As there were many predictors it was not possible to include each in the regression models. The independent variables which correlated most strongly with listening were used to build LC models for both subgroups. Additionally, in later hierarchical regression analyses, the order of entry of the independent variables into each LC model was changed to provide further information on the contribution of each variable. Interestingly, the PET listening and the CET4 listening correlated much more strongly for Group 2 (r = .73, p < .01) than for Group 1 (r = .25, p <.01). There may be two reasons for this: 1) the CET4 listening scores were calculated differently for each subgroup because I did not have the formula for the weighting of the different components; 2) the PET listening was not equally reliable for both groups. For Group 1 the reliability was low (Cronbach's Alpha = .54), but the reliability for Group 2 was high (Cronbach's Alpha = .84).

For Group 1, when the scores from the PET listening were entered, the highest significant correlation was between LC and WRS (r = .20, p < .05), although this was very weak, followed by the correlation between LC and OPT listening (r = .18, p < .05), and the correlation between LC and OPT grammar (r = .18, p < .05). When the scores from the CET4 listening were entered, the highest significant correlation remained between LC and WRS (r = .51, p < .01). The correlation between LC and LVLT (r = .42, p < .01) and the correlation between LC and TROG-2 (r = .31, p < .01) followed. The highest correlation between each pair of independent variables was between WRS and LVLT (r = .71, p < .01), followed by the correlation between WRS and TROG-2 (r = .61, p < .01) and the correlation between WRS and OPT listening (r = .51, p < .01). PBQ had no significant correlation with the PET listening, the CET4 listening, or any other observed variables, neither did OGJ.

For Group 2, when the scores from the PET listening were entered, the significant correlations between LC and the predictors ranged from .74 (p < .01) to .36 (p < .05) and were higher than those for Group 1, which ranged from .20 (p < .05) to .17 (p < .05). The highest correlation was between LC and OPT grammar. The correlation between LC and TROG-2 (r = .74, p < .01) and the correlation between LC and WRS (r = .69, p < .01) followed. When the scores from the CET4 listening were entered, OPT grammar, TROG-2, and WRS retained the highest correlations with LC (r = .64, .63, .56, respectively p < .01). The highest correlation between each pair of the predictors was from WRS and TROG-2 (r = .74, p < .01), followed by TROG-2 and OPT grammar (r = .72, p < .01), and LVLT and

OPT grammar (r = .72, p < .01). OGJ was found to have no significant correlation with the PET listening, the CET4 listening, or any other variables.

Table 5.7 Pearson correlations for all observed variables for Group 1 (*N*=147) above the diagonal, and Group 2 (*N*=40) below the diagonal

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. PET listening	1	.25**	.02	.18*	.20*	.05	.17*	.18*	.11	.07	.05	-.01	.12	.07	.03
2. CET4 listening	.73**	1	.14	.26**	.51**	.42**	.31**	.24**	.15	-.02	.08	.21*	.10	-.12	.07
3. WST	.37*	.37*	1	.18*	.29**	.30**	.25**	.28**	.06	.13	.17*	.27**	.16	.06	-.00
4. OPT listening	.36*	.15	.37*	1	.51**	.31**	.41**	.27**	-.12	-.02	.04	.14	.10	.16*	.04
5. WRS	.69**	.56**	.21	.31	1	.71**	.61**	.45**	-.09	.06	.16	.15	.18*	.02	.06
6. LVLT	.65**	.55**	.32*	.20	.62**	1	.49**	.24**	-.15	.10	.07	.15	.01	-.03	.08
7. TROG-2	.74**	.63**	.28	.43**	.74**	.44**	1	.37**	-.15	.10	.00	.15	.18*	.17*	-.01
8. OPT grammar	.74**	.64**	.41**	.48**	.71**	.72**	.72**	1	.07	.16	.15	.07	.21**	.13	-.00
9. OGJ	-.12	.03	-.21	-.11	-.10	-.21	-.07	-.20	1	.01	.05	-.04	.08	.04	.04
10. WMCF	.22	.22	.18	.20	.40*	.39*	.20	.39*	-.20	1	.09	.12	.02	-.01	-.04
11. WMCB	.20	.16	.08	.11	.35*	.11	.23	.19	.13	.20	1	.09	.45**	.08	.15
12. WMEF	.38*	.23	.10	.43**	.41**	.40*	.41**	.54**	-.23	.21	.16	1	.32**	.09	-.08
13. WMEB	.47**	.37*	.12	.33*	.54**	.38*	.45**	.52**	.13	.30	.58**	.37*	1	.08	-.02
14. Raven's SPM	.40*	.51**	.17	-.10	.37*	.24	.30	.36*	.28	.06	.02	-.03	.15	1	-.04
15. PBQ	.31	.38*	.04	.12	.59**	.39*	.34*	.39*	.18	.37*	.24	.12	.39*	.36*	1

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

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

The correlation between the frequency of the use of English in daily life and the length of study time in the UK for the entire group of participants was analysed (see Appendix 21). The results indicated that the frequency of using the L2 significantly correlated with the length of study time in the L2 environment (r = .84, p < .01).

5.2 Regression analyses

Before testing against the hypothesized SEM models of LC, regression analyses were carried out in order to discover to what extent the predictors could explain the variance in LC for all participants and for each subgroup. Since two different listening tasks were used to measure participants' LC, two regression analyses were undertaken for the entire group (N = 187): one with the dependent variable of the PET listening and the other with the CET4 listening. Two regression models were built based on the results from the correlation analyses (see Table 5.6). The analyses are presented as follows: firstly, with the PET listening as the dependent variable for the entire group (see Section 5.2.1); secondly, with the CET4 listening as the dependent variable for the entire group (see Section 5.2.2); and finally, for each subgroup (see Section 5.2.3).

5.2.1 Predicting LC as measured with the PET listening task

As can be seen in Table 5.6, there were significant correlations between the PET listening task and the predictors: WST; OPT listening; WRS; LVLT; TROG-2; OPT grammar; WM tasks; Raven's SPM; and, PBQ. Therefore, these predictors were entered into a regression model. According to Field (2017), if the values of tolerance are less than 0.01 and the values of VIF are greater than 10, it means that there is a problem of multicollinearity. As can be seen in Appendix 22, no problem with multicollinearity existed. In addition, Figure 5.1 shows that residuals were distributed normally and the correlation between them and the predicted variables was not linear. Therefore, the assumptions for carrying out a regression analysis were met. Since I am interested in knowing whether any of those variables could explain unique variance, hierarchical multiple regression analysis was used rather than a normal regression analysis which contains shared variance.

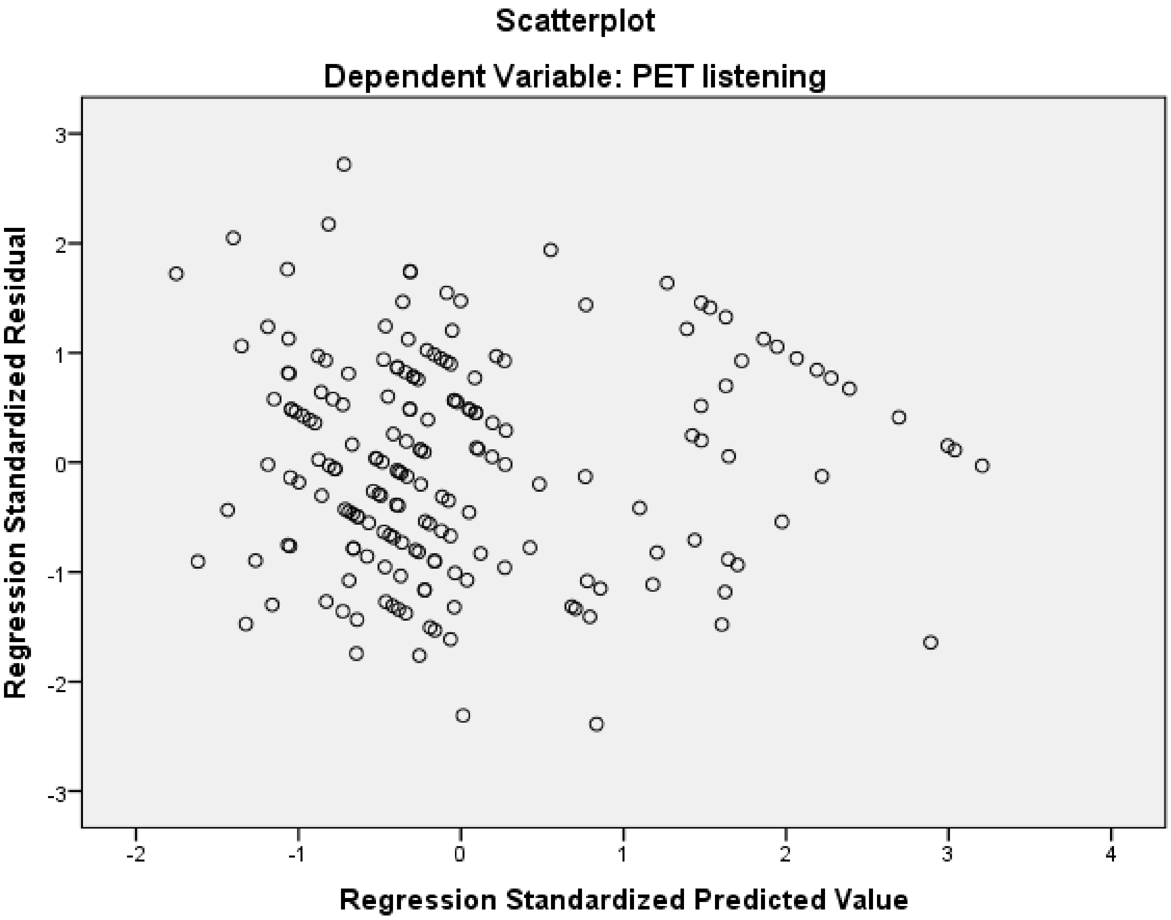


Figure 5.1 Multiple linear regression: standardized predicted values and residuals

Although we have seen in Chapter 2 that decoding is a key problem for Chinese learners of English, it is not clear which decoding processes are most difficult: word segmentation (WST) or word recognition from speech (WRS). Therefore, a hierarchical regression analysis was undertaken to analyse the unique contribution of WST and WRS to LC. In the first regression analysis, I inputted WRS in the first step and WST in the second step. Then in the second regression analysis, I reversed the order of entry to WRS and WST. As can be seen in Appendix 23, the first model was significant (F (1, 185) = 98.13, p < .001), and WRS explained 34.7% variance in LC (β = .59, p < .001) (see Appendices 24 & 25). When WST was entered, it did not explain any additional variance, although the model remains significant (F (2, 184) = 48.83, p < .001) (see Appendices 23 & 24). In the second regression analysis, WST was entered in the first step. Appendices 26 and 27 indicated that the model was significant (F (1, 185) = 9.17, p < .05) and WST explained 4.7% of the variance in LC (β = .22, p < .05). When, in a second step, WRS was added a further 29.9% variance was explained (β = .58, p < .001). The addition of the explained variance made a significant difference to the model, and the overall model was still significant (F (2, 184) = 48.83, p < .001) (see Appendices 26 & 28). The two variables explained 34.7% of the variance in LC. The results suggest that WRS explained the variance in LC over and above the contribution of WST, because WRS uniquely explained 29.9% of the variance in LC. However, WST did not explain unique variance over and above the contribution of WRS.

Next, the focus turned to an exploration of the contribution of variables representing PK (WST, OPT listening, and WRS), LVLT, and GK (OPT grammar and TROG-2), to explaining variance in LC. As can be seen in Table 5.6, the correlation between LC and WRS was the strongest among the other variables. Therefore, in the first regression analysis, the variables representing PK were entered first, followed by LVLT, and finally the variables representing GK. Then the order of the variables representing PK and LVLT was reversed for the second analysis. Appendix 29 indicates that the first model was significant (F (3, 183) = 33.28, p < .001), and WST, OPT listening, and WRS together explained 35.3% of the variance in LC (see Appendix 30). When LVLT was entered, no additional variance was explained in the model, but the model remained significant (F (4, 182) = 24.84, p < .001) (see Appendices 29 - 31). When OPT grammar and TROG-2 were entered, a further 0.9% variance was explained. However, the addition of the explained variance did not make a significant change to the model although the overall model remained significant (F (6, 180) = 17.07, p < .001). In the second regression analysis, LVLT was entered in the first step. Appendices 32 and 33 indicated that the model was significant (F (1, 185) = 53.75, p < .001), and LVLT explained 22.5% of the variance in LC (β = .47, p < .001). When, in a second step, WST, OPT listening, and WRS were added, a further 12.8% variance was explained. The addition of the explained variance made a significant change to the model, and the overall model was significant (F (4, 182) = 24.84, p < .001) (see Appendices 32 - 34). When, in a third step, OPT grammar and TROG-2 were added to the model a further 0.9% variance was explained, but the addition of the explained variance did not make a significant change to the model. However, the overall model was still significant (*F* (6, 180) = 17.07, *p* < .001). Together the variables explained 36.3% of the variance in LC. The results suggest that together WST, OPT listening, and WRS explained unique variance in LC over and above the contribution of LVLT. The variables representing PK uniquely explained 12.8% of the variance in LC. In addition, although LVLT is a significant predictor on its own, when entered with other variables it does not explain unique variance. The variables representing GK were not significant predictors. In other words, phonological knowledge has much more predictive power for LC than vocabulary or grammar knowledge.

As can be seen in Table 5.6, WRS and PBQ have the same strong correlation with LC (*r* = .59, *p* < .01). To discover whether PBQ explains any variance in LC over and above the contribution of the variables representing PK, in the first regression analysis, I inputted PBQ in the first step, then WST, OPT listening, and WRS in the second step, followed by the other variables in next steps. Then, in the second regression analysis, I reversed the order of entry of the variables: variables representing PK were entered first, then PBQ, and then the other variables.

In the first regression analysis, when only PBQ was entered as a predictor, the model was significant (*F* (1, 185) = 99.11, *p* < .001), and PBQ explained 34.9% of the variance in LC (*β* = .59, *p* < .001) (see Appendices 35 - 37). When, in a second step, WST, OPT listening, and WRS were added a further 6.2% of the variance was explained. The addition of this variable made a significant change to the explained variance of the model. When OPT grammar and TROG-2 were added a further 2.2% variance was explained. When the other variables were added further to the model in the next steps, the addition of each explained variance (1% together) was very little and did not make significant change to the model. The variables together explained 44.2% of the variance in LC.

In the second regression analysis, I entered WST, OPT listening, and WRS as a predictor in the first instance. This model was significant (*F* (3, 183) = 33.28, *p* < .001) and the three variables explained 35.3% of the variance in LC (see Appendices 38 and 39). When, in a second step, PBQ was added a further 5.8% of the variance was explained. The addition of the explained variance made significant change to the model, and the model was significant (*F* (4, 182) = 31.73, *p* < .001). Therefore, the change in R^2, produced by adding PBQ, was almost of the same magnitude as the change in R^2 that was produced by adding the above three variables in the previous model (6.2%). The results indicate that most of the variance explained by PBQ is shared with the other variables, but there is a small contribution of PBQ over and above the other variables. PBQ uniquely explained 5.8% of the variance in LC. When OPT grammar and TROG-2 were added a further 2.2% variance was explained. The addition of the additional variables made a significant change to the model, and the model was still significant (F (6, 180) = 22.85, $p < .001$). When the other variables were added to the model, no additional variance was explained. Table 5.6 shows that only OGJ was not significantly correlated with LC, therefore, OGJ was not originally included in the analyses described above. To verify whether it would make a difference to the model, it was added in a final step. Results indicate that, as expected, the variable did not significantly contribute to the model (see Table 5.8). All variables together explained 44.3% of the variance in LC (see Appendices 40 to 42 for details).

Table 5.8 Regression models explaining LC measured with the PET listening (N = 187)

Model	R	R square	Adjusted R square	Std. error of the estimate	R squared changed	Sig. F change
1	.594[a]	.353	.342	3.336	.353	< .001
2	.641[b]	.411	.398	3.193	.058	< .001
3	.658[c]	.432	.413	3.151	.022	< .05
4	.658[d]	.433	.411	3.158	.001	> .05
5	.662[e]	.439	.403	3.178	.006	> .05
6	.665[f]	.442	.403	3.178	.003	> .05
7	.666[g]	.443	.401	3.183	.001	> .05

a. Predictors: (Constant), WRS, WST, OPT listening

b. Predictors: (Constant), WRS, WST, OPT listening, PBQ

c. Predictors: (Constant), WRS, WST, OPT listening, PBQ, OPT grammar, TROG-2
d. Predictors: (Constant), WRS, WST, OPT listening, PBQ, OPT grammar, TROG-2, LVLT
e. Predictors: (Constant), WRS, WST, OPT listening, PBQ, OPT grammar, TROG-2, LVLT, WMCF, WMCB, WMEF, WMEB
f. Predictors: (Constant), WRS, WST, OPT listening, PBQ, OPT grammar, TROG-2, LVLT, WMCF, WMCB, WMEF, WMEB, Raven's SPM
g. Predictors: (Constant), WRS, WST, OPT listening, PBQ, OPT grammar, TROG-2, LVLT, WMCF, WMCB, WMEF, WMEB, Raven's SPM, OGJ

To summarise, together observed variables explained 44.3% of the variance in LC. Only WST, OPT listening, WRS, PBQ, OPT grammar, TROG-2 and LVLT were predictors of LC, but other variables were not significant predictors. The results mean that phonological knowledge, frequency of L2 use, vocabulary knowledge and grammar knowledge are significant predictors of LC, and the first two factors have approximately the same predicting power. Phonological knowledge could uniquely explain 12.8% of the variance in LC, followed by frequency of L2 use, which uniquely explained 5.8% of the variance. Vocabulary knowledge could explain overall 22.5% of the variance in LC and grammar knowledge explained 2.2% of the variance in LC. Results also indicate that when only those variables which represent linguistic knowledge were entered into the model, the variables together could explain 36.3% of the variance in LC and that phonological knowledge had the greatest predictive power. Vocabulary knowledge as measured in the current study is a significant predictor on its own. In addition, results show that WRS uniquely explained 29.9% of the variance in LC. This is over and above the contribution of WST, which means learners' ability to recognise words from speech has more predictive power of LC than learners' ability to segment words.

5.2.2 Predicting LC as measured with the CET4 listening task

In this section, hierarchical regression analyses were undertaken to analyse the contribution of different variables to explaining the variance in LC as measured with the CET4. It is possible that different variables are responsible for explaining the variance in listening if the

CET4 is used instead of the PET. First of all, the aim was to discover whether decoding problems were caused by the fact that learners are unable to segment words, or by the fact that learners were unable to recognise words from speech. In the first regression analysis, I inputted WRS in the first step and WST in the second step. Then in the second regression analysis, I reversed the order of entry to WST and then WRS. As can be seen in Appendix 43, the first model was significant (F (1, 185) = 27.90, $p < .001$), and WRS explained 13.1% of the variance in LC ($\beta = .36$, $p < .001$) (see Appendices 44 & 45). When WST was entered, a further 0.4% variance was added, but the addition of the explained variance did not make a significant change to the model although the overall model remained significant (F (2, 184) = 14.37, $p < .001$) (see Appendices 43 & 44). In the second regression analysis, WST was entered in the first step. Appendices 46 and 47 indicate that WST explained 3.5% of the variance in LC ($\beta = .19$, $p < .05$), and that the model was significant (F (1, 185) = 6.64, $p < .05$). When in a second step WRS was added a further 10.0% variance was explained ($\beta = .34$, $p < .001$). The addition of the explained variance made a significant change to the model, and the model was significant (F (2, 184) = 14.37, $p < .001$) (see Appendices 46 - 48). Together, the two variables explained 13.5% of the variance in LC. WRS uniquely explained 10.0% of the variance in LC. The results suggest that WRS explained the variance in LC over and above the contribution of WST, but WST did not contribute to explaining any unique variance in listening. The results for both dependent variables (PET listening and CET4) were, therefore, similar in this respect.

Next, the analyses were undertaken in order to discover to what extent variables representing linguistic knowledge source could explain the variance in LC, and to what extent each type of linguistic knowledge could explain the variance in LC. Since the correlation between WRS and LC ($r = .36$, $p < .01$) was the same as that between LVLT and LC, and because WRS was one of the variables representing PK, the variables representing PK were entered first, followed by LVLT. Then, TROG-2 and OPT grammar were entered in the first regression analysis. In the second analysis, the order of the variables representing PK and LVLT was reversed. As can be seen in Appendix 49, the first model was significant (F (3, 183) = 9.53, $p < .001$), when WRS, WST, and OPT listening were entered, and the three variables explained 13.5% of the variance in LC (see Appendices 50 & 51). When LVLT was entered, a further 1% of the variance was explained, but the addition of this variable did not make significant change to the overall model, even though it remained significant (F (4, 182) = 7.74, $p < .001$) (see Appendices 49 & 50). When TROG-2 and OPT grammar were entered, a further 3.3% variance was explained. The addition of these variables made a significant change to the model, and the overall model was also still significant (F (6, 180) = 6.53, $p < .05$).

In the next model, the order of entry of the variables in the second and third steps was reversed. After the variables representing PK, TROG-2, and OPT grammar were entered and then LVLT. Results show that when TROG-2 and OPT listening were added a further 3.3% of the variance was explained, and the model was significant (see Appendices 52 - 54). In the third regression analysis, LVLT was entered in the first step. Appendices 55 and 57 indicate that the model was significant (F (1, 185) = 27.60, $p < .001$), and LVLT explained 13% of the variance in LC ($\beta = .36$, $p < .001$). When in a second step WRS, WST, and OPT listening were added a further 1.6% of the variance was explained, but the addition of the phonological variables did not make a significant change to the model. The model was significant (F (4, 182) = 7.74, $p < .001$) (see Appendices 55 - 57). The results mean that phonological knowledge is a significant predictor on its own, but when entered with other variables, it did not explain unique variance. The results also mean that vocabulary knowledge is a significant predictor on its own. When in a third step TROG-2 and OPT grammar were added to the model a further 3.3% variance was explained. The addition of the explained variance made a significant change to the model, and the overall model was still significant (F (6, 180) = 6.53, $p < .001$). Then, the order of entry of variables representing PK and variables representing GK was reversed. Results indicate that when TROG-2 and OPT grammar were added a further 4.7% of the variance was explained. The addition of these variables made a significant change to the model, and the overall model was significant (see Appendices 58 -- 60). The variables representing PK explained a further 0.2% of the variance, but the addition of explained variance did not make a significant change to the model. Together the variables representing linguistic knowledge explained 17.9% of the variance in LC. The results (see Appendices 53 & 59) indicate that phonological knowledge and vocabulary knowledge, each on its own, explained approximately the same amount of overall variance in LC.

The results for both dependent variables were therefore different in this respect. When LC was measured with the CET4 listening, phonological knowledge and vocabulary knowledge were significant predictors on their own: the former explained 13.5% of the overall variance and the latter explained 13% of the overall variance in LC. Grammar knowledge explained 4.7% of the variance in LC. The above variables together explained 17.9% of the variance in LC. However, when LC was measured with the PET listening, phonological knowledge explained the variance in LC over and above the contribution of vocabulary knowledge. Phonological knowledge uniquely explained 12.8% of the variance in LC. Vocabulary knowledge is a significant predictor on its own. Grammar knowledge was a significant predictor. Linguistic knowledge explained 36.3% of the variance in LC.

As can be seen in Table 5.6, only WRS, LVLT, the OPT grammar, TROG-2, the OPT listening, WMEF, and WST significantly correlated with LC. However, it would be useful to explore the contribution of Raven's SPM, OGJ, and PBQ to explaining the variance in LC, even though they did not correlate significantly with LC. Therefore, Raven's SPM, OGJ, and PBQ were added to the model with three other variables used to measure working memory capacity. In the first regression analysis, Appendices 61 to 63 show that variables representing PK explained 13.5% of the variance in LC. The addition of the variables which represent GK, PBQ, and OGJ each made significant change to the model. Then OGJ and PBQ were entered before LVLT which had been entered to the model just after the variables representing GK but the explained variance did not make significant change to the model (see Appendix 62), and the model was reanalysed. Results are presented in Table 5.9. Results show that together the variables explained 28.2% of the variance in LC. Phonological knowledge explained 13.5% of the variance in LC, followed by grammar knowledge explaining 3.3% of the variance, the frequency of use of English explaining 5.1%, aural sentence processing speed explaining 2.5%, and vocabulary knowledge explaining 2.1% of the variance (see Appendices 64 - 66 for details). The explained variance by WM capacity and Raven' SPM (together 1.7%) did not make significant changes to the model, which means that variables to measure working memory capacity and reasoning ability were not predictors of LC. The results suggest that variables that represent learners' linguistic knowledge contribute significantly to the model, and variables that represent learners' phonological knowledge contribute the most to the model.

Table 5.9 Regression models explaining LC measured with the CET4 listening (N = 187)

Model	R	R square	Adjusted R square	Std. error of the estimate	R squared changed	Sig. F change
1	.368[a]	.135	.121	.89	.135	< .001
2	.410[b]	.168	.145	.88	.033	< .05
3	.468[c]	.219	.193	.85	.051	< .05
4	.494[d]	.245	.215	.84	.025	< .05
5	.515[e]	.265	.232	.83	.021	< .05
6	.529[f]	.280	.230	.83	.015	> .05
7	.531[g]	.282	.228	.83	.002	> .05

a. Predictors: (Constant), WRS, WST, OPT listening

b. Predictors: (Constant), WRS, WST, OPT listening, OPT grammar, TROG-2

c. Predictors: (Constant), WRS, WST, OPT listening, OPT grammar, TROG-2, PBQ

d. Predictors: (Constant), WRS, WST, OPT listening, OPT grammar, TROG-2, PBQ, OGJ

e. Predictors: (Constant), WRS, WST, OPT listening, OPT grammar, TROG-2, PBQ, OGJ, LVLT

f. Predictors: (Constant), WRS, WST, OPT listening, OPT grammar, TROG-2, PBQ, OGJ, LVLT, WMCF, WMCB, WMEF, WMEB

g. Predictors: (Constant), WRS, WST, OPT listening OPT grammar, TROG-2, PBQ, OGJ, LVLT, WMCF, WMCB, WMEF, WMEB, Raven's SPM

The results for both dependent variables, therefore, are similar to each other in some respects but differ from each other in other respects. When LC was measured with the CET4 listening, the variables together explained 28.2% of the variance in LC. WRS, WST, and OPT listening explained 13.5% of the variance, followed by: OPT grammar and TROG-2, 3.3%; PBQ, which explained 5.1%; OGJ, 2.5%; LVLT, 2.1%; WM capacity, 1.5%; and Raven’s SPM, 0.2%. Phonological knowledge, grammar knowledge, vocabulary knowledge, frequency of use of English, and sentence processing speed were predictors of LC. Variables which represent working memory capacity and Raven’s SPM were not predictors of LC. Phonological knowledge and vocabulary knowledge were significant predictors on their own. Word recognition from speech explained uniquely 10.0% of the variance in LC. However, when LC was measured with the PET listening, the variables together explained a larger share of the variance, namely 44.3%. WRS, WST, and OPT listening explained 35.3% of the variance in LC, followed by PBQ which explained 5.8% of the variance, and TROG-2 and OPT grammar which explained 2.2% of the variance. LVLT explained 0.1%, followed by WM capacity, which explained 0.6%; Raven’ SPM, 0.3%; and OGJ, 0.1%. Phonological knowledge, grammar knowledge, vocabulary knowledge and frequency of use of English were predictors of LC. Sentence processing speed, WM capacity and reasoning ability were not predictors. Phonological knowledge uniquely explained 12.8% of the variance in LC. Word recognition from speech uniquely explained 29.9% of the variance in LC. For both dependent variables, the variables representing phonological knowledge contributed the most to LC models, followed by frequency of use of English, and grammar knowledge. Vocabulary knowledge was also a significant predictor. In addition, variables representing working memory capacity and reasoning ability were not significant predictors of LC in either model. Results indicate that sentence processing speed only significantly contributed to explaining the variance in LC when LC was measured with the CET4 listening, but it was not a significant predictor of LC when LC was measured with the PET listening. Word recognition from speech was clearly a more important predictor than word segmentation from speech stream when LC was measured with both listening tasks.

5.2.3 Predicting LC as measured with the CET4 listening for each subgroup

In order to discover learner variables which could explain differences in LC between two subgroups, separate hierarchical multiple regression analyses were undertaken for each subgroup. The dependent variable was LC as measured with the CET4 listening. The CET4 listening scores were chosen, rather than the PET listening scores, because for Group 1 the variance explained by the observed variables when LC was measured with the PET listening (10%) was much lower than the explained variance when LC was measured with the CET4 listening (36.4%). This may be because the CET4 is such a high stakes test in China, so Group 1 really had to do their best. Comparatively, there was no reward for them to do the PET listening. However, the PET listening remains reliable as its internal reliability for the entire group (N = 187) is .77. Since Group 2 comprised a small number of participants (N = 40), it was not possible to include all assumed learner variables in the analysis for each group. Therefore, for Group 1, variables that were significantly correlated with LC (WRS, OPT listening, LVLT, TROG-2, OPT grammar, WMEF) were entered into a regression model. Since PBQ and OGJ were found to be significant predictors of LC when measured with the CET4 listening for the entire group (N = 187) (see Section 5.2.2), the two variables were also entered into the model. For Group 2, only variables which had significant correlations with LC greater than .50 were taken into consideration for a model due to the small number of participants (WRS, OPT grammar, TROG-2, LVLT, and Raven's SPM). Although both OPT grammar and TROG-2 correlated significantly with LC, only OPT grammar was chosen because both variables measured learners' grammar knowledge, and correlation between OPT grammar and LC (r = .64, p < .01) was stronger than that between TROG-2 and LC (r = .63, p < .01). Therefore, WRS, OPT grammar, LVLT, and Raven's SPM were chosen as independent variables for Group 2.

Predicting LC for Group 1

WRS and OPT listening were entered into a LC model in the first step. Then, LVLT was added to the model in the second step, followed by the other variables in the next steps. Results in Appendices 67 to 69 indicate that when WRS and OPT listening were entered, the model was significant (F (2, 144) = 24.97, p < .001), but the addition of other variables, except for OGJ, did not significantly contribute to the model. Therefore, in the next analyses, the order of entry of OGJ was adjusted. OGJ was entered to the model just after WRS and OPT listening and the model was reanalysed. Results show that together the variables explained 32.8% of the variance in LC, among which WRS and OPT listening explained 25.7%, followed by OGJ which explained 3.9% (see Table 5.10). Other variables each explained very little variance in LC and the adding of explained variance (together 3.2%) did not make significant change to the model (see Appendices 70 to 72). The results (Table 5.10) mean that in this group learners' phonological knowledge, measured with WRS and OPT listening, and learners' sentence processing speed, measured with OGJ, are significant predictors of LC. These results also mean that learners' vocabulary knowledge, grammar knowledge, working memory capacity, and frequency of English use are not significant predictors of LC. For this group, phonological knowledge is the most important predictor of LC.

Table 5.10 Regression models explaining LC measured for Group 1 (*N* = 147)

Model	R	R square	Adjusted R square	Std. error of the estimate	R squared changed	Sig. F change
1	.507[a]	.257	.247	.82	.257	< .001
2	.545[b]	.297	.282	.80	.039	< .05
3	.556[c]	.309	.290	.79	.013	> .05
4	.556[d]	.309	.280	.80	.000	> .05
5	.572[e]	.327	.293	.79	.018	> .05
6	.573[f]	.328	.289	.79	.001	> .05

a. Predictors: (Constant), OPT listening, WRS

b. Predictors: (Constant), OPT listening, WRS, OGJ

c. Predictors: (Constant), OPT listening, WRS, OGJ, LVLT

d. Predictors: (Constant), OPT listening, WRS, OGJ, LVLT, OPT grammar, TROG-2

e. Predictors: (Constant), OPT listening, WRS, OGJ, LVLT, OPT grammar, TROG-2, WMEF

f. Predictors: (Constant), OPT listening, WRS, OGJ, LVLT, OPT grammar, TROG-2, WMEF, PBQ

Predicting LC for Group 2

OPT grammar was entered into a model in the first step because it has the strongest correlation with LC among the other variables (r = .64, p < .01). In the second step, WRS was entered, followed by LVLT and Raven’s SPM, based on the strength of their correlation with LC. Results (see Table 5.11) show that when OPT grammar was entered, it explained 40.4% of the variance in LC, and the model was significant (F (1, 38) = 25.73, p < .001). Then when WRS and LVLT were entered in the next two steps, the addition of the variance explained by WRS and LVLT did not make significant change to the model. However, when Raven’s SPM was added, a further 8.3% of the variance was explained in LC, and the addition of the explained variance made significant change to the model (see Appendices 73 - 75). In order to discover to what extent Raven’s SPM could explain the variance in LC, the order of entry of the variable was changed. It was entered into the model just after the entry of the OPT grammar and the model was reanalysed. Results in Table 5.11 indicate that the four variables together explained 52.2% of the variance in LC, among which OPT grammar explained 40.4%, followed by Raven’s SPM which explained 9.2%. The addition of explained variance by WRS and LVLT, together 2.7%, did not significantly contribute to explaining variance in LC (see Appendices 76 - 78). The results mean that, in this group, learners’ grammar knowledge and reasoning ability, measured with the OPT grammar and Raven’s SPM, respectively, are significant predictors of LC. Results also suggest that learners’ ability to recognise words from speech and their vocabulary knowledge measured with WRS and LVLT are not significant predictors of LC. For this group, grammar knowledge is the most important predictor of LC.

Table 5.11 Regression models explaining LC measured for Group 2 (N = 40)

Model	R	R square	Adjusted R square	Std. error of the estimate	R squared changed	Sig. F change
1	.636[a]	.404	.388	.77	.404	< .001
2	.704[b]	.495	.468	.72	.092	< .05
3	.712[c]	.507	.466	.72	.012	> .05
4	.723[d]	.522	.468	.72	.015	> .05

a. Predictors: (Constant), OPT grammar

b. Predictors: (Constant), OPT grammar, Raven's SPM

c. Predictors: (Constant), OPT grammar, Raven's SPM, WRS

d. Predictors: (Constant), OPT grammar, Raven's SPM, WRS, LVLT

The differences between the analyses for the two groups can be summarised as follows: For Group 1, learners' phonological knowledge and aural sentence processing speed are significant predictors of LC, but learners' vocabulary knowledge, grammar knowledge, working memory capacities, and frequency of use of English are not significant predictors of LC. The above variables together explained 32.8% of the variance in LC, of which learners' phonological knowledge explained 25.7%, followed by learners' sentence processing speed at 3.9%. For Group 1 phonological knowledge is the most important predictor of LC. For Group 2, learners' grammar knowledge and reasoning ability can significantly predict LC, but learners' word recognition ability and vocabulary knowledge cannot. The above variables together explained 52.2% of the variance in LC, of which learners' grammar knowledge explained 40.4%, followed by learners' reasoning ability at 9.2%. Grammar knowledge is the most important predictor of LC for Group 2.

5.3 Checking assumptions for testing hypothesized structural equation models of LC

Two hypothesized structural equation models of LC were built based on Andringa et al. (2012) (see Figures 3.1 & 3.2). In the two models, observed variables remained the same, but the dependent variables were measured with two different listening tasks.

First, multivariate normality was checked on the basis of the results for skewness and kurtosis for each variable. According to Kline (2016), if skewness is greater than 3 and kurtosis greater than 10, the assumption of multivariate normality is violated. Appendix 79 shows that the absolute values of skewness were less than 3 and kurtosis less than 10. According to Bentler (2005), as a normalized estimate, if Mardia's coefficient is less than 5, it indicates that there is no problem with multivariate nonnormality. As can be seen in Appendix 79, Mardia's coefficient was 1.55 which means that the assumption of multivariate normality was not violated. Appendix 22 indicates that there was no problem of multivariate collinearity between observed variables. The assumption of collinearity was not violated. As can be seen in Figures 5.2 and 5.3, the assumption of homoscedasticity was not violated.

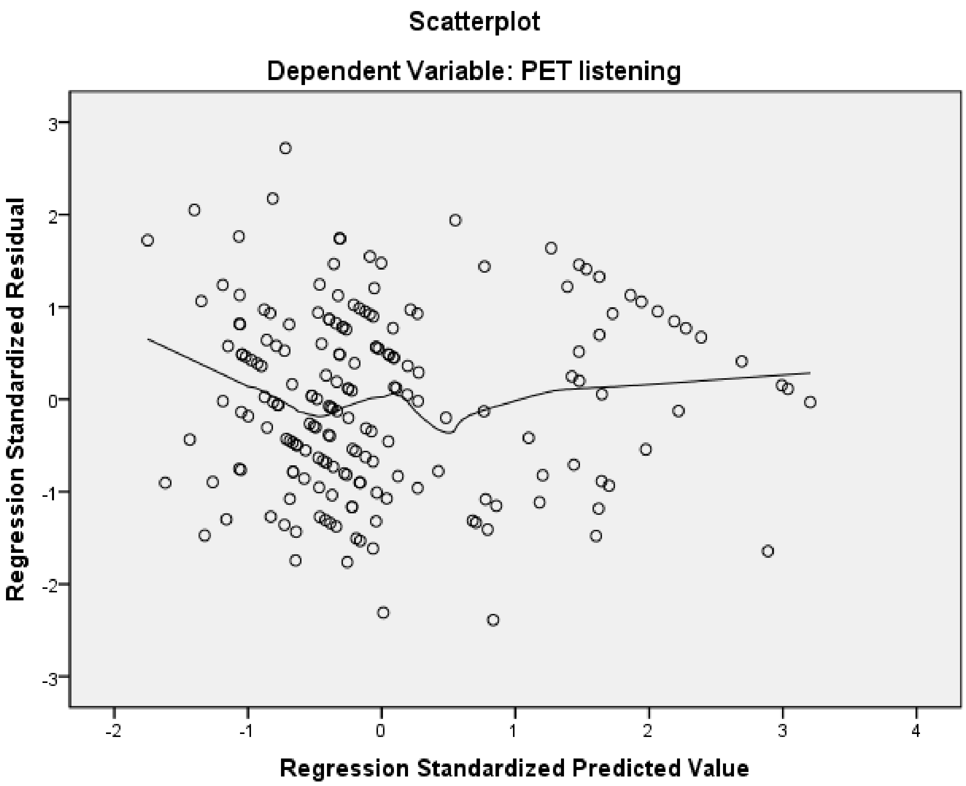


Figure 5.2 A normal plot of the regression standardized residuals with PET listening

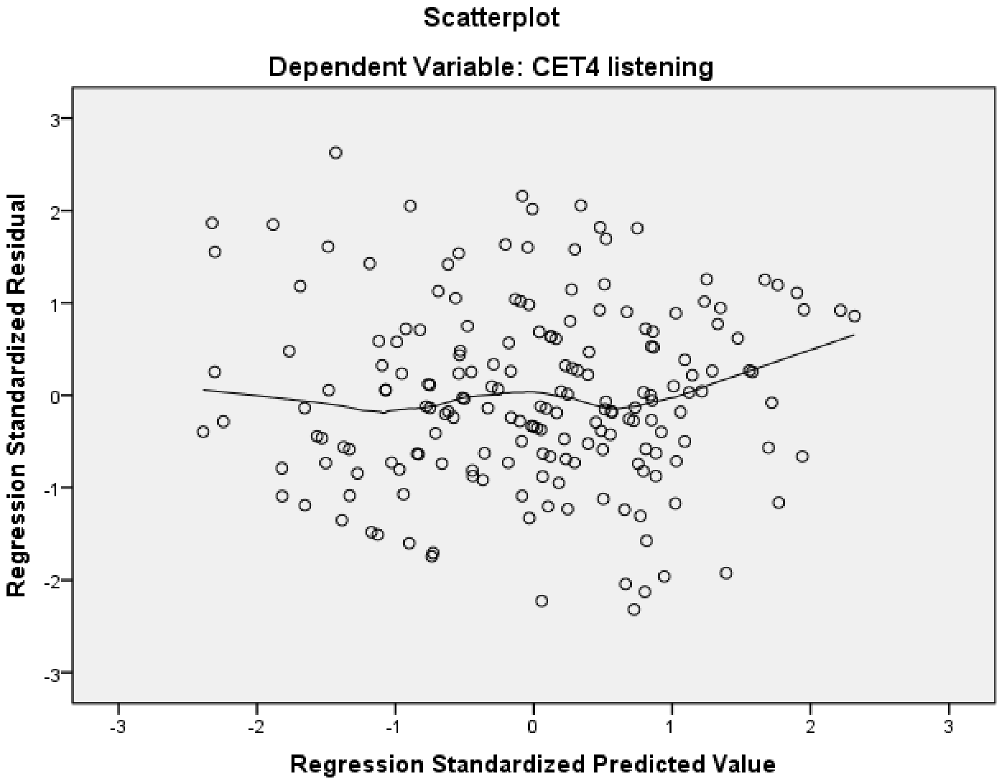


Figure 5.3 A normal plot of the regression standardized residuals with CET4 listening

Although there is a basic rule that the minimum sample size for a study using SEM is about 200 (Boomsma, 1987), and the sample in the current study, therefore, appeared to be too limited, the sample size in this study was checked through an A-priori Sample Size Calculator for Structural Equation models (https://www.danielsoper.com/statcalc/calculator.aspx?id=89). In the calculator, the necessary parameter values were set as follows: the anticipated effected size was 0.3; the desired statistical power level was 0.8; the number of latent variables was eight (excluding the latent variable of LC); the number of observed variables was 34; and the probability level was 0.05. The results of the calculation showed that the minimum sample size to detect effect is 177, and the minimum sample size for the model structure is 91. Since there were 187 participants in the study, the assumption of the minimum sample size was not violated. Positive definiteness was checked by analysing whether the determinant of the correlation matrix was equal to 0; the results indicated that the determinant did not equal 0 (Determinant = 4.011E-9).

Since it is preferred that each construct be indicated by at least three observed variables (Wu, 2010), the tasks in this study were split into different subsets. In order to ascertain whether the different subsets for each construct reflected the same pre-determined construct, the reliability of each construct was checked separately. LC was measured with the four sections of the PET listening. PK was measured with WST, OPT listening, and WRS. Each of the three tasks was split into two subsets, one of which contained the odd-numbered and one the even-numbered items. In total, there were, therefore, six subsets. VK was measured with LVLT. The task was split into six subsets based on the six parts in LVLT. GK was measured with TROG-2 and OPT grammar. TROG-2 was randomly split into three subsets, and the OPT grammar task was split into two subsets based on the two parts of the task (the third part had been deleted in order to improve the reliability of the task). SPS was measured with four randomly split subsets of OGJ. WM was measured with WMCF, WMCB, WMEF, and WMEB. RA was measured with the five parts of Raven's SPM. EU was measured with the four subsets of PBQ, which was split randomly in the questionnaire. Results of the reliability analyses for each latent construct are presented in Table 5.12.

Table 5.12 The results of analyses on the reliability of each latent construct

Latent Construct	Number of subsets in each construct	Reliability of each latent construct	Notes
LC	4	.75	
PK	6	.77	
VK	6	.86	
GK	5	.81	
SPS	4	.84	
WM	3	.57	The subset of WMCF was deleted (see Appendices 80 – 81)
RA	4	.59	Section A of Raven's SPM was deleted (see Appendices 82 – 84)
EU	4	.98	

Since two observed variables had been deleted from two individual constructs, the hypothesized structural equation models of LC (Figures 3.1 & 3.2) were changed accordingly. The new hypothesized structural models of LC were built based on the results of the reliability of each latent construct (see Figures 5.4 and 5.5).

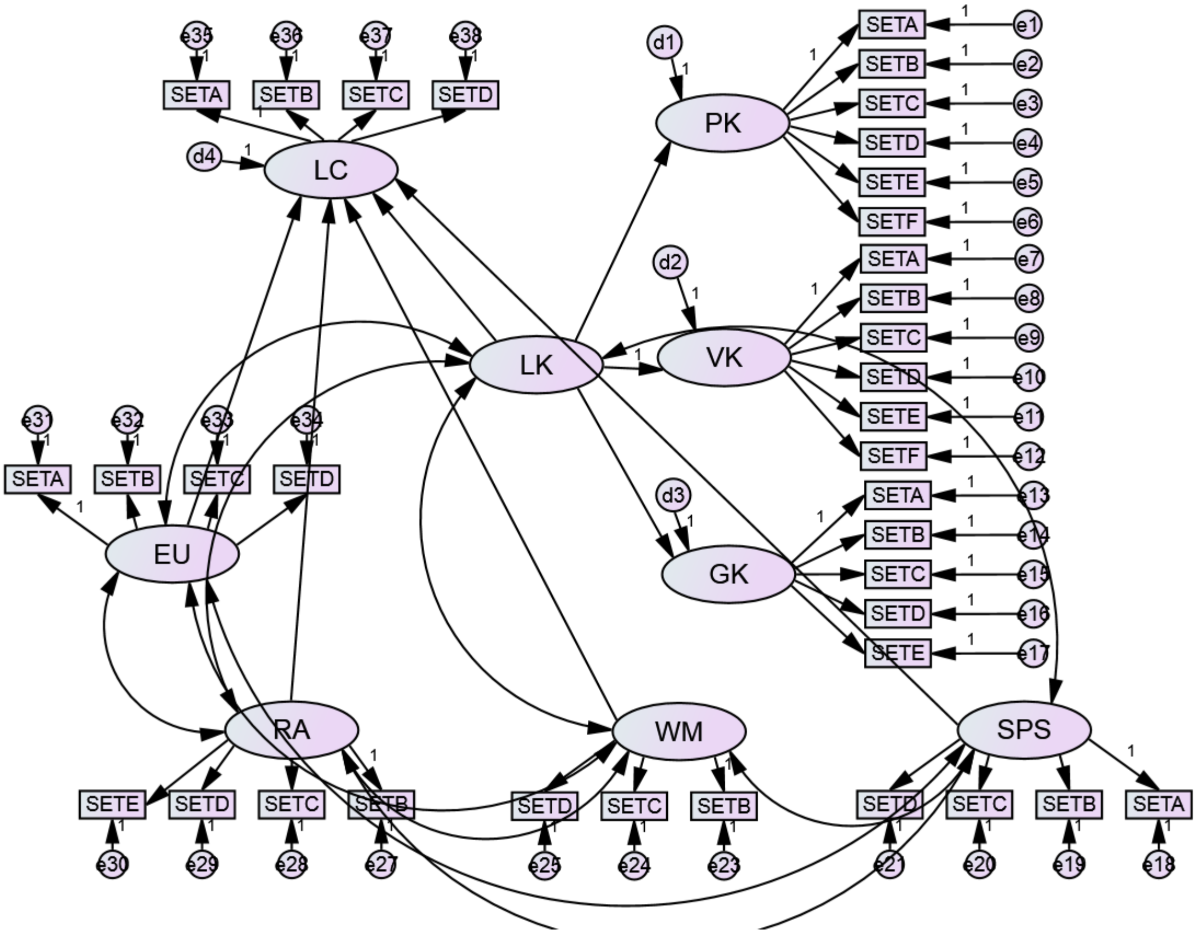


Figure 5.4 A new hypothesized structural equation model of LC as measured with the PET listening

In the hypothesized structural equation model of LC, as measured with the PET listening (Figure 5.4), there are 45 exogenous variables and 40 endogenous variables (see Table 5.13). The exogenous variables included the five latent variables of LK, SP, WM, RA, and EU, the 36 error terms, and the four residuals. The endogenous variables included the 36 observed variables and the four latent variables of LC, PK, VK, and GK.

Table 5.13 Variable counts

Number of variables in the model	85
Number of observed variables	36
Number of unobserved variables	49
Number of exogenous variables	45
Number of endogenous variables	40

The model identification was verified next. The number of observations and parameters was counted (see Tables 5.14 & 5.15). Since there were 576 degrees of freedom, the requirement that the degrees of freedom should be larger than zero was met, and the hypothesized model was overidentified.

Table 5.14 Parameter summary

	Weights	Covariances	Variances	Total
Fixed	49	0	0	49
Unlabelled	35	10	45	90
Total	84	10	45	139

Table 5.15 Computation of degrees of freedom

Number of distinct sample moments	666
Number of distinct parameters to be estimated	90
Degrees of freedom	576

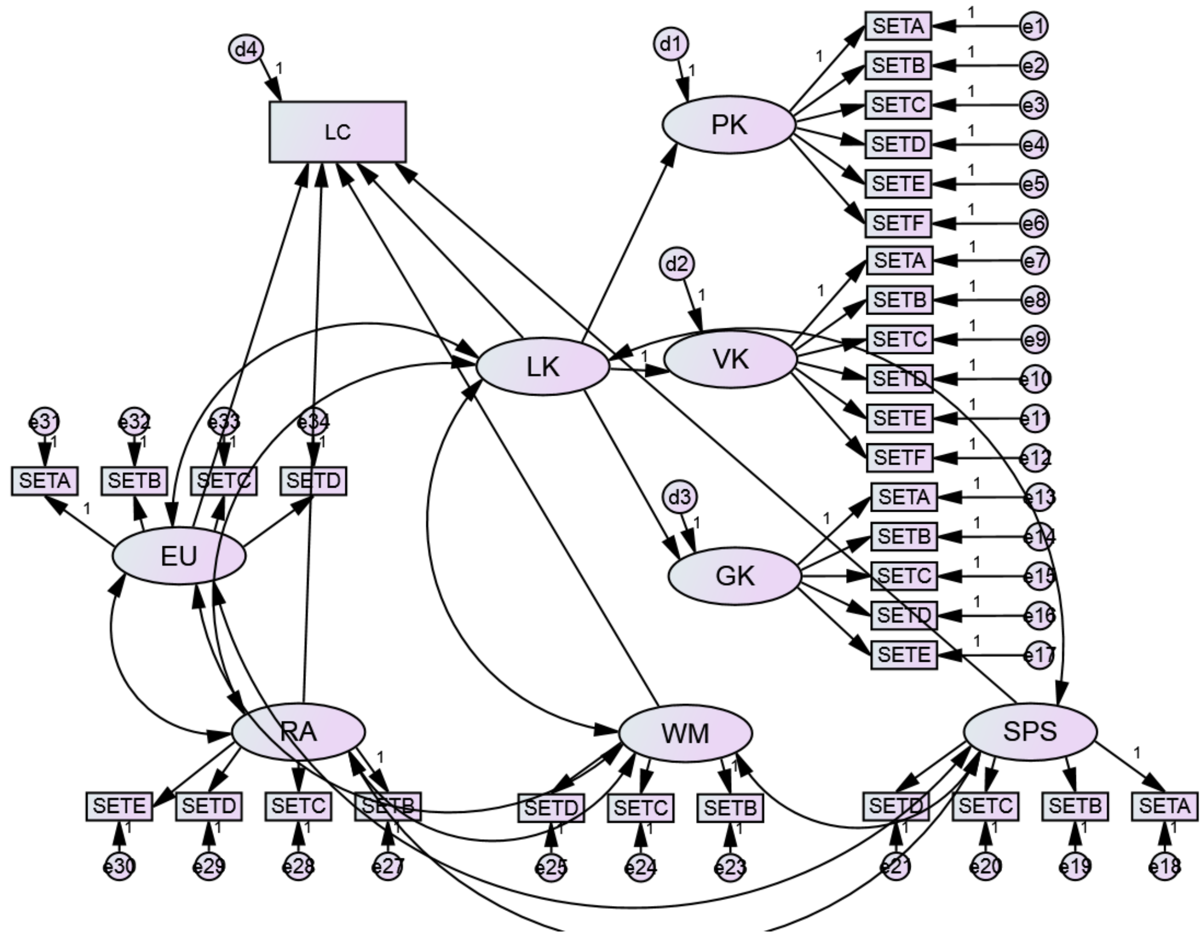


Figure 5.5 A new hypothesized structural equation model of LC as measured with the CET4 listening

In the hypothesized structural equation model of LC as measured with the CET4 listening (Figure 5.5), there are 40 exogenous variables and 36 endogenous variables (see Table 5.16). The exogenous variables included the five latent variables of LK, SP, WM, RA, and EU, the 32 error terms, and the three residuals. The endogenous variables included the 33 observed variables and the three latent variables of PK, VK, and GK.

Table 5.16 Variable counts when LC was measured with the CET4 listening

Number of variables in the model	76
Number of observed variables	33
Number of unobserved variables	43
Number of exogenous variables	40
Number of endogenous variables	36

The model identification was verified next. The number of observations and parameters was counted (see Tables 5.17 & 5.18). Since there were 478 degrees of freedom, the requirement that the degrees of freedom should be larger than zero was met, and the hypothesized model was overidentified.

Table 5.17 Parameter summary

	Weights	Covariances	Variances	Total
Fixed	44	0	0	44
Unlabelled	32	10	41	83
Total	76	10	41	127

Table 5.18 Computation of degrees of freedom

Number of distinct sample moments	561
Number of distinct parameters to be estimated	83
Degrees of freedom	478

Next, the factor loadings between each construct and its items for each measurement model were checked. The measurement models were analysed in AMOS. In each model the value of the greatest loading weight between the construct and the observed variables was constrained to the value of 1. In this way, the largest value for all the factor loadings was unitised to 1, and the other factor loadings would be lower than 1. By doing this, it could be found which variable loaded most strongly onto a construct from the unstandardized regression weights. After that, each measurement model was checked to ascertain whether the observed variables reflected the constructs, as assumed.

The measurement model of LC was checked first as it was the key dependent variable in the study. The model comprised LC and four observed variables (see Figure 5.6). As shown in Appendix 85, the loading weight between item LC3 and the construct was the greatest among the four loading weights. Then, the greatest loading weight was changed to the value of 1 and the loading weights were calculated again (see Appendix 86). The results of the analysis of the model fit indicate it was acceptable (χ^2 = 4.83, p = .089, RMSEA = .087, SRMR = .03, CFI = .98). Based on the results of the analysis, the four observed variables were considered as valid variables to measure LC and they were kept in the model.

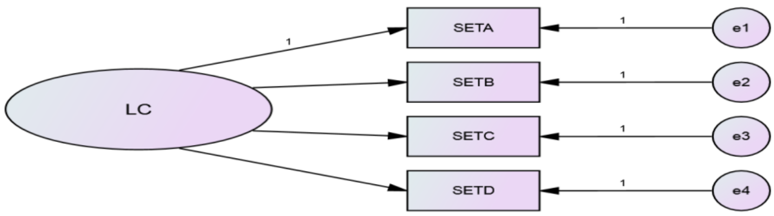


Figure 5.6 The measurement model of LC with four subsets from the PET listening test

The measurement model of PK was analysed second to discover which observed variables should be kept to measure the construct. The model comprised PK and six observed variables (see Figure 5.7).

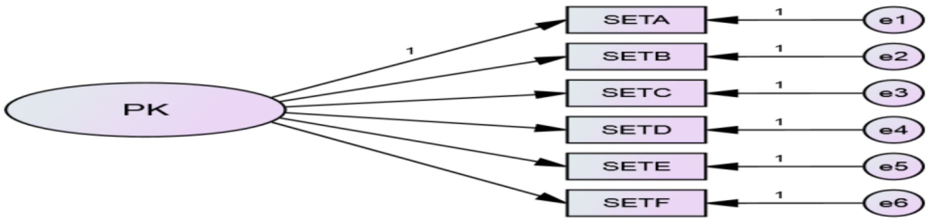


Figure 5.7 The measurement model of PK

As can be seen in Appendix 87, the unstandardized regression weight from PK to PK5 was the greatest among the other regression weights. Therefore, the value of the loading weight between PK and PK5 was changed to 1. Then the model was analysed again and the values of the other regression weights were less than 1 (see Appendix 88). The goodness-of-fit of the measurement model was also tested, but results show that the fit was not good (χ^2 = 76.83, p < .05, RMSEA = .20) (see Appendices 89 to 92). To improve the goodness-of-fit of this measurement model, some changes were made. First, the items which had factor loading weights less than .50 were deleted one by one. Each time an item was deleted the model was analysed in order to find out whether the change made any difference to the model fit. Based on the results of analysis on the loading weights, PK1 was deleted from the model as it had the lowest factor among the six observed variables (factor loading = .29). It was found that the model fit was improved after the change was made (χ^2 = 31.84, p < .05, RMSEA = .17). Then another item, PK2, which loaded poorly onto PK (factor loading

< .50), was deleted. However, the results showed that although the χ^2 reduced from 31.84 to 28.28, the RMSEA increased from .17 to .27 (preferably < .05, acceptable if it is less than .08). Therefore, the item PK2 was kept in the model. Second, since loading weights should ideally be between .50 and .95, items with loading weights greater than .95 were deleted. Since the item of PK5 had a factor loading of .98, it was deleted. The loading weight between PK6 and PK was changed to 1 and the model was analysed again (see Appendix 93). As can be seen in Appendices 94 to 97 the model fit was good (χ^2 = 3.37, p > .05, RMSEA = .06, SRMR = .03, CFI = .99). Therefore, PK2, PK3, PK4, and PK6 were considered to be reflective of PK and were kept in the measurement model.

The measurement model of VK was tested in third place. The measurement model comprised VK and six observed variables (see Figure 5.8). The results showed that the loading weight between VK and VK6 was the greatest among the other observed variables (see Appendix 98). Therefore, the value of the loading weight between VK and VK6 was changed to 1 and the measured model was analysed again in Amos (see Appendix 99). As can be seen in Appendices 100 to 103, the model fit was good (χ^2 = 13.17, p > .05, RMSEA = .05, SRMR = .02, CFI = .99). Therefore, the six subsets were considered as valid variables to measure VK and were kept in the measurement model.

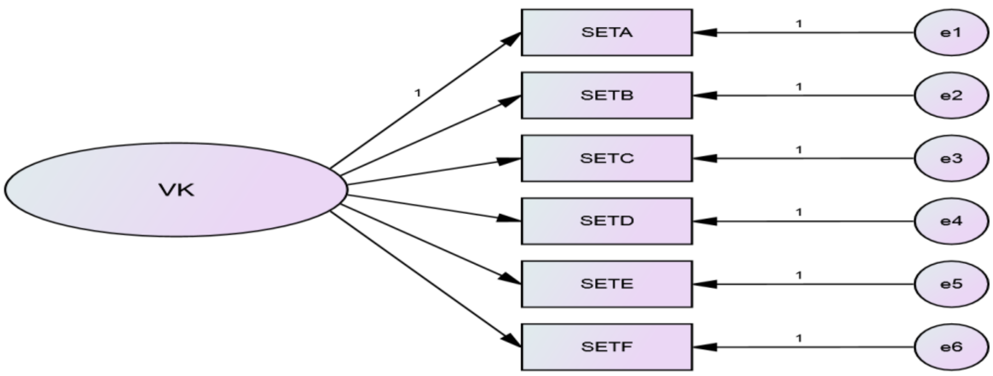


Figure 5.8 The measurement model of VK

The measurement model of GK was tested in fourth place. The model comprised GK and five measured variables (see Figure 5.9). As can be seen in Appendix 104, GK2 had the greatest loading weight with GK among the five measured variables. Therefore, the value of the loading weight of GK2 was changed to 1 and the model was analysed again (see Appendix 105). Appendices 106 to109 indicate that the model fit was not good (χ^2= 69.42, p < .05, RMSEA = .26). Since the loading weight between GK5 and GK was less than .50, GK5 was deleted from the model and the model was analysed again. The results show that new model fit was good (χ^2 = 1.96, p > .05, RMSEA =.00, SRMR = .02, CFI = 1) (see Appendices 110 - 113)

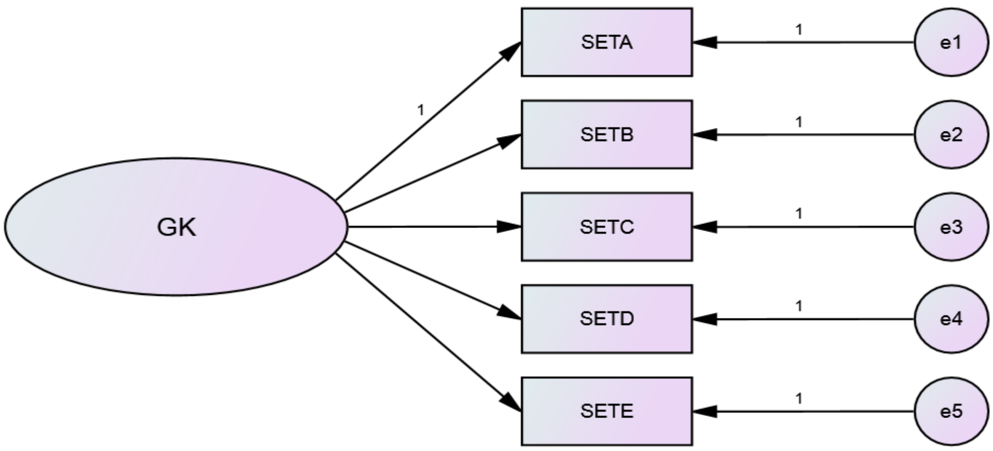


Figure 5.9 The measurement model of GK

The measurement model of SPS was tested in fifth place. The model comprised SPS and four measured variables (see Figure 5.10). As can be seen in Appendix 114, the greatest regression weight among the four regression weights was between SPS and SPS2. Therefore, the factor loading between SPS2 and SPS was changed to 1 and the model was analysed again (see Appendix 115). Appendices 116 to 119 show that the model fit was good (χ^2 = 2.18, p > .05, RMSEA =.02, SRMR = .01, CFI = 1). Therefore, the four items were considered as valid variables to measure SPS and were kept in the model.

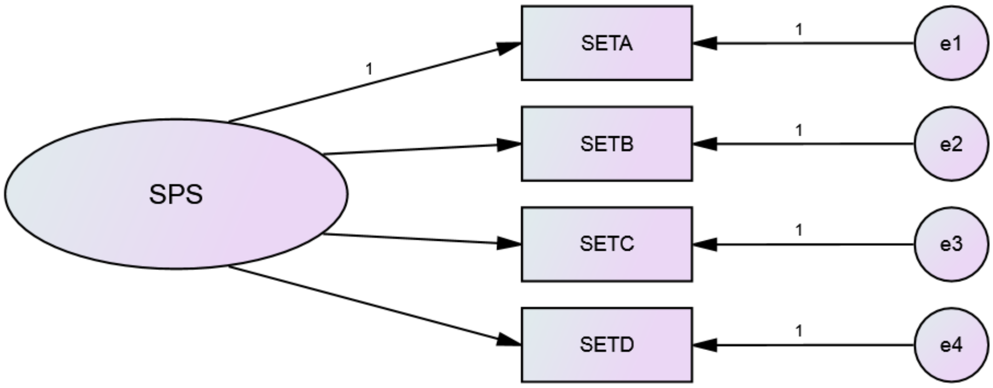


Figure 5.10 The measurement model of SPS

The measurement model of WM was tested in sixth place. The model comprised WM and four measured variables (see Figure 5.11). The results of analysis show that the greatest regression weight was between WM and WM4 (see Appendix 120). Therefore, the regression weight between WM4 and WM was changed to 1 and the model was analysed again (see Appendix 121). Appendices 122 and 125 indicate that the model fit was on the borderline of acceptable fit (χ^2 = 6.17, p = .05, RMSEA = .106, SRMR = .05, CFI = .96). Although the loading weight between WM1 and WM was low (factor loading = .080), WM1 was not deleted from the model. If the item was deleted the model would be a saturated model in which $\chi^2 = 0$, and the other indicators for the model fit would not be calculated. Based on the results of the analysis, the four variables were considered as valid measured variables for WM and were kept in the model.

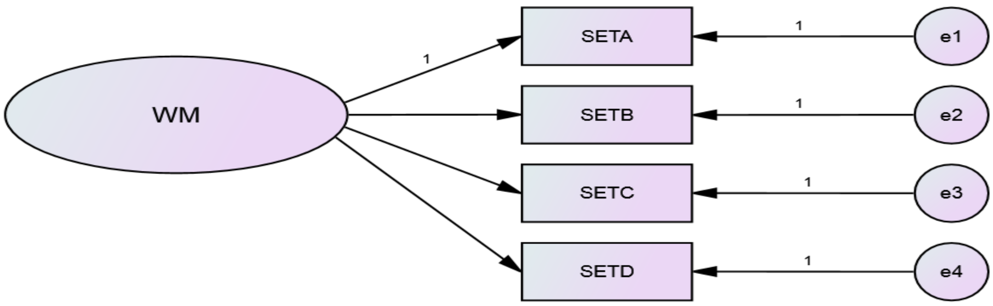


Figure 5.11 The measurement model of WM

The measurement model of RA was tested in seventh place. The model comprised RA and four measured variables (the item of RA1 had been deleted to improve the reliability of the individual construct) (see Figure 5.12). As can be seen in Appendix 126, RA4 had the greatest loading weight among the four loading weights with the construct. Therefore, the loading weight between RA4 and RA was changed to 1 and the model was analysed again (see Appendix 127). Appendices 128 to 131 show that the model fit was acceptable ($\chi^2 = 5.3$, $p > .05$, RMSEA = .095, SRMR = .05, CFI = .96). Based on the results of analysis for the model, the four items were considered as valid variables to measure RA and were kept in the model.

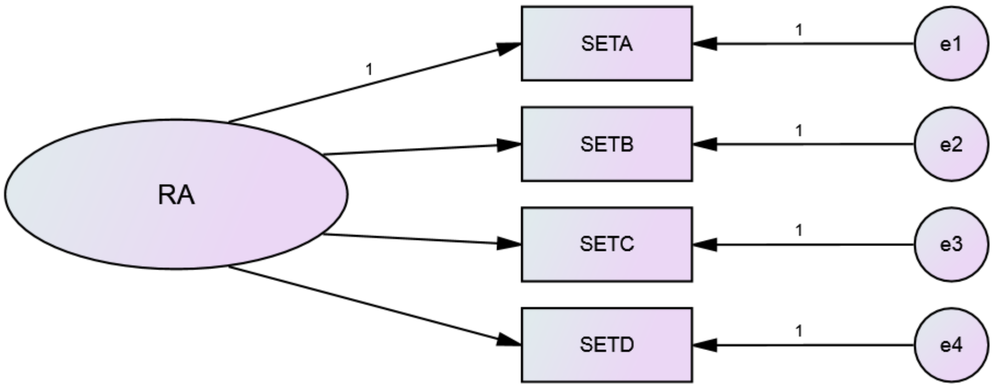


Figure 5.12 The measurement model of RA

The measurement model of EU was tested in eighth place. The model comprised EU and four measured variables (see Figure 5.13). As can be seen in Appendix 132, EU2 had the greatest loading weight with EU among the four observed variables. Therefore, the loading weight between EU2 and EU was changed to 1 and the model was analysed again (see Appendix 133). Appendices 134 to137 indicate that the model fit was good (χ^2 = 1.44, $p > .05$, RMSEA = .00, SRMR = .00, CFI = 1). Based on the results of the analysis, the four observed variables were considered as valid variables to measure EU and were kept in the model.

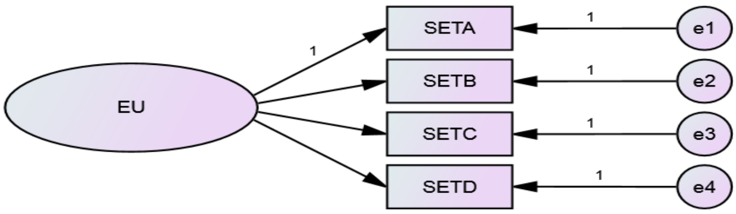


Figure 5.13 The measurement model of the EU

The loading weights between each construct and its observed variables were tested for the models. The eight latent constructs were LC, PK, VK, GK, SPS, WM, RA, and EU. The greatest loading weight between each construct and its observed variables in each measurement model was unitised to the value of 1. The observed variables which were used to measure the latent constructs were found to reflect LC, VK, SPS, WM, RA, and EU and were kept in the measurement models. For the other two latent constructs (PK and GK), two out of six observed variables in PK and one out of five observed variables in GK were deleted from the two models. The goodness-of-fit for each measurement model was tested and the results indicated that the goodness-of-fit of the measurement models was good or acceptable.

Confirmatory factor analysis (CFA) was undertaken to ascertain whether the observed variables were in fact loading onto the constructs. Two different validities were checked: convergent validity, which confirms whether the items load onto a construct; and discriminant validity, which confirms whether each construct is measuring a different aspect.

For the convergent validity, average variance extracted (AVE) (preferably > .50, Wu, 2010) and composite reliability (CR) (preferably > .60, Wu, 2010) were calculated (see Table 5.19). The convergent validity of LC was not calculated because LC was the dependent variable, while others were independent variables. The results indicate that four out of seven constructs (VK, GK, SPS, and EU) had convergent validity, while the other three (PK, WM, and RA) did not.

Table 5.19 The results of AVE and CR of each latent construct

Convergent validity	PK	VK	GK	SPS	WM	RA	EU
AVE (value >.5)	.434	.560	.607	.574	.368	.334	.939
CR (value >.6)	.493	.739	.857	.843	.620	.623	.984
Convergent validity*	NE	E	E	E	NE	NE	E

* NE = Not established; E = Established

Since convergent validity was violated for each of PK, WM, and RA, changes were made to these three latent constructs in order to improve convergent validity. Based on the results of analyses of the measurement model for PK, since PK2 had factor loading weights of .327 with PK, lower than the preferable value of .50, the item was deleted from the measured model. The CR and the AVE were recalculated. The results showed that the CR was improved from .493 to .781 and the AVE was improved from .434 to .544. For WM, since the loading weight between WM1 and WM was much lower than .50 (factor loading = .099), WM1 was deleted from the measurement model. Then the CR and the AVE of WM were recalculated. The results showed that the new CR was .733 and the new AVE was .537, which means that both were improved. For RA, since RA2 had lower loading weight with RA (factor loading = .234), RA2 was deleted from the measured model. The new CR was .683 and the new AVE was .458 which was on the borderline of the preferable value of

AVE (preferably > .50, Wu, 2010). The results of reanalyses of CR and AVE values can be found in Table 5.20. The results indicate that six out of seven constructs did not violate the assumption of convergent validity. The convergent of RA was considered acceptable because the value of AVE (.458) was just below the preferable value of .50. Therefore, convergent validity of each construct was confirmed.

Table 5.20 The results of AVE and CR of each construct with changes

Convergent validity	PK	VK	GK	SPS	WM	RA	EU
AVE (value > .50)	.544	.560	.607	.574	.537	.458	.939
CR (value > .60)	.781	.739	.857	.843	.733	.683	.984
Convergent validity*	E	E	E	E	E	NE (but accept-able)	E

* E = established; NE = Not established

To ascertain whether the latent variables measured different constructs, discriminant validity was verified by pairwise comparisons between the squared correlations and AVE scores for each construct (see Table 5.21). As can be seen in Table 5.21, 18 out of 21 pairs of constructs did not violate the assumption of discriminant validity, which means the 18 different latent variables measured different constructs because each construct was compared against another in the model. When PK was compared against VK, the two constructs showed a very high correlation ($r > .80$). When PK was compared against GK, the two constructs also showed a very high correlation ($r > .80$). High correlation between two different constructs violated the assumption of discriminant validity. This might be due to the assumption of a common construct (LK), which included the constructs of PK, VK, and GK, as assumed in the hypothesized structural equation model of LC (Figures 5.4 & 5.5). LK was checked in confirmatory factor analysis (CFA) and is reported in the next section. PK and EU strongly correlated with each other (*r* > .70). However, since PK is subsumed under LK, the strong correlation between PK and EU would not have much direct impact on the model as they are two different constructs. In addition, reported in the next section, the discriminant validity of LK and EU was verified. The new results indicate that LK and EU were different constructs.

Table 5.21 The results of discriminant validity

Discriminant validity	Factor correlation	Correlation squared	AVE1	AVE2	Discriminant validity*
			AVEs should be > r^2		
PK <--> SPS	.003	.000	.544	.574	E
PK <-->WM	.515	.265	.544	.537	E
PK<-->RA	.130	.017	.544	.458	E
PK<-->GK	.801	.642	.544	.607	NE
PK<-->VK	.887	.787	.544	.560	NE
PK<-->EU	.764	.584	.544	.939	NE
VK<-->SPS	-.076	.006	.560	.574	E
VK<-->WM	.357	.127	.560	.537	E
VK<-->RA	.094	.009	.560	.458	E
VK<-->GK	.696	.484	.560	.607	E
VK<-->EU	.651	.424	.560	.939	E
GK<-->SPS	-.084	.007	.607	.574	E

GK<-->WM	.434	.188	.607	.537	E
GK<-->RA	.259	.067	.607	.458	E
GK<-->EU	.478	.228	.607	.939	E
SPS<-->WM	.124	.015	.574	.537	E
SPS<-->RA	.116	.013	.574	.458	E
SPS<-->EU	.140	.020	.574	.939	E
WM<-->RA	.111	.012	.537	.458	E
WM<-->EU	.466	.217	.537	.939	E
RA<-->EU	.157	.025	.458	.939	E

* E = Established; NE = Not established

First-order CFA was conducted among PK, VK, and GK to test whether each construct had acceptable convergent validity and whether each pair of constructs had acceptable discriminant validity. Firstly, CR and AVE of each construct were tested in order to ascertain whether observed variables in each construct could measure each underlying latent variable. As can be seen in Table 5.22, the convergent validity of each construct was confirmed. Second, discriminant validity was checked between each pair of constructs. As can be seen in Table 5.23, the assumptions of discriminant validity between PK and VK and between PK and GK were not established. The absence of discriminant validity might be due to a higher-order common construct (see also the hypothesized structural equation model of LC in Figures 5.4 & 5.5). Therefore, a second-order CFA was conducted next in order to ascertain whether LK could be reflected in PK, VK, and GK. A second order CFA is applied in order to confirm that a higher-order latent construct loads into underlying sub-constructs (Awang, 2012). The results showed that the standard regression weight between LK and PK was greater than 1 (factor loading = 1.023), which was considered unreasonable. After checking PK and its three observed variables, it was found that the problem was caused by the great regression weight between PK6 and PK (factor loading = .95). Therefore, PK6 was deleted from the model and the second-order model was reanalysed. Based on the results of the reanalyses, CR and AVE were calculated to verify the underlying latent variable of LK. As can be seen in Table 5.24, LK was verified to be the common latent construct which can represent the three latent constructs of PK, VK, and GK. As can be seen in Appendices 138 to 141 the model fit was good (χ^2 = 100.971, p < .05, χ^2/DF = 1.98, RMSEA = .073).

Table 5.22 The results of CR and AVE from first-order CFA

Measured variables	Loading weights	R squared	Error variances	CR (preferable >.60)	AVE (preferable >.50)
PK3	.547	.299	.701		
PK4	.593	.352	.648		
PK6	.948	.899	.101		
				.750	.517
VK1	.50	.25	.75		
VK2	.824	.679	.321		
VK3	.798	.637	.363		
VK4	.782	.611	.388		
VK5	.708	.501	.499		
VK6	.825	.681	.319		
				.882	.560
GK1	.826	.682	.318		
GK2	.873	.762	.238		
GK3	.824	.679	.321		
GK4	.565	.319	.681		
				.86	.611

Table 5.23 The results of analysis on discriminant validity in first-order CFA

Discriminant validity	Factor correlation	Correlation squared	AVE1 (AVEs should be > r^2)	AVE2	Discriminant validity*
VK <--> GK	.693	.480	.560	.611	E
PK <-->GK	.81	.656	.517	.611	NE
PK <--> VK	.896	.803	.517	.560	NE

* E = Established; NE = Not established

Table 5.24 The results of analysis on CR and AVE in second-order CFA

Measured variables/construct	Loading weights	R squared	Error variances	CR (preferable >.60)	AVE (preferable >.50)
PK3	.728	.530	.470		
PK4	.775	.601	.399		
				.722	.565
VK1	.513	.263	.737		
VK2	.823	.677	.323		
VK3	.803	.645	.355		
VK4	.783	.613	.387		
VK5	.711	.506	.494		
VK6	.814	.663	.337		
				.883	.561

GK1	.827	.684	.316		
GK2	.879	.773	.227		
GK3	.818	.669	.331		
GK4	.562	.316	.684		
				.859	.610
PK	.762	.581	.419		
VK	.795	.632	.368		
GK	.870	.757	.243		
				.851	.657

Since LK was found to be the underlying latent variable which can represent PK, VK, and GK, it was important to check the discriminant validity of each pair of the latent constructs which were assumed to contribute to LC, namely: LK, SPS, WM, RA, and EU. CR and AVE of each construct were calculated first (see Table 5.25).

Table 5.25 Convergent validity for the five latent constructs assumed to explain LC

Convergent validity	LK	SPS	WM	RA	EU
AVE (value>.50)	.647	.574	.443	.421	.939
CR (value>.60)	.845	.844	.688	.666	.984
Convergent validity*	E	E	NE	NE	E

* E = Established; NE = Not established

As can be seen in Table 5.25, LK, SPS, and EU had separate convergent validity, which means that the observed variables in each construct reflected the construct that the observed variables measured. However, WM and RA did not have separate convergent validity. Changes were made to improve AVEs of the two constructs by eliminating observed variables which had lower factor loadings onto each construct (factor loading < .50). For WM, WM3 was deleted because the variable had the lowest loading weight onto WM (factor loading = .467). For RA, RA3 had a loading weight of .48 and was deleted for the same reason. Then AVE and CR of the constructs were recalculated (see Table 5.26). The convergent validity of WM was verified, but not for RA.

The discriminant validity between each pair of the five latent constructs was tested. The correlation between each pair of constructs was squared and the R squared of each pair was compared with AVEs of the two constructs involved. As can be seen in Table 5.27, the assumptions of discriminant validity of the five constructs were confirmed. The results of analysis of the model fit showed that χ^2 = 380.042, p < .05, χ^2/DF = 1.59, RMSEA =.056, which means that the overall model fit was good.

Table 5.26 The new results of convergent validity of the five latent constructs

Convergent validity	LK	SPS	WM	RA	EU
AVE (value>.5)	.646	.574	.569	.460	.939
CR (value>.6)	.845	.843	.715	.626	.984
Convergent validity*	E	E	E	NE	E

* E = Established; NE = Not established

Table 5.27 Discriminant validity of each pair of the five latent constructs

Discriminant validity	Factor correlation	Correlation squared	AVE1 AVE2 (AVEs should be > r^2)		Discriminant validity*
RA <--> WM	.100	.01	.460	.569	E
SPS <-->WM	.137	.019	.574	.569	E
SPS <-->RA	.158	.025	.574	.460	E
SPS <-->EU	.139	.019	.574	.939	E
SPS <-->LK	-.100	.01	.574	.646	E
EU <-->WM	.458	.21	.939	.569	E
WM <-->LK	.447	.200	.569	.646	E
RA <-->EU	.191	.036	.460	.939	E
RA <-->LK	.159	.025	.460	.646	E
EU <-->LK	.694	.482	.939	.646	E

* E = Established

To summarise, the observed variables were verified as reflective of the latent constructs in first-order and second-order models, except that the value of convergent validity for RA was on the borderline (see Table 5.26 for details). LK was the underlying latent variable of PK, VK, and GK. The assumptions of discriminant validity for each pair of latent constructs were verified. Based on the results of the analyses, the overall model is valid. Table 5.28 provides an overview of the latent constructs and the observed variables which were retained in the model.

Table 5.28 Overview of latent constructs and measured variables retained in the model

Latent construct	Observed variables	Number of measured variables	Notes
PK	PK3, PK4	2	
VK	VK1, VK2, VK3, VK4, VK5, VK6	6	
GK	GK1, GK2, GK3, GK4	4	
LK			Comprising PK, VK and GK
SPS	SPS1, SPS2, SPS3, SPS4	4	
WM	WM2, WM4	2	
RA	RA4, RA5	2	
EU	EU1, EU2, EU3, EU4	4	
Total		24	

Next, unit loading identification (ULI) of the entire measurement model was verified to ascertain whether the entire measurement model was a good fit. First, the item in each measurement model which had the greatest loading weight with a construct was constrained to the value of 1. Then the model fit was tested. As can be seen in Appendices 142 to 145, the overall model fit was good (χ^2 = 380.042, $p < .05$, χ^2/DF = 1.59, RMSEA =.056), which means the entire measurement model fits the data.

5.4 Testing the structural equation model with the PET listening test

5.4.1 Testing the structural model fit

The structural model fit was tested to ascertain whether the data for the study were suitable for the hypothesized structural equation model of LC (see Figure 5.14).

Firstly, LC with four measured variables was added to the second-order measurement model. The new hypothesized structural model of LC was formed by adding single-headed arrows from the five latent constructs of LK, SPS, WM, RA, and EU to LC (see Figure 5.14). The addition of the single-headed arrows reflected the assumed associations between each latent construct and LC. The data were put into the model (see Figures 5.15 & 5.16). Then, the residuals of observed variables were tested to ascertain whether there were discrepancies between the hypothesized model and the estimated model. The preferable value for each standard residual covariance was 2.58 (on an absolute scale) (Wu, 2010); only the residual covariance, between VK1 and GK3, was found to be greater than this. Therefore, it was considered that not many assumptions of residuals were violated and the hypothesized model and the estimated model contained no discrepancies. Next, the significance of actual path estimates between the five latent constructs and LC were tested to ascertain whether the associations between the constructs were significant. As can be seen in Appendix 146, the *p*-value of regression weight between EU and LC was significant, and the *p*-value of regression weight between LK and LC was significant, which means EU and LK were predictors of LC. However, the *p*-values of regression weight between LC and the other three constructs were not significant which means the three latent variables were not key predictors of LC. Based on the modification indices (MI), the regressions from PK3 to SPS2 and from RA to SPS1 (the most important two changes to be made on regression weights) were added to the structural model. However, the changes were not very effective, i.e., the significance of the estimates of regression weight between LC and the three constructs SPS, WM, and RA was not changed. The model was not significantly improved with this change. Therefore, it was decided not to make any changes to the model and the insignificant path coefficients were kept in the entire structural model.

The results of estimates can be seen in Appendices 147 to 150. The correlations between each of the five latent constructs and LC were analysed (see Figure 5.17). Results of the correlations between each of the five latent constructs and LC and the standardized regression weights of each latent factor onto LC are presented in Table 5.29 (see Appendices 147 & 151). The model explained 65.7% of the variance in LC. The model fit was good (χ^2 = 490.92, p < .05, χ^2/DF = 1.479, RMSEA = .051, SRMR = .05, CFI = .95) (see Appendices 152 -155 for details).

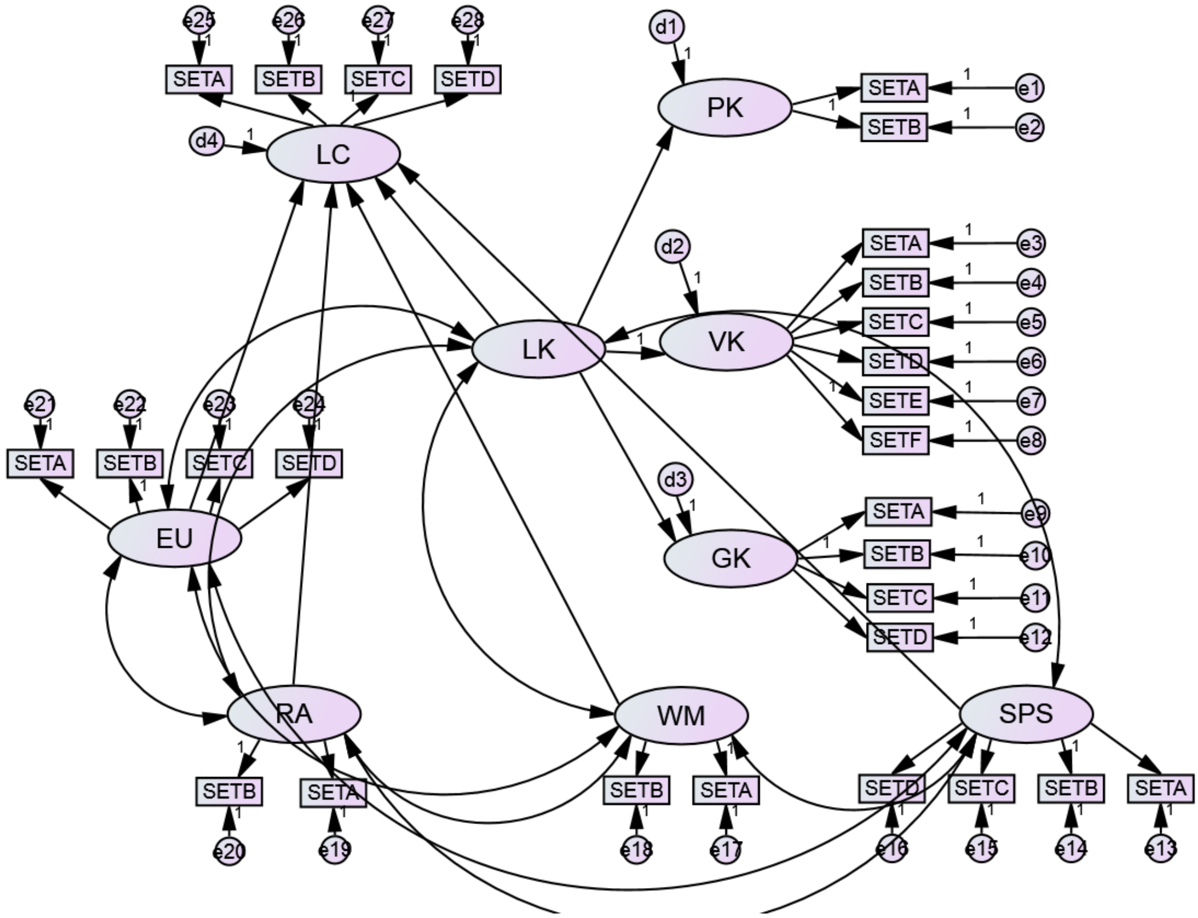


Figure 5.14 The new hypothesized structural equation model of LC

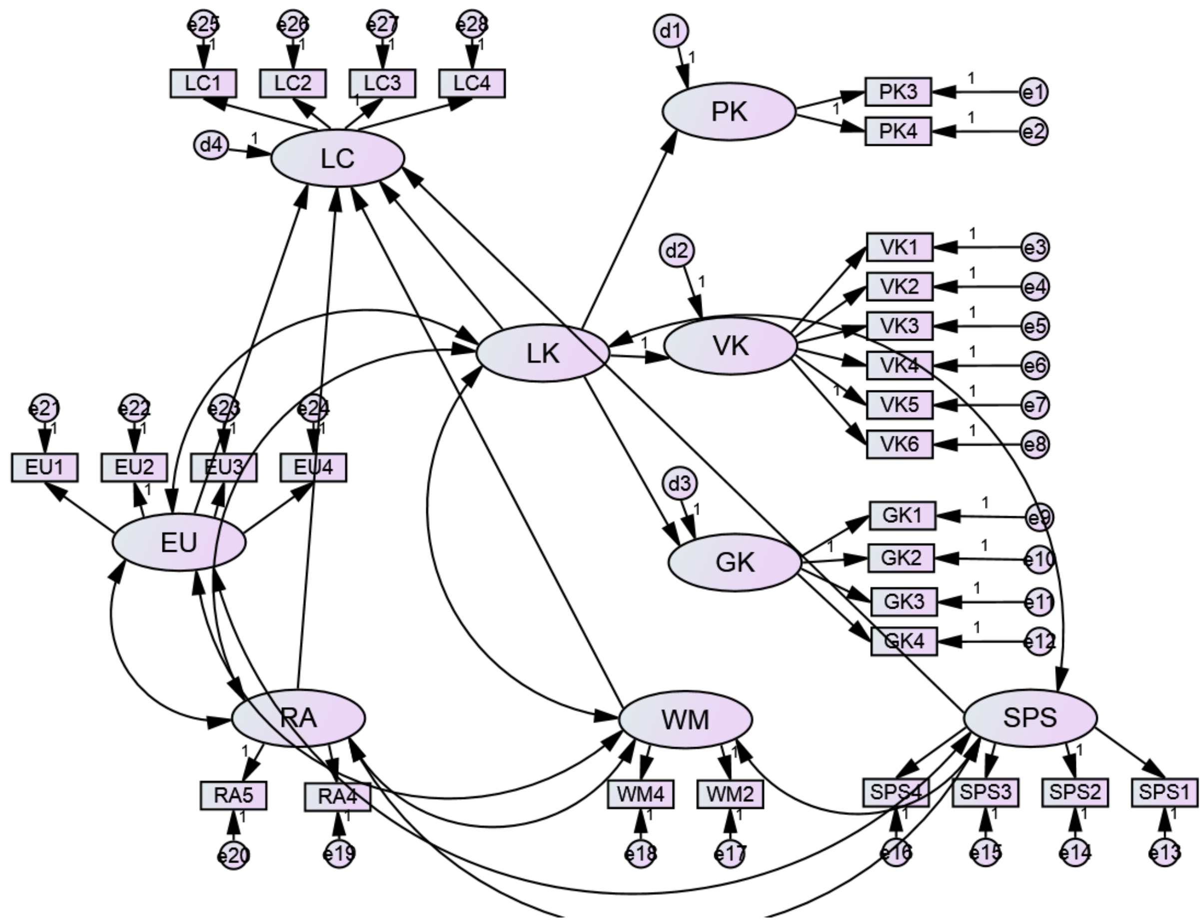


Figure 5.15 The hypothesized structural equation model of LC with data

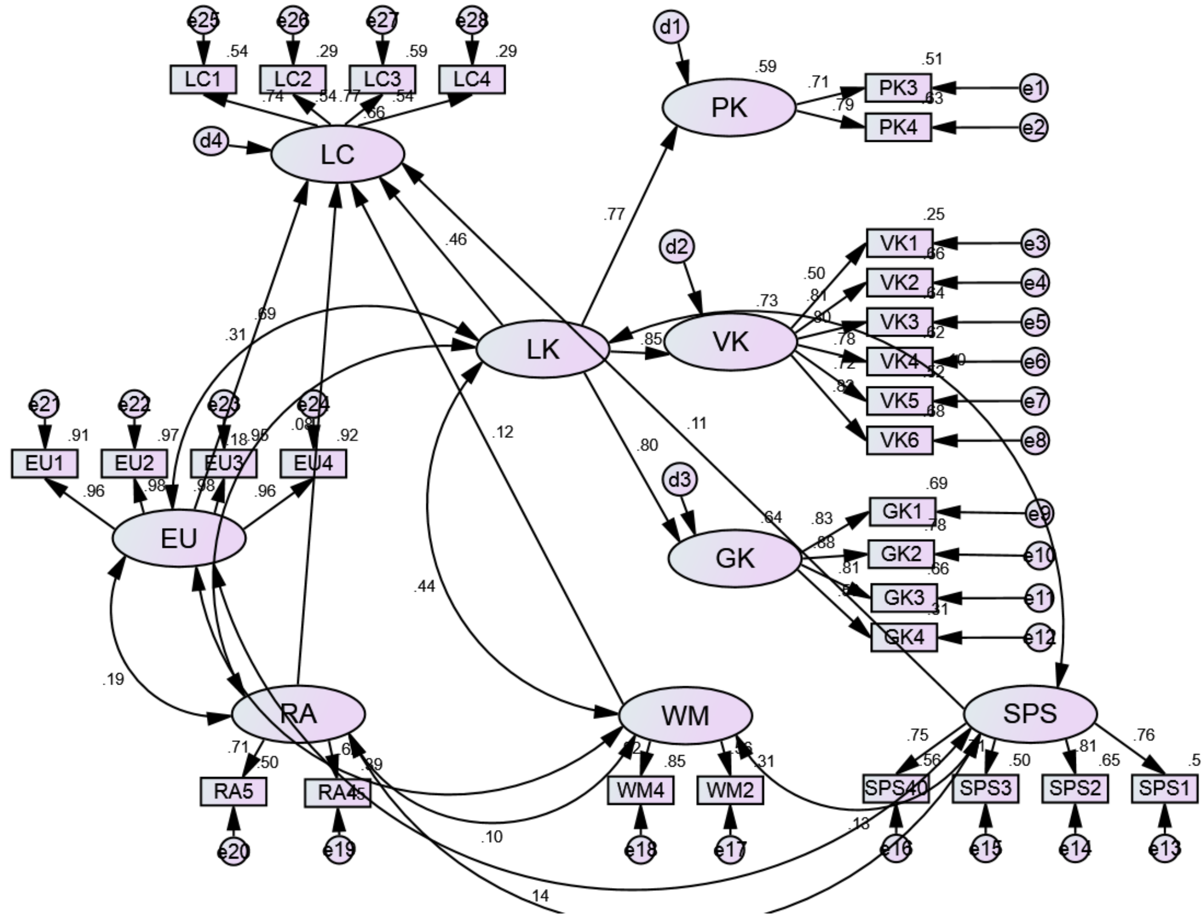


Figure 5.16 The hypothesized structural equation model of LC

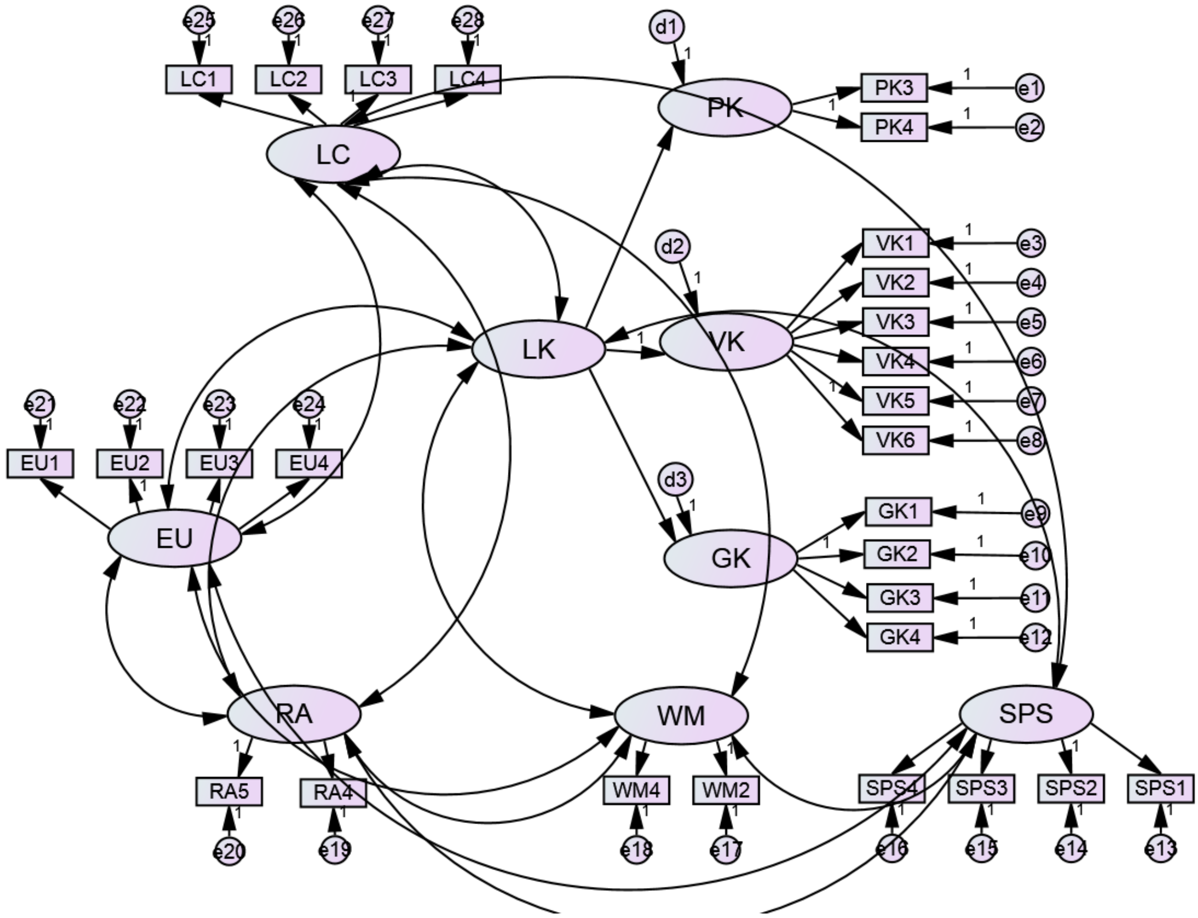


Figure 5.17 The structural model of LC with all correlations between each latent factor and LC

Table 5.29 Implied correlations and standardized regression weights of the latent factors with LC in the structural model of LC

Predictors	Correlations	Stand. Regression weights
LK	.734	.464
SPS	.139	.114
WM	.492	.125
RA	.250	.077
EU	.716	.308
Variance explained		65.7%

5.4.2 Model evaluation

The structural model of LC was tested with composite scale indicators to ascertain whether the structural model was plausible. First, the CRs of the latent constructs were calculated (see Table 5.30). The CRs of these constructs, except for LC, were recalculated as LC was added into the structural model, which might lead to changes in the CRs previously calculated. Next, the structural model was turned into a composite scale model (see Figure 5.18). Differing from the hypothesized structural model, the composite scale model was formed with only one item per latent construct. The one item was the mean of all items in each measurement model. The new variables replaced the old variables in the composite scale model (see Figure 5.19). Then, the factor loadings and error variances of each of these latent constructs on the composite scale indicators were calculated. The following formulas were used for the calculation: Factor loading = root CR; and, error variance = 1 - CR. The results of this calculation can be seen in Table 5.31. The calculated factor loadings and error variances were put into the composite scale model. Since the regression weight from GK onto LK was greater than that from VK onto LK or from PK onto LK, the factor loading from GK onto LK was set to 1 (see Figure 5.20). The composite scale model was analysed in Amos (see Figure 5.21).

Table 5.30 The results of analysis on CR of each latent construct

Convergent validity	LC	LK	PK	VK	GK	SPS	WM	RA	EU
CR	.746	.847	.724	.882	.859	.843	.724	.616	.984

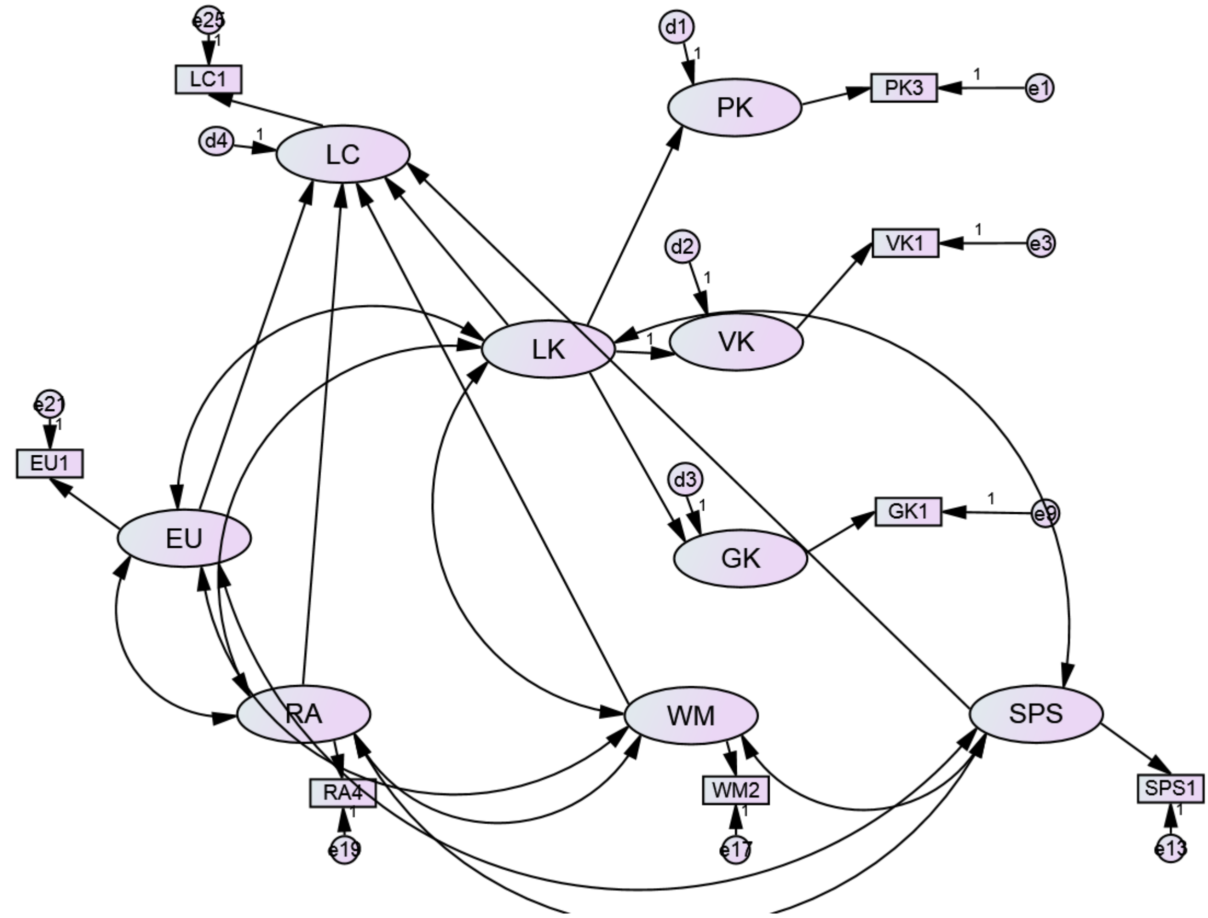


Figure 5.18 The composite scale model of LC

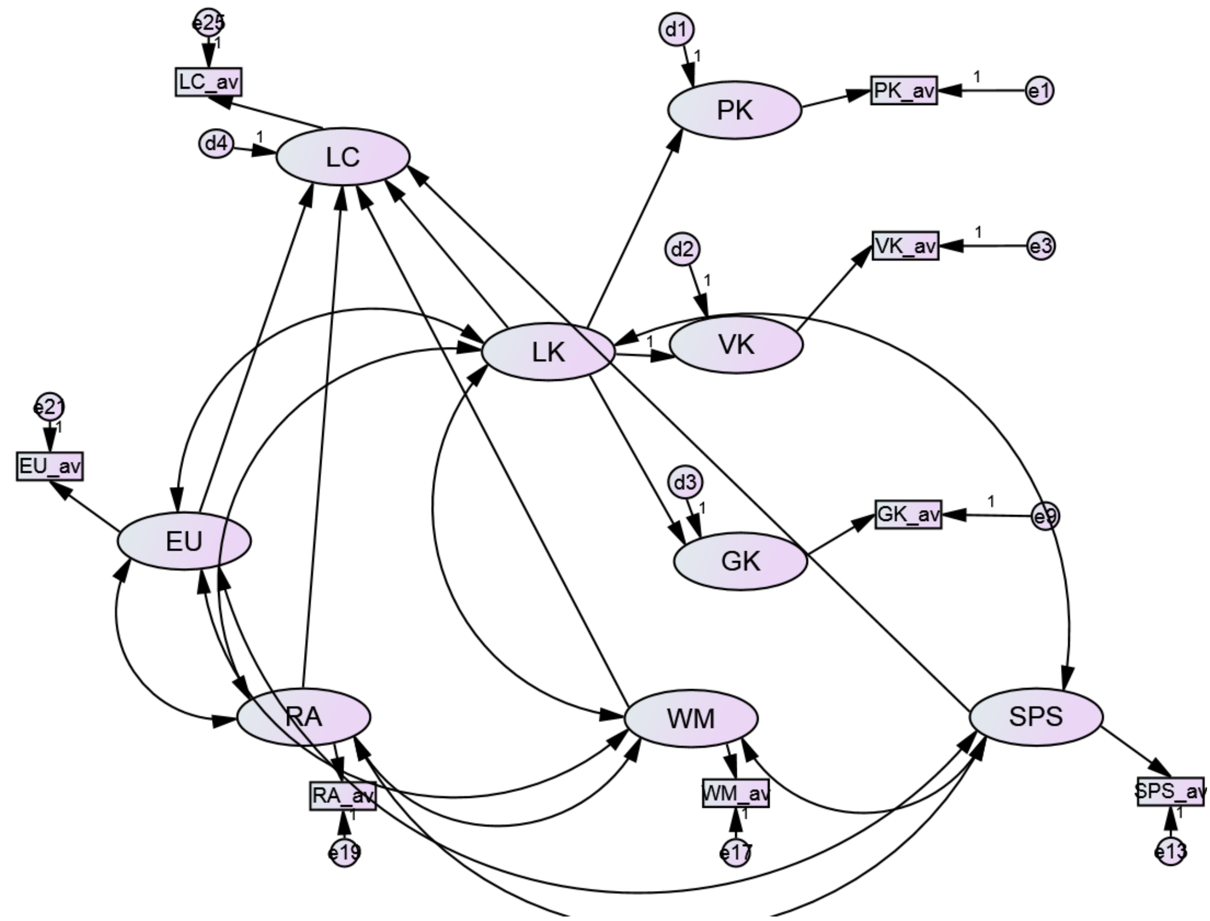


Figure 5.19 The composite scale model of LC with data

Table 5.31 The results of analyses of factor loadings and error variances of each latent construct

Convergent validity	LC	LK	PK	VK	GK	SPS	WM	RA	EU
CR	.746	.847	.724	.882	.859	.843	.724	.616	.984
Factor loading (root CR)	.864	.910	.851	.939	.927	.911	.851	.785	.992
Error variance (1-CR)	.254	.153	.276	.118	.141	.157	.276	.384	.016

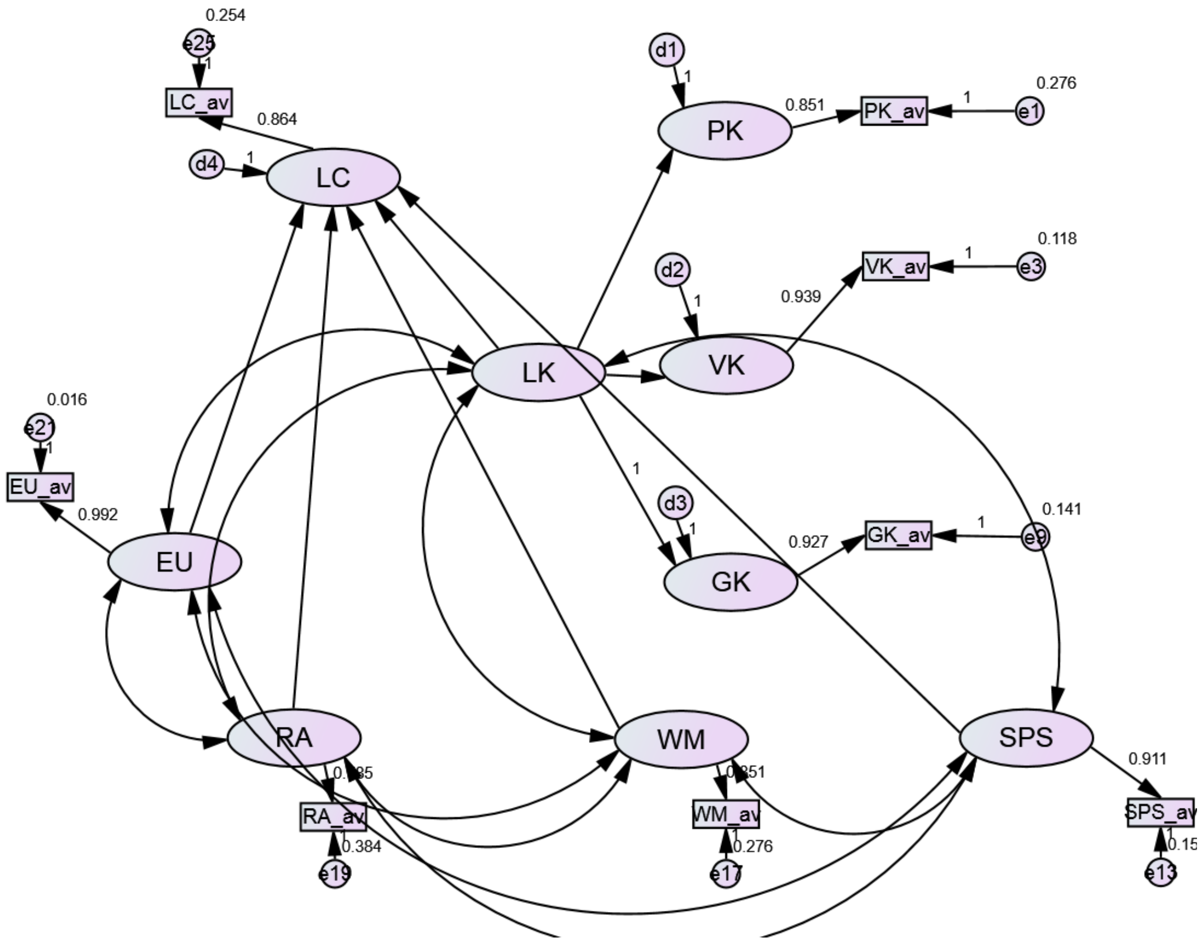


Figure 5.20 The composite scale model of LC with factor loadings and error variances for each latent construct

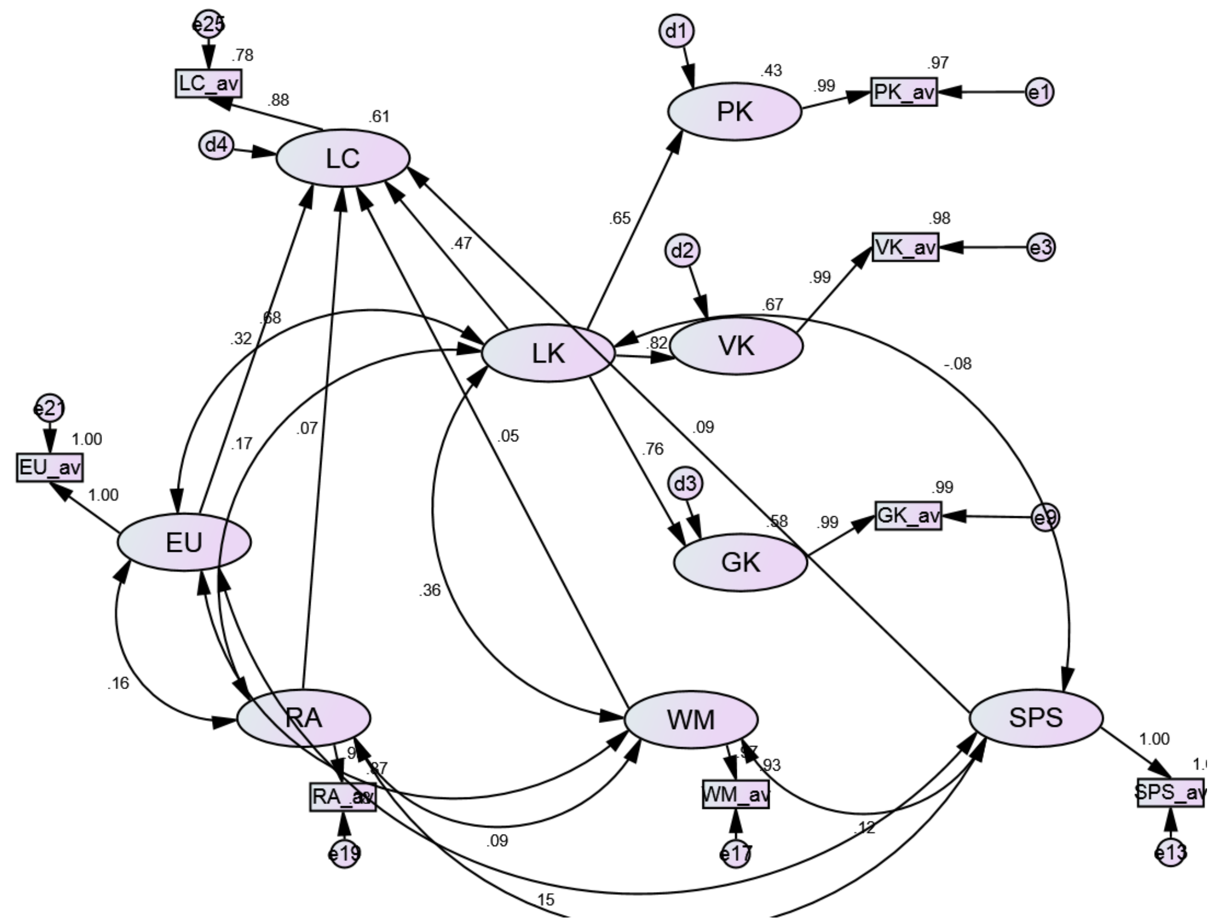


Figure 5.21 The composite scale model of LC with the results of analysis in Amos

The results show that the model fit was not good (χ^2 = 32.061, p < .05, χ^2/DF = 3.206, RMSEA = .109). As can be seen in Appendix 156, three of five loading paths from the latent constructs to LC remained not significant. However, Appendix 157 on standardised residual covariances shows that all the covariances were less than 2.58, and this indicates that the model was accurate in examining the associations between the constructs. MI was checked to ascertain whether any suggestions were given for significant changes between constructs to be made. Based on the suggestion from MI, two loading paths (from EU to GK and from RA to GK) were added to the model; the model fit changed significantly (see Figure 5.22) to become very good after these changes (χ^2 = 7.034, p > .05, χ^2/DF = .879, RMSEA = .000, CFI = 1.00) (see Appendices 158 -161).

It can be concluded, from the results of the evaluation of the structural model with composite scale indicators, that the hypothesized structural model is an accurate model. The overall model fit is good (χ^2 = 490.92, χ^2/DF = 1.479, RMSEA = .051, SRMR = .054) except that the model is significant. Therefore, further steps were taken to modify the structural model based on the suggestions from MI.

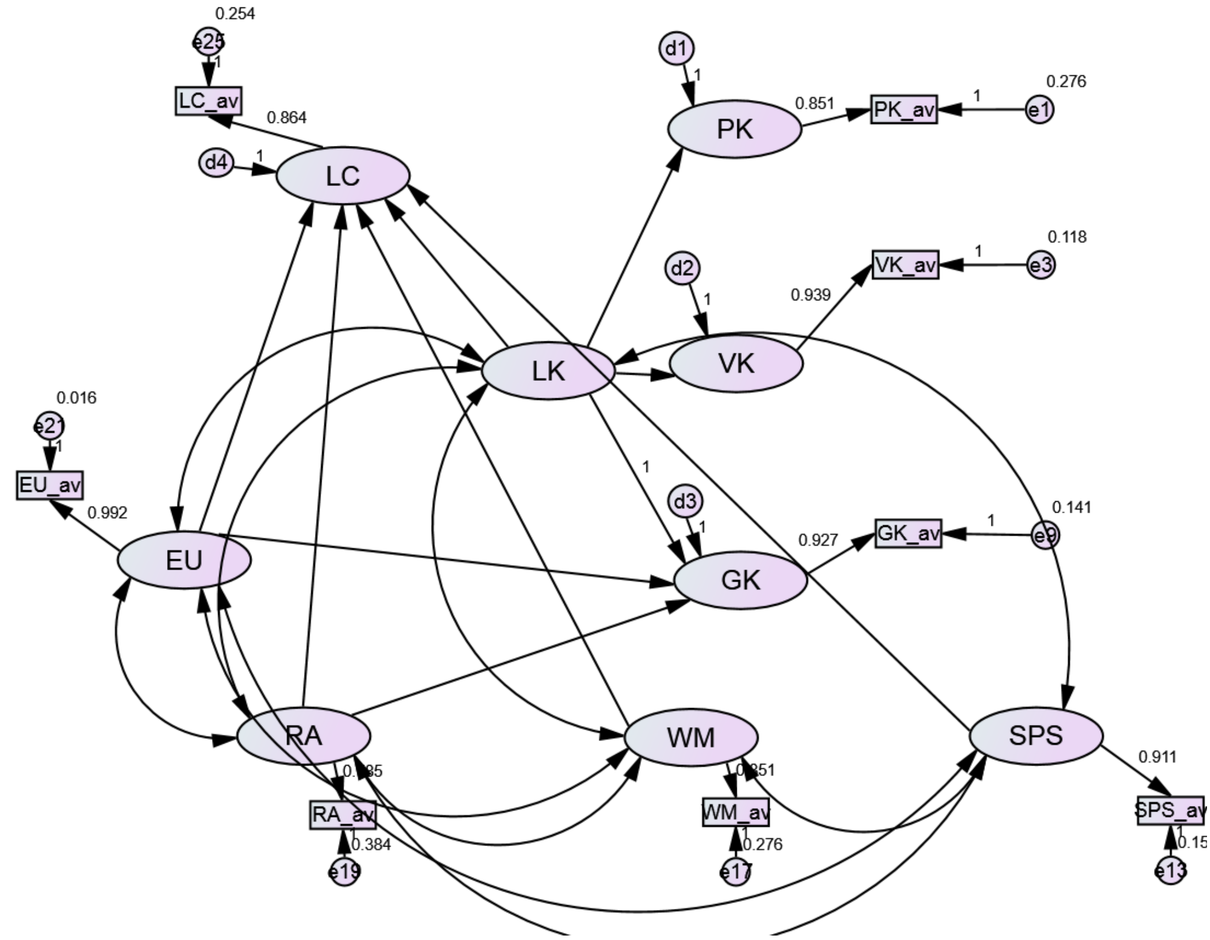


Figure 5.22 The composite scale model of LC with two more loading paths from EU and RA to GK

5.4.3 Model modification

Model modification was tried to improve the model fit based on MI (see Appendix 162). First, the loading path from PK3 to SPS2 was added to the model (see Figure 5.23). It was found that the model was not significantly improved with this change (χ^2 = 472.957, p < .05, χ^2/DF = 1.429, RMSEA = .048, CFI = .958) (see Appendices 163-166). Based on the new MI, the loading path from RA5 to VK6 was added to the model (see Appendix 167 & Figure 5.24). The results indicated that the model was not significantly improved with this change either (χ^2 = 459.208, p < .05, χ^2/DF = 1.392, RMSEA = .046, CFI = .962) (see Appendices 168 -171). Since the overall model was not significantly improved with the modification, I decided not to make changes to the entire model.

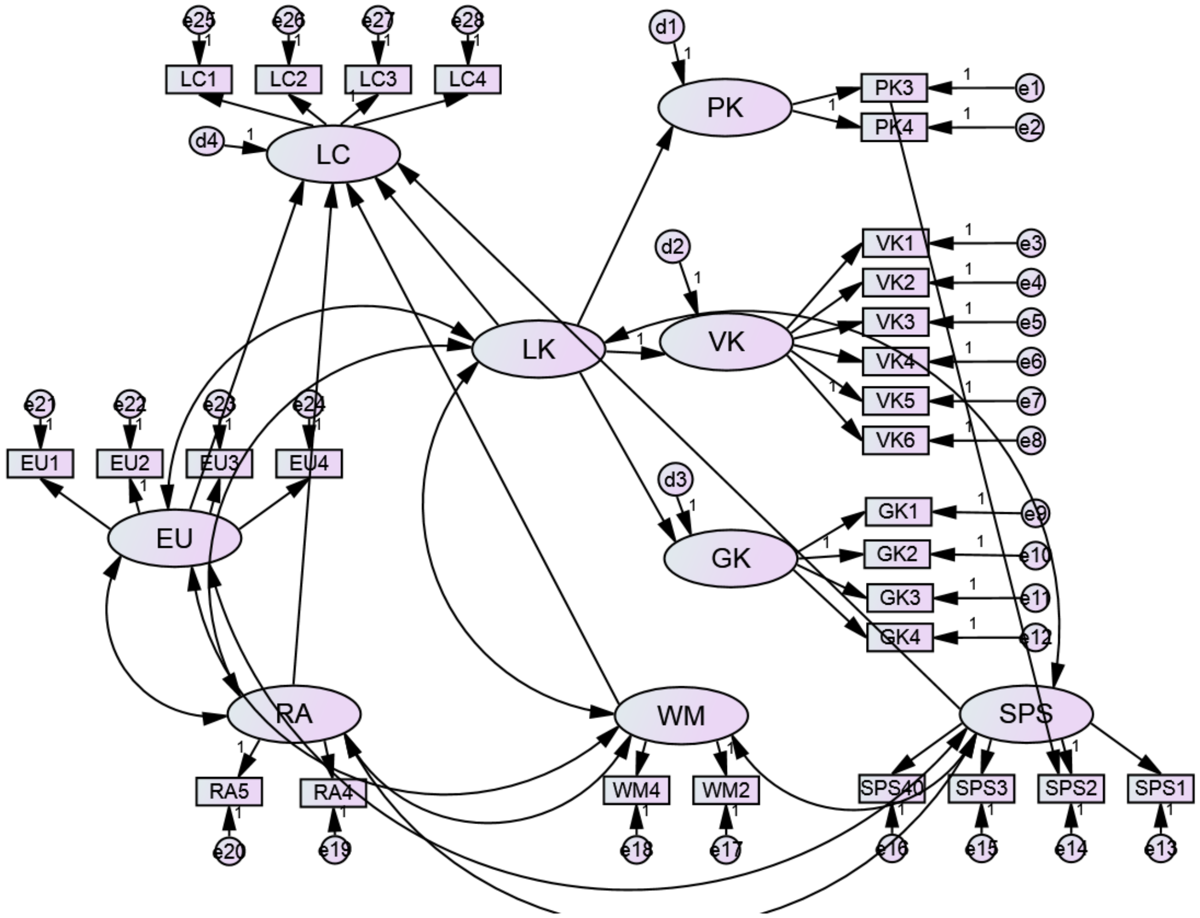


Figure 5.23 The structural mode of LC with the loading path from PK3 to SPS2 added

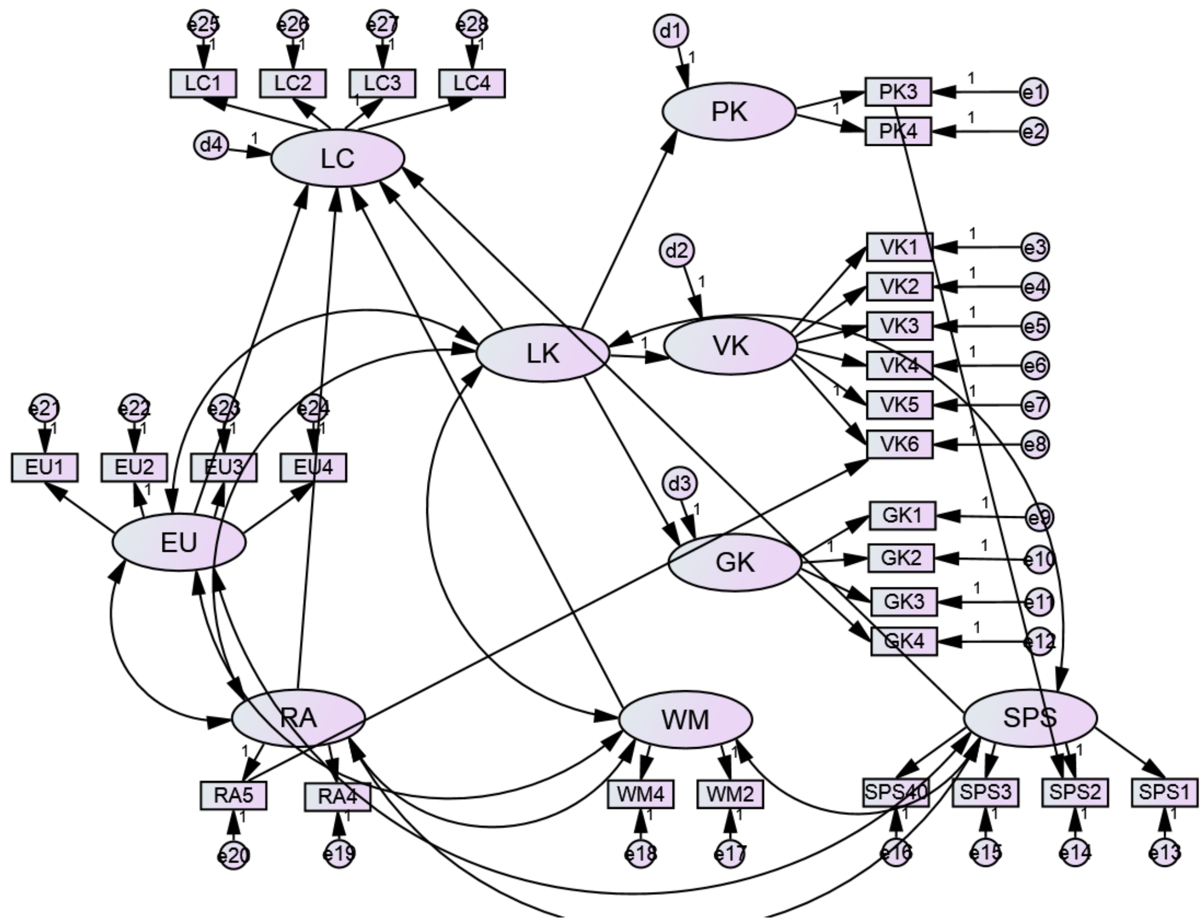


Figure 5.24 The structural mode of LC with the loading path from RA5 to VK6 added

5.5 Testing the structural equation model with the CET4 listening test

5.5.1 Testing the structural model fit

The structural model fit was checked in order to ascertain whether the data for the study were suitable for the hypothesized structural equation model of LC with the CET4 listening measuring LC.

Firstly, the dependent variable of LC was added to the second-order measurement model. The hypothesized structural model of LC was formed by adding single-headed arrows from the five latent constructs of LK, SPS, WM, RA, and EU to LC (see Figure 5.25). The structural model was run. Since the regression weight from RA4 onto RA was larger than 1 and the regression weight from RA5 onto RA was smaller than 1, the regression weight from RA4 onto RA was set to 1 and the model was run again (see Figure 5.26). As can be seen in Appendix 170, the assumption of multivariate normality was met (Mardia's coefficient = 4.28). Then, the residuals of measured variables were checked to ascertain whether there were discrepancies between the hypothesized model and the estimated model. It was found that only two of the residuals were greater than 2.58 (the residual covariance between VK1 and GK3 and the residual covariance between GK4 and RA4). Therefore, it was considered that not many assumptions of residuals were violated and the hypothesized model and the estimated model contained no discrepancies. Then, the significance of actual path estimates between the five latent constructs and LC was checked in order to ascertain whether the associations between the constructs were significant. The results showed that the *p*-values of unstandardized regression weights between LC and each of the three constructs of LK, SPS, and EU were significant, this means the three latent variables were predictors of LC. However, the *p*-values of the estimates of regression weights between LC and RA and WM were not significant (see Appendix 173), which means RA and WM were not predictors of LC. Based on the modification indices (MI), the regressions from PK3 to SPS2 and from RA to VK6 (the most important two changes to be made on regression weights) were added to the structural model (see Figure 5.27). However, the changes were not very effective, i.e., the significance on the estimates of regression weight between LC and RA and WM was not changed. The model was not significantly improved with this change. Therefore, I decided not to make any changes to the model and the insignificant path coefficients were kept in the entire structural model.

The results of analyses of the model can be seen in Appendices 174 to177. The correlations between each of the five latent constructs and LC were analysed (see Figure 5.28). The results of these correlations and the standardized regression weights of each latent factor onto LC can be seen in Table 5.32 (see Appendices 174 & 178 for details). The SEM model of LC explained 35.4% of the variance in LC. The model fit was good (χ^2 = 417.006, $p < .05$, χ^2/DF = 1.616, RMSEA = .058, SRMR = .055, CFI = .95) (see Appendices 179 -182).

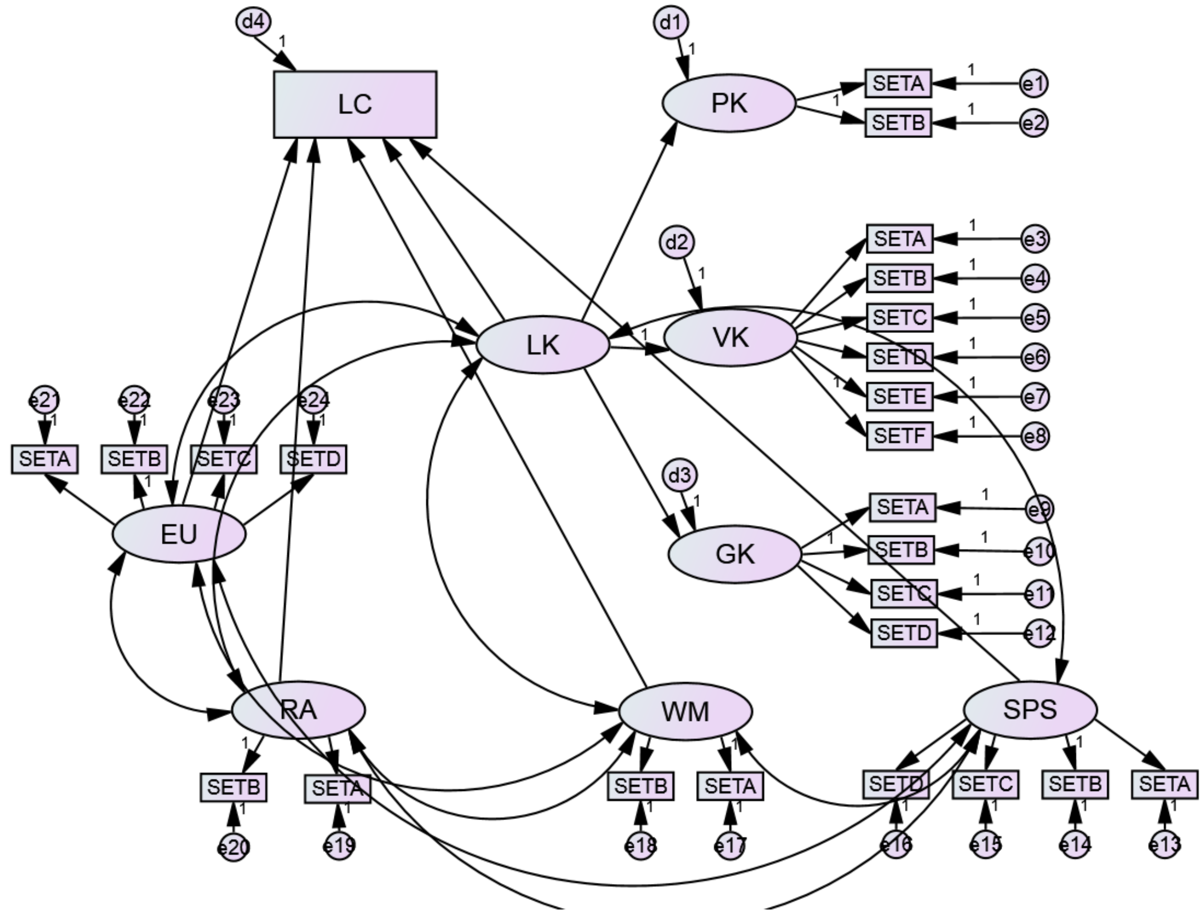


Figure 5.25 The hypothesized structural model of LC with CET4 listening measuring LC

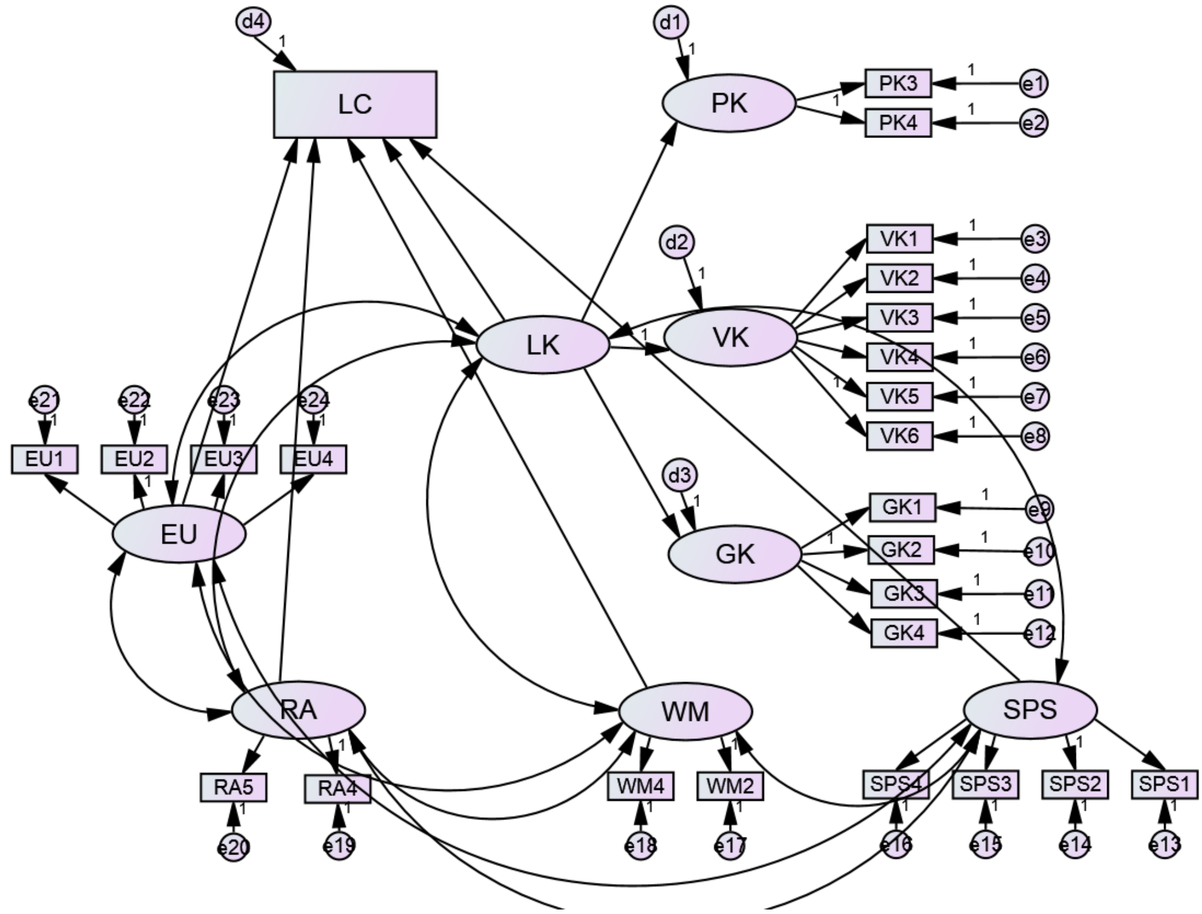


Figure 5.26 The hypothesized structural model of LC with data

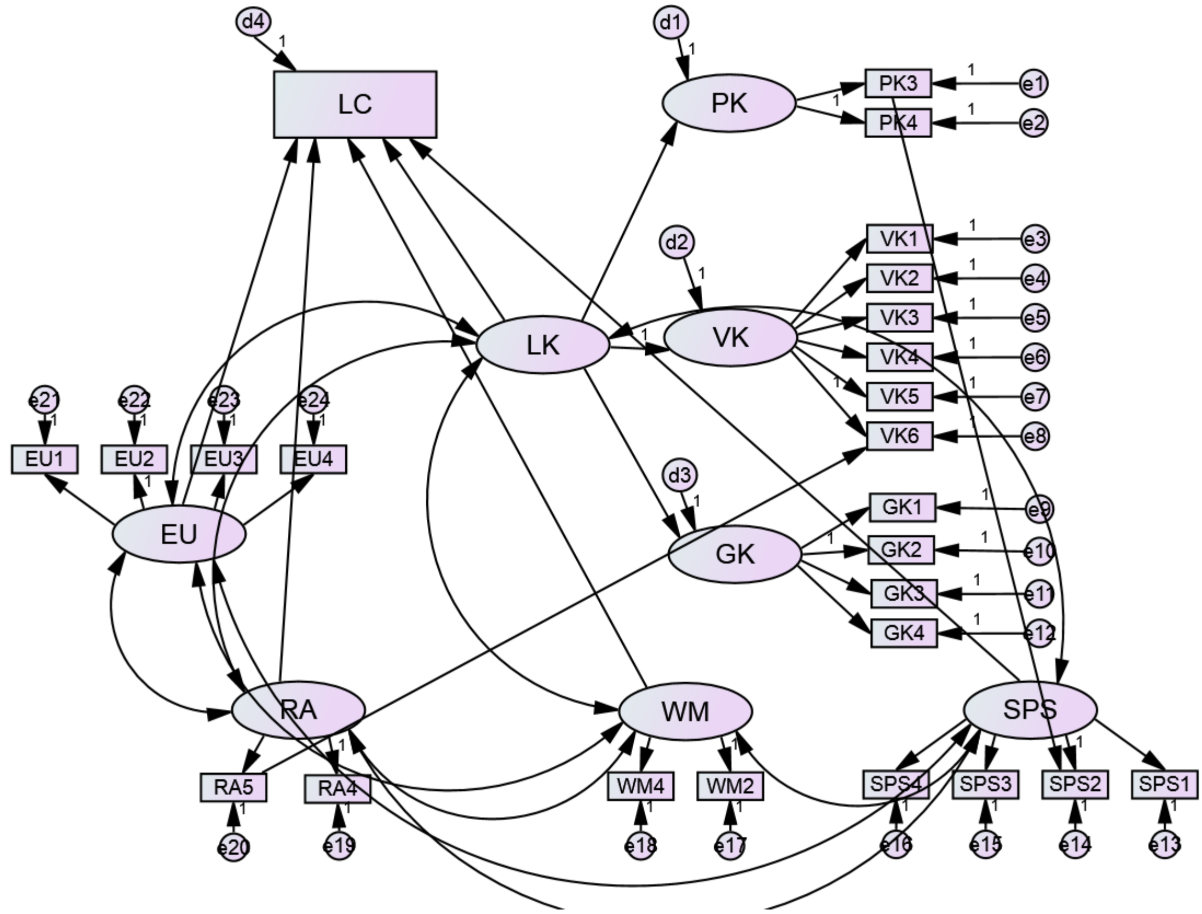


Figure 5.27 The structural model with two loading paths added from PK3 to SPS2 and from RA5 to VK6

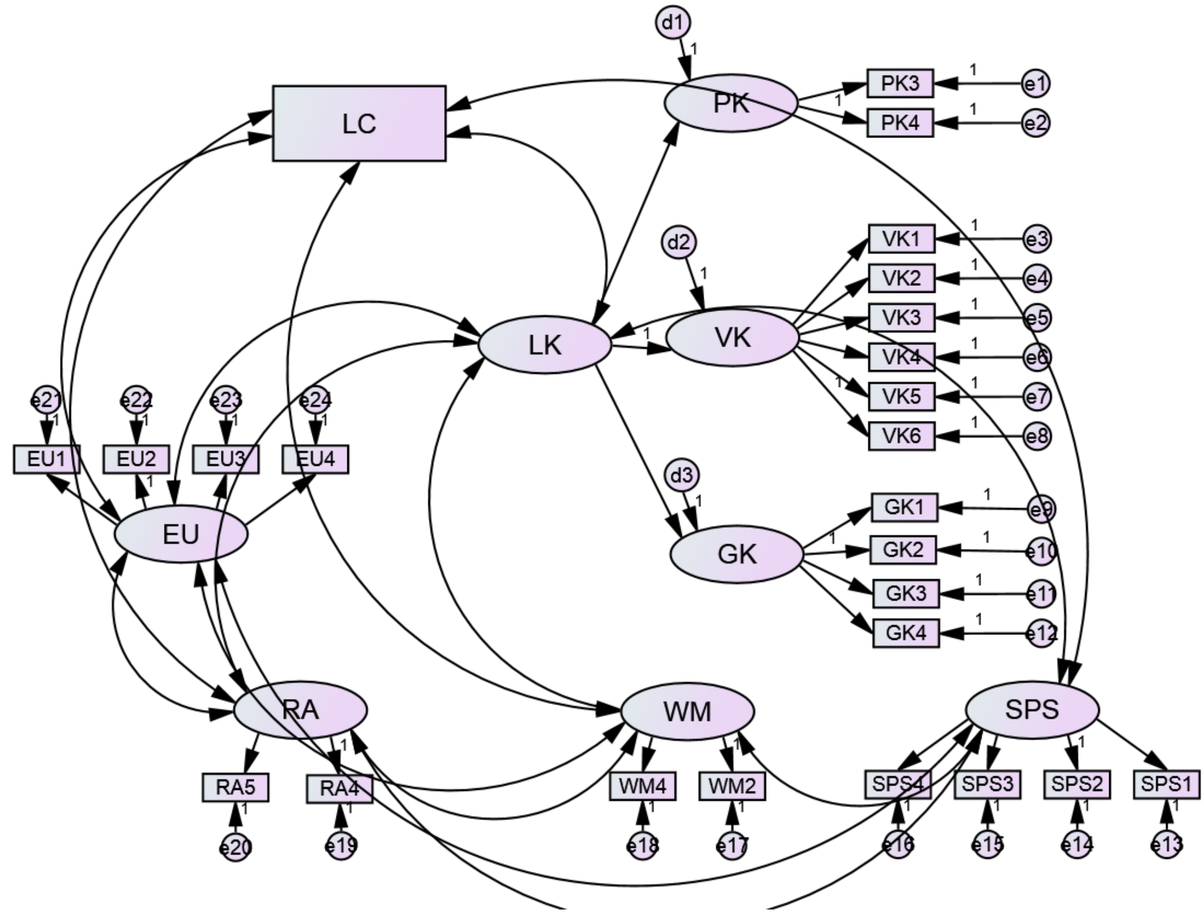


Figure 5.28 The structural model with all correlations between each of the five latent constructs and LC

Table 5.32 Implied correlations and standardized regression weights of the latent factors with LC

Predictors	Correlations	Stand. Regression weights
LK	.434	.838
SPS	.121	.291
WM	.160	-.001
RA	-.014	-.099
EU	.088	-.516
Variance explained		35.4%

5.5.2 Model evaluation

The structural model was tested with composite scale indicators to ascertain whether the structural model was plausible. First the CRs of the latent constructs were calculated (see Table 5.33). Although the loading weights from EU, RA, and WM were negative, when the three constructs were deleted from the structural model, the variance explained in the model was much reduced. Therefore, they were kept in the model. Next, the structural model was turned into a composite scale model (see Figure 5.29) where the new variables of means replaced the old variables (see Figure 5.30). Then, the factor loadings and error variances of each of these latent constructs onto the composite scale indicators were calculated. The results of this calculation can be found in Table 5.34. The calculated factor loadings and error variances were put into the composite scale model and the model was analysed in Amos. Since the loading from GK onto LK was greater than 1 and was greater than the loadings from VK onto LK or from PK onto LK, the loading weight from GK to LK was set to 1. (see Figures 5.29-5.30).

Table 5.33 The results of the analysis on CR of each latent construct

Convergent validity	LK	PK	VK	GK	SPS	WM	RA	EU
CR	.841	.726	.882	.859	.844	.712	.613	.984

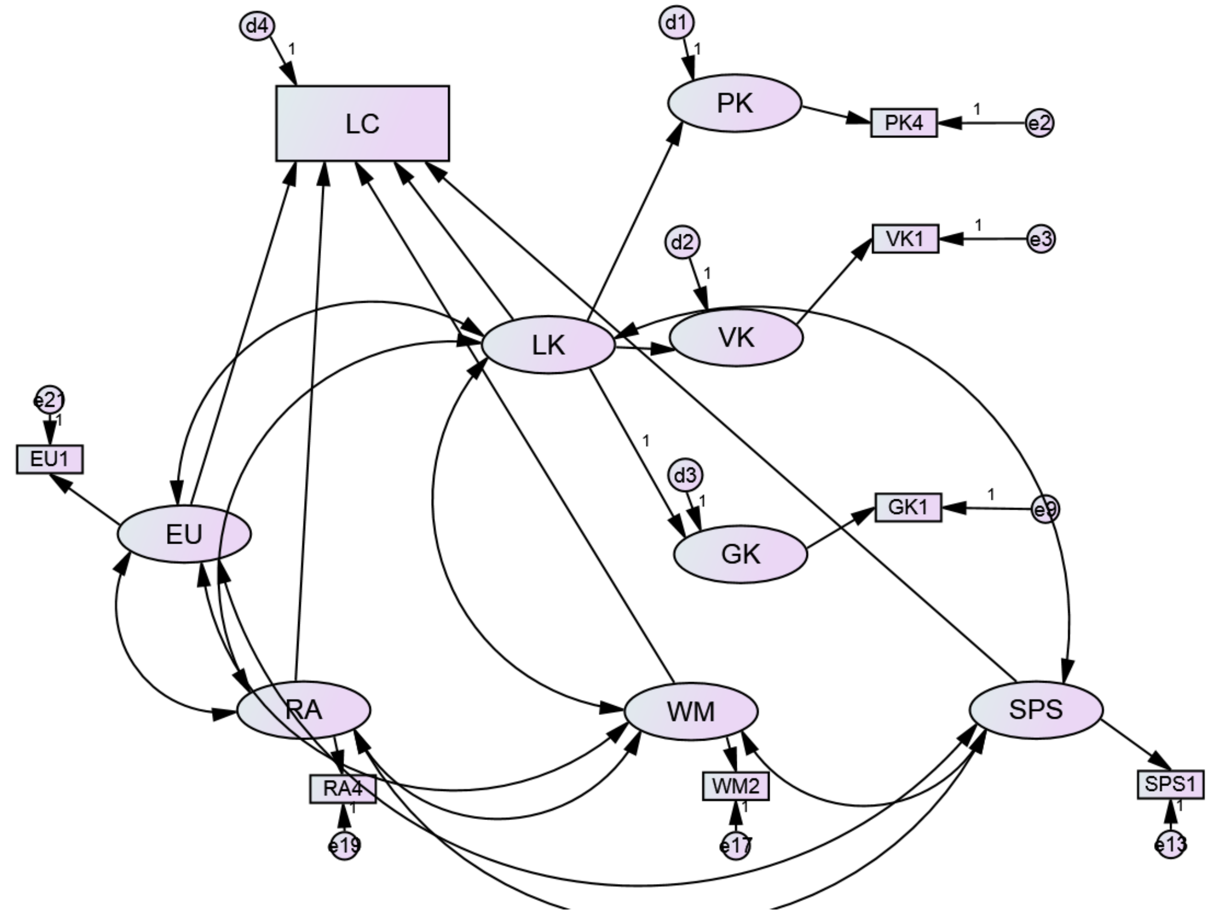


Figure 5.29 The composite scale model of LC with CET4 measuring LC

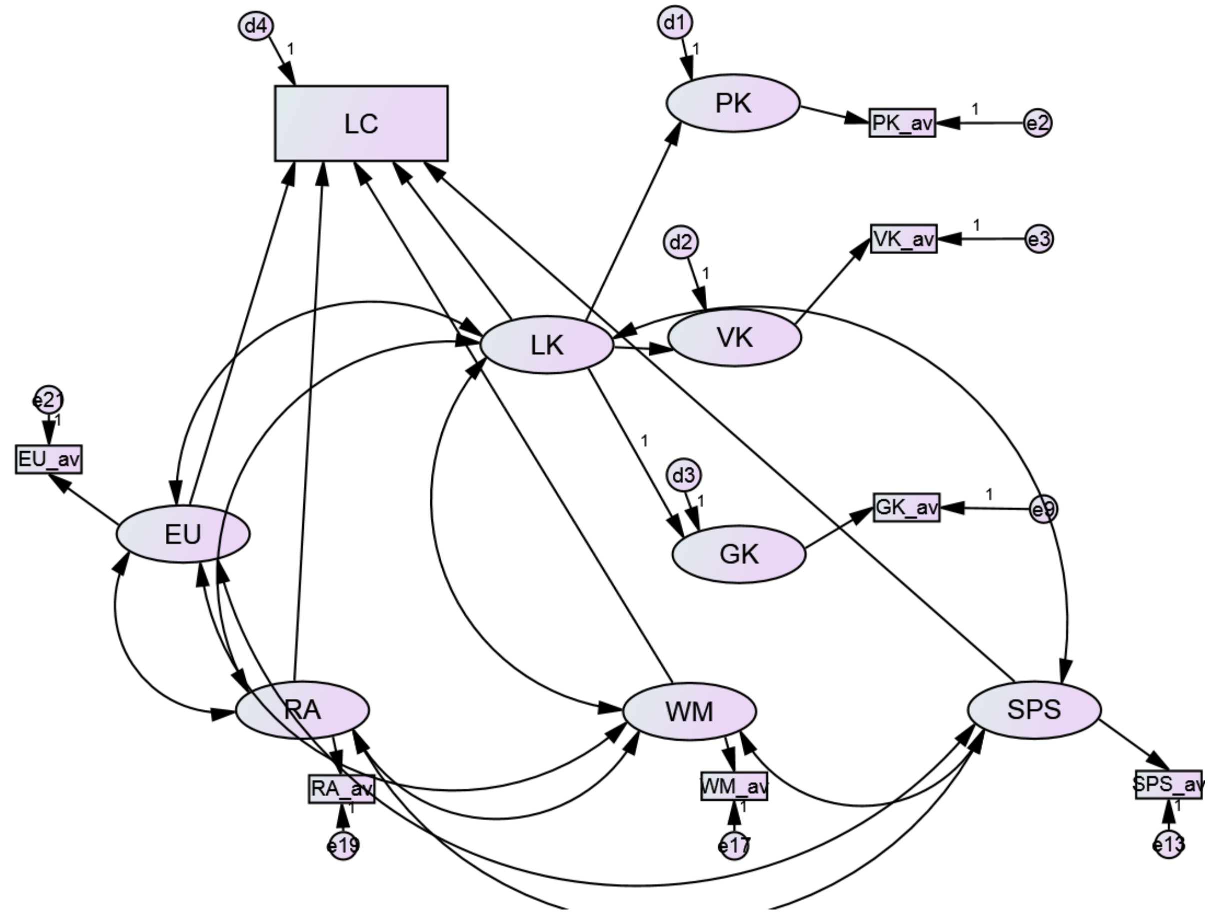


Figure 5.30 The composite scale model of LC with CET4 listening with data in

Table 5.34 Analyses of factor loadings and error variances for each latent construct with the CET4 listening

Convergent validity	LK	PK	VK	GK	SPS	WM	RA	EU
CR	.841	.726	.882	.859	.844	.712	.613	.984
Factor loading (root CR)	.917	.852	.939	.927	.919	.844	.783	.992
Error variance (1-CR)	.159	.274	.118	.141	.156	.288	.387	.016

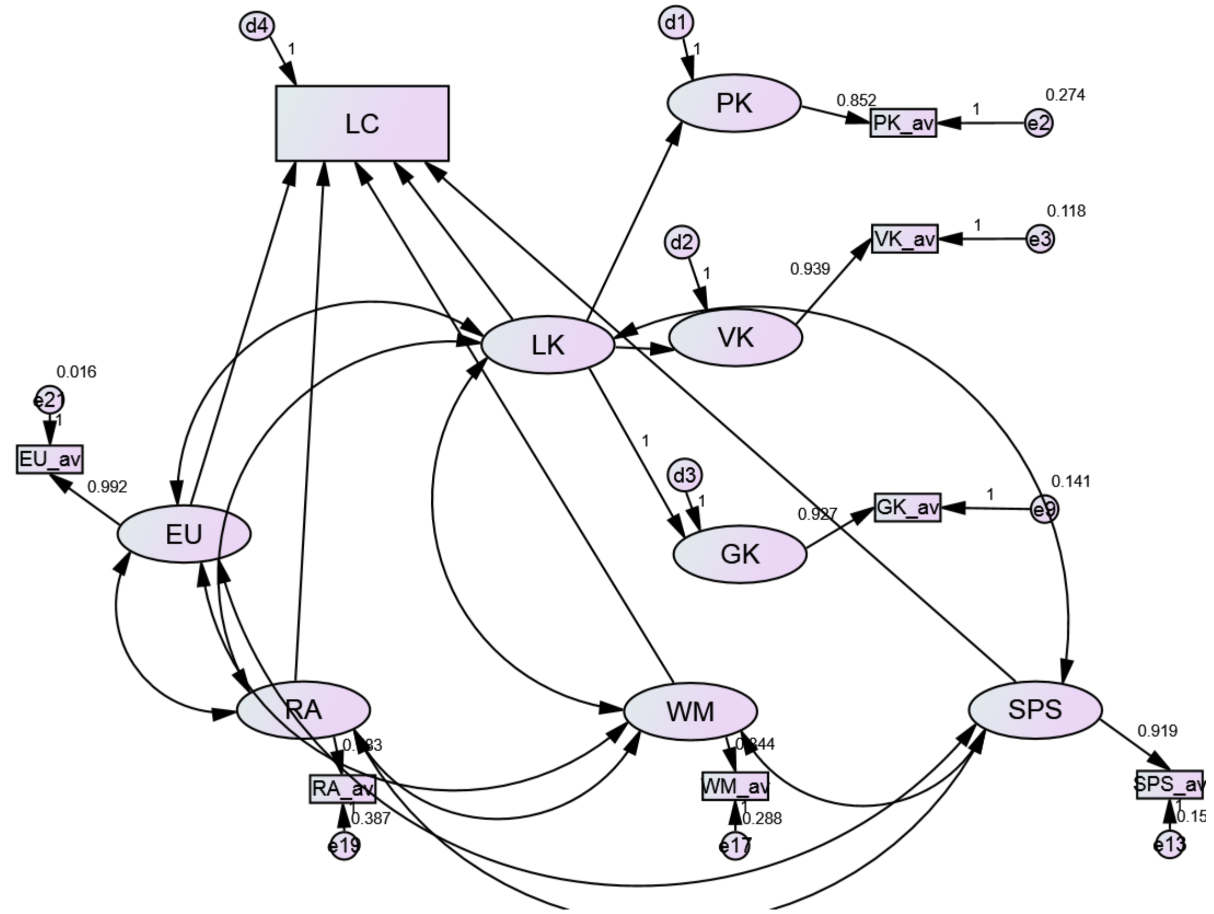


Figure 5.31 The composite scale model of LC with loading weights and error variances

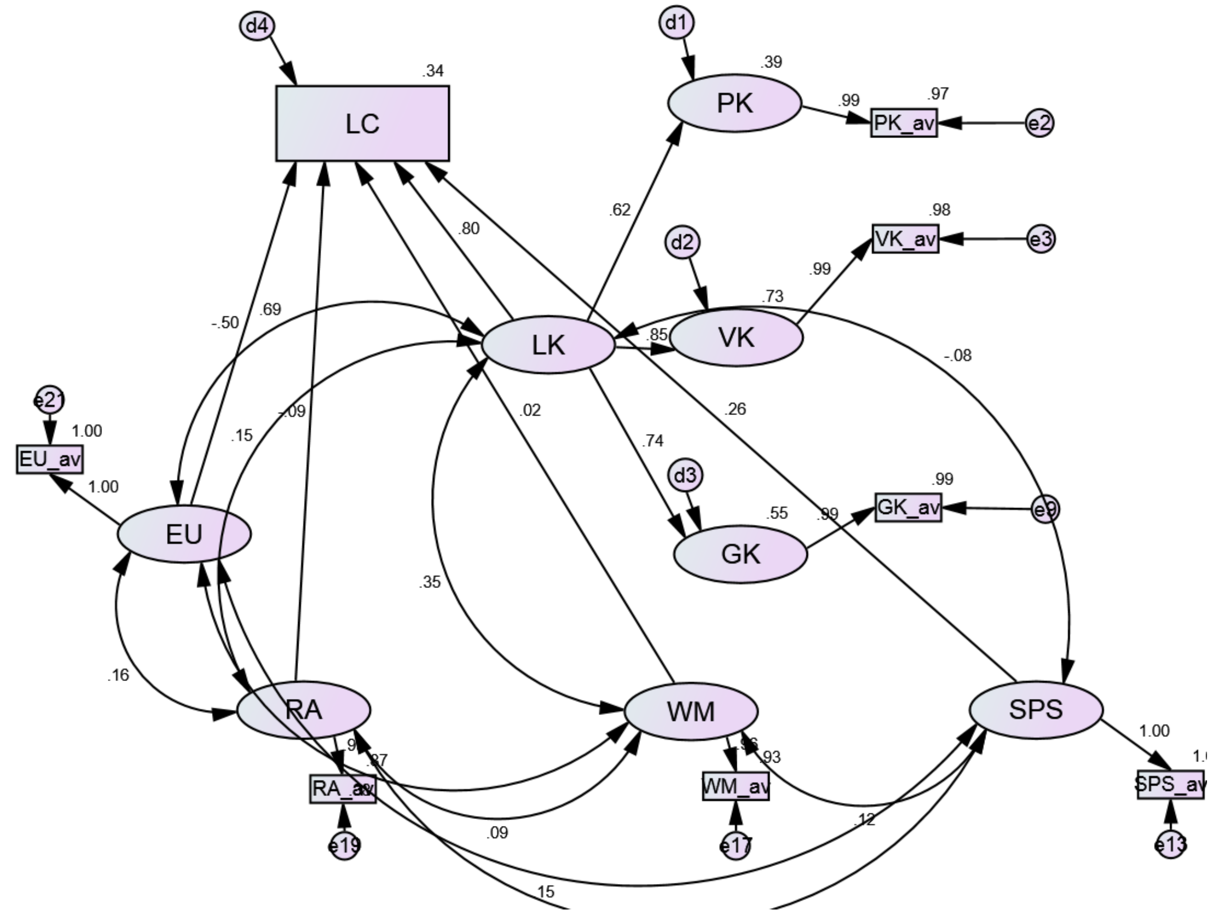


Figure 5.32 The composite scale model analysed under Amos

The results of analyses show that the model fit was not good (χ^2 = 29.717, p < .05, χ^2/DF = 2.972, RMSEA = .103). As can be seen in Appendix 183, two of the five loading paths from the latent constructs to LC remained not significant. However, Appendix 184 on standardised residual covariances shows that all the covariances were less than 2.58, which indicates that the model is accurate for examining the associations between the constructs and LC. MI was checked to ascertain whether any suggestions were given to make significant changes between constructs. Based on the suggestion from MI, two loading paths (from RA-average to GK-average and from PK-average to GK-average) were added to the model (see Figure 5.33). It was found that the model fit had changed significantly and was very good after the changes (χ^2 = 11.84, p > .05, χ^2/DF = 1.480, RMSEA = .051, CFI = .989) (see Appendices 185 -188).

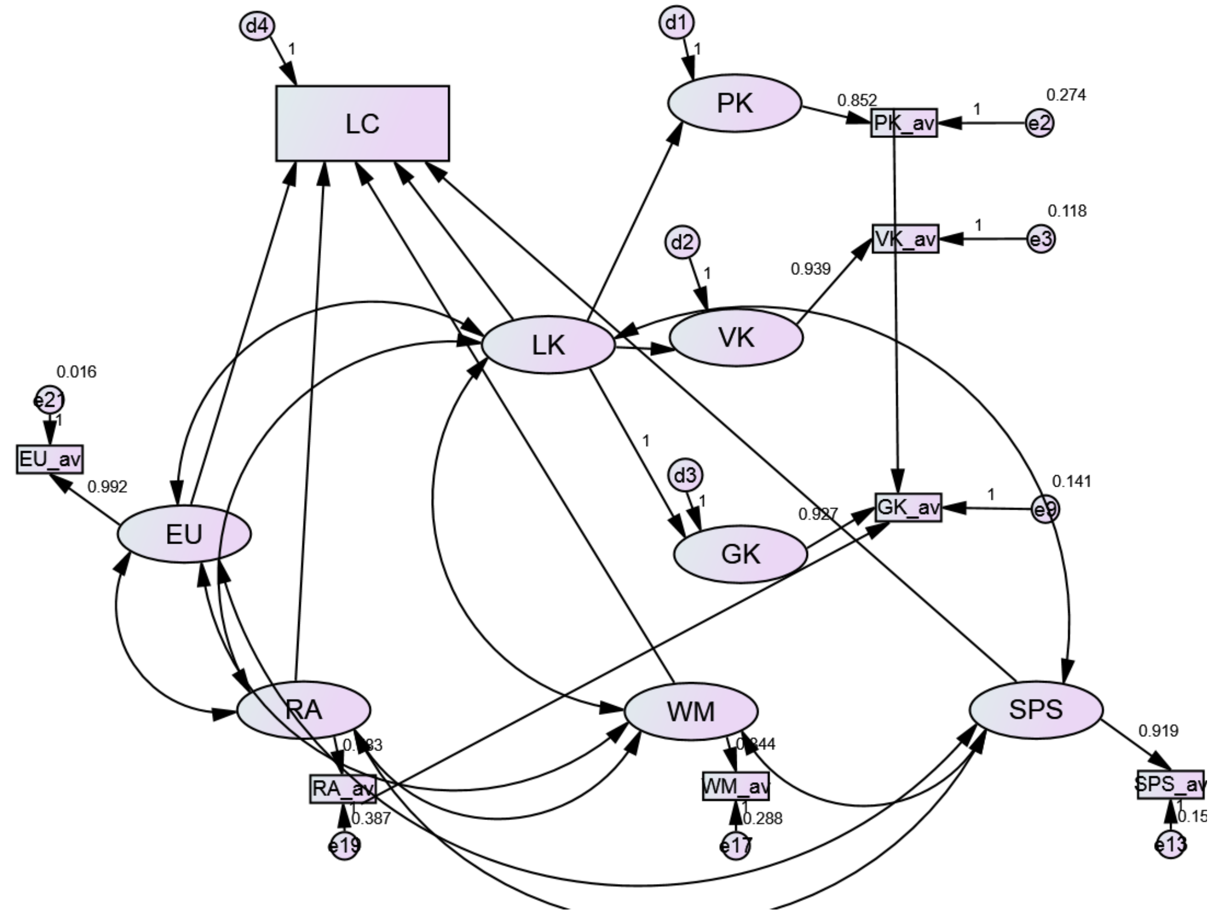


Figure 5.33 The composite scale model with two loading paths added

From the results of the evaluation of the structural model with composite scale indicators, the following conclusion was made: the hypothesized structural model of LC was an accurate model. The overall model fit was good. But further steps were taken to modify the structural model based on the suggestions from MI.

5.5.3 Model modification

Model modification was tried to improve the model fit based on MI (see Appendix 189). First, the loading path from PK3 to SPS2 was added to the model (see Figure 5.34). Appendices 190 to 193 indicate that the model was not significantly improved with this change (χ^2 = 399.436, $p < .05$, χ^2/DF = 1.554, RMSEA = .055). Based on the new MI, as can be seen in Appendix 194, the loading path from RA5 to VK6 was added (see Figure 5.35). The results indicated that the model was not significantly improved either (χ^2 = 385.803, $p < .05$, χ^2/DF = 1.507, RMSEA = .052, CFI = .959) (see Appendices 195 -198). Since the overall model was not significantly improved with the modification, I decided not to make changes to the entire model.

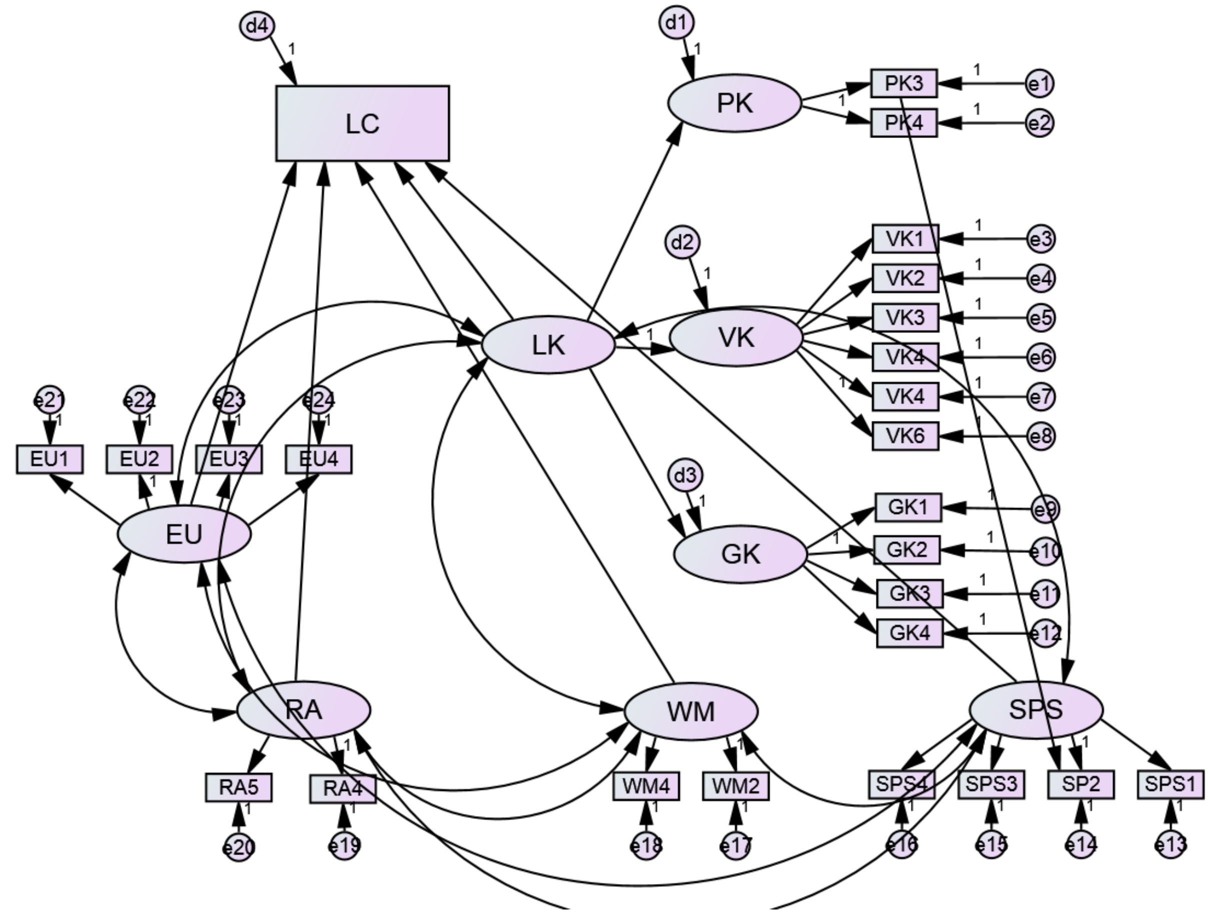


Figure 5.34 The structural model with the loading path from PK3 to SPS2 added

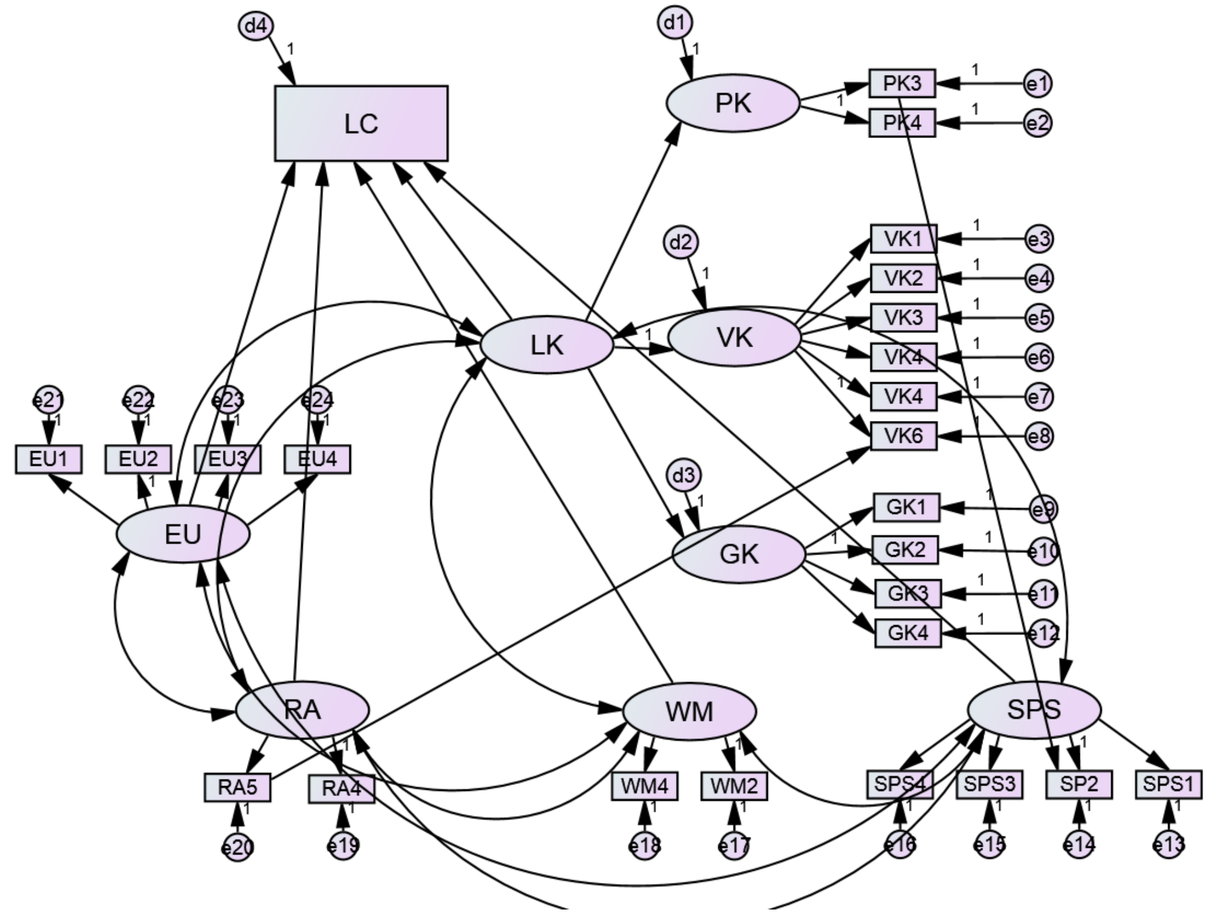


Figure 5.35 The structural model of LC with the loading path from RA5 to VK6

5.6 Answers to RQs

To answer the RQs, five factors on learners' individual differences were assumed to explain variance in L2 LC among 187 adult learners of English. Since two different tasks were used to measure learners' LC and the scores of the two tasks could not be combined and used in one model, two different SEM models of LC were built, one for the CET listening task and one for the PET listening task. In order to compare with which analysis method more variance can be explained, structural equation modelling and hierarchical regression method were adopted.

RQ1: Which individual differences can explain variance in L2 LC among adult Chinese learners of English?

Results from the structural equation modelling analyses show that when LC was measured with the PET listening, together the five factors (linguistic knowledge, aural sentence processing speed, working memory capacity, reasoning ability, and frequency of the use of English) can explain 65.7% of the variance in LC among adult Chinese learners of English. Linguistic knowledge and the frequency of use of English were found to have significant associations with LC and are key predictors. Linguistic knowledge is more important in predicting success in L2 LC than the frequency of use of English. Sentence processing speed, working memory capacity and reasoning ability are not predictors of LC.

Results from the structural equation modelling analyses show when LC was measured with the CET4 listening, together the five factors can explain 35.4% of the variance in LC among adult Chinese learners of English. Linguistic knowledge, the frequency of use of English, and sentence processing speed were found to have significant associations with LC and are key predictors. Linguistic knowledge is found to be the most important predictor of L2 LC. Working memory capacity and reasoning ability are not predictors of LC.

Results from hierarchical regression analyses indicate that when LC was measured with the PET listening, all observed variables could explain 44.3% of the variance in LC. This is much less than the variance the SEM model could explain. Linguistic knowledge explained 36.3% of the variance in LC, thus linguistic knowledge is more important in predicting success in L2 LC than the other four factors. When all variables were entered to the model together, phonological knowledge explained 35.3% of the variance in LC, followed by frequency of the use of English, which explained 5.8% of the variance, and grammar knowledge which explained 2.2% of the variance. LVLT explained 0.1%, followed by WM capacity, which explained 0.6%; Raven' SPM, 0.3%; and OGJ, 0.1%. Linguistic knowledge, phonological knowledge, frequency of the use of English, vocabulary knowledge and grammar knowledge are key predictors of L2 LC. In addition, phonological knowledge explained uniquely 12.8% of the variance in LC, thus, phonological knowledge explained unique variance in LC over and above the contribution of vocabulary knowledge. Also, phonological knowledge explained much more variance in LC than vocabulary knowledge and grammar knowledge. Vocabulary knowledge is a significant predictor on its own which explained overall 22.5% of the variance. Word recognition from speech which uses phonological knowledge is more important in predicting the variance in L2 LC than word segmentation from the speech stream. For the entire group, sentence processing speed, reasoning ability, and working memory capacity are not key predictors of L2 LC.

When LC was measured with the CET4 listening, together all observed variables explained 28.2% of the variance in LC. This variance is much less than what the SEM model could explain and is much less than that explained for the PET listening. This will be discussed in detail in Chapter 6. Together, variables representing linguistic knowledge explained 17.9% of the variance in LC, thus, linguistic knowledge is more important in predicting success in L2 LC than the other variables. When all variables were entered into the model, phonological knowledge explained 13.5% of the variance in LC, followed by grammar knowledge explaining 3.3% of the variance, the frequency of use of English explaining 5.1%, aural sentence processing speed explaining 2.5%, vocabulary knowledge explaining 2.1% of the variance, working memory capacity explaining 1.5% and reasoning ability explaining 0.2%. The key predictors are linguistic knowledge, phonological knowledge, frequency of the use of English, sentence processing speed, grammar knowledge, and vocabulary knowledge. Word recognition from speech explained the variance in LC over and above the contribution of word segmentation from the speech stream. Word recognition from speech using phonological knowledge is more important in predicting success in L2 LC than word segmentation from the speech stream. In addition, phonological knowledge and vocabulary knowledge are a significant predictor on their own: the former explained 13.5% of the variance in LC and the latter explained 13% of the variance. For the entire group, reasoning ability and working memory capacity are not key predictors of L2 LC.

RQ2: What is the contribution of the following variables in explaining the variance of L2 LC: linguistic knowledge; aural sentence processing speed; working memory; reasoning ability; and frequency of use of an L2?

The analyses of regression models of LC measured with the CET4 listening and the PET listening showed that phonological knowledge was the most reliable predictor. In addition, the findings showed that word recognition from speech was more important in predicting success in LC for Chinese learners of English than word segmentation from the speech stream.

When LC was measured with the PET listening, the overall variance explained by word recognition from speech was 34.7%, but some of the variance was shared with other variables. Word recognition from speech uniquely explained 29.9% of the variance in LC. However, word segmentation from the speech stream did not explain unique variance over and above the contribution of word recognition from speech. The overall variance explained by word segmentation from the speech stream was 4.7%, but the variance was shared with other variables and it did not explain any unique variance in LC. The overall variance explained by phonological knowledge was 35.3%, but some of the variance was shared with other variables. Phonological knowledge uniquely explained 12.8% of the variance in LC. The overall variance explained by vocabulary knowledge was 22.5%. Vocabulary knowledge is a significant predictor on its own. The overall variance explained by frequency of the use of English was 34.9%, but some of the variance was shared with other variables. Frequency of the use of English uniquely explained 5.8% of the variance in L2 LC.

When LC was measured with the CET4 listening, the overall variance explained by word recognition from speech was 13.1%, but some of the variance was shared with other variables. Word recognition from speech explained uniquely 10% of the variance in LC. The overall variance explained by word segmentation from the speech stream was 3.5%, but the variance was shared with other variables. Word segmentation did not explain unique variance in LC. The overall variance explained by phonological knowledge was 13.5%, but the variance was shared with other variables. The overall variance explained by vocabulary knowledge was also 13%, but the variance was also shared with other variables. Therefore, phonological knowledge and vocabulary knowledge were significant predictors on their own.

3. To what extent do Chinese students in China and in the UK differ from each other in L2 LC? Which factors can explain these differences?

The study found that LC of Chinese learners of English in China was significantly lower than that of learners in the UK. For learners in China, together the variables of word recognition from speech, distinguishing similar sounds, vocabulary knowledge, aural and written grammar knowledge, working memory capacity, frequency of the use of English in daily life and sentence processing speed explained 32.8% of the variance in LC, among which phonological knowledge (word recognition from speech and distinguishing similar sounds) explained 25.7% of the variance, followed by sentence processing speed, which explained 3.9% of the variance in LC. Other variables each explained very little variance in LC and the adding of explained variance (together 3.2%) did not make significant change to the model. The results mean that, in this group, learners' phonological knowledge and aural sentence processing speed are significant predictors of LC. Phonological knowledge is the most important predicting variable of LC. However, learners' vocabulary knowledge, grammar knowledge, working memory capacity, and frequency of the use of English in daily life are not significant predictors of LC.

For learners in the UK, results indicate that together word recognition from speech, grammar knowledge, vocabulary knowledge, and reasoning ability explained 52.2% of the variance in LC, of which grammar knowledge explained 40.4%, followed by reasoning ability, which explained 9.2% of the variance in LC. The addition of explained variance by word recognition from speech and vocabulary knowledge, together 2.7%, did not significantly contribute to the change of the model. The results mean that, in this group, learners' grammar knowledge and reasoning ability are significant predictors of LC. Grammar knowledge was the most important predicting variable for this group. Results also suggest that learners' ability to recognise words from speech and their vocabulary knowledge are not significant predictors of LC.

To summarise, there are important differences in the results for the two subgroups. For Group 1, learners' phonological knowledge and aural sentence processing speed are significant predictors of LC, but learners' vocabulary knowledge, grammar knowledge, working memory capacities, and frequency of the use of English in daily life are not significant predictors of LC. For Group 2, learners' grammar knowledge and reasoning ability can significantly predict LC, but learners' word recognition ability and vocabulary knowledge cannot.

Chapter 6 Discussion

In this chapter, I summarize the results of the current study and discuss the findings as they compare with other related studies. Findings for the entire group (*N* = 187) are discussed first. In Section 6.1, I compare the findings of the current study on word segmentation from the speech stream with findings from other studies. In section 6.2, models of LC used in this and other studies are compared and similarities and differences between the models are explored. In Section 6.3, how the findings of the current study underline the importance of the lower-level processes in LC is discussed. In the fourth section, Chinese learners' listening problems are examined. In Section 6.5, findings on learner variables in this and other studies are compared. In Section 6.6, factors which can explain differences in LC between two subgroups are discussed. In Section 6.7, a discussion of the CET4 listening test ensues as it is compared to the PET listening test. In the final section, the implications of the results are discussed.

6.1 A comparison of studies on word segmentation from the speech stream

Although recognising word from speech was found to be the biggest problem for Chinese learners of English, the findings about learners' ability to segment words from the speech stream in the current study indicate that Chinese learners have segmenting problems in processing English speech input. The findings suggest that Chinese learners have the lowest mean accuracy on all items (69.4%) in the word segmentation test compared with that found in Altenberg (2005) (76.3%), in Ito and Strange (2009) (83.8%), and in Shoemaker (2014) (74.6%). The findings mean that word segmentation and word recognition from speech are indeed more difficult for Chinese learners of English. The current study also found that Chinese learners had less accuracy on Double cue stimuli (61.54%) than on Glottal stimuli (67.28%), or on Aspiration cue stimuli (74.14%). Altenberg (2005) found that L2 learners had relatively higher accuracy on both negative glottal stop stimuli (77%) and positive glottal stop stimuli (99.1%). Ito and Strange (2009) found that L2 learners had better performance on double cue pairs (94% accuracy) than on glottal stop pairs (91% accuracy) and aspiration pairs (73% accuracy). Shoemaker (2014) found that L2 learners had better performance on positive glottal stop stimuli than on other types of stimuli. Although the stimuli used in the current study and in Altenberg (2005), Ito and Strange (2009), and Shoemaker (2014) are the same, the findings of the current study differ. There are several possible reasons for this. The first may be that the learners' L1 differs from those in the other studies. The current study looked at Chinese learners of English, but Altenberg (2005) looked at Spanish learners of English, Ito and Strange (2009) looked at Japanese learners of English, and Shoemaker (2014) looked at French learners of English. For Chinese learners of English, it seems listening is more difficult than for these other learners. The second reason might be that the stimuli utterances were produced by a different speaker. The stimuli were read by a female English speaker who spoke Standard British English. In Altenberg (2005) the stimuli were read by a native male speaker of New York English. Shoemaker (2014) and Ito and Strange (2009) used the same recordings which were produced by a female native speaker of American English with a New York dialect. Xu (2014) mentioned that some Chinese learners are unable to adapt to sound differences between British and American English which means that speakers' dialects influence L2 learners' listening performance. The third reason might be that Chinese learners in China have no study abroad experience and use very little English outside of their English classes. In Altenberg (2005), L2 learners were college or university students in New York area. On average, they had studied English for 5.1 years and had been in the USA for 5.8 years. In Ito and Strange (2009), L2 learners were native speakers of Japanese who were living in the USA. Learners' mean immersion experience, i.e., their education experience in English-speaking countries, was 2.5 years and ranged from 0 to 10 years. The mean length of residence in English-speaking countries was four years and ranged from two weeks to 12 years. In Shoemaker (2014), learners were native French university students enrolled in English language studies in France. One group of learners had studied English for a mean 8.7 years and the other group for a mean 11.2 years. Findings in this and the above studies suggest that immersion in an L2 environment is very important to improving L2 learners' listening proficiency. Findings in the current study on Chinese learners' segmentation problems when listening fill some gaps in the literature on L2 listening and have important pedagogical implications for the teaching of listening to Chinese learners of English in the UK and in China.

6.2 A comparison of models of L2 listening comprehension

One of the aims of this current study was to explain individual differences in LC among Chinese learners of English. The study is a partial replication of Andringa et al. (2012) who built a SEM model of LC for native speakers and L2 learners of Dutch. They found that for native speakers the key variables which explained 91% of the variance in LC are linguistic knowledge and processing speed; for L2 learners the key variables are linguistic knowledge and reasoning ability which together explained 96% of the variance in LC. Findings of the current study differ from those of Andringa et al. for L2 learners. In the current study, for SEM models L2 learners' linguistic knowledge and frequency of the use of English in daily life were key variables when LC was measured with the PET listening; learners' linguistic knowledge, frequency of the use of English in daily life and aural sentence processing speed were key variables when LC was measured with the CET4 listening. There may be several reasons for the differences in the findings between the two studies. The first is that the L2s differ. They looked at the LC of Dutch and the current study looked at LC of English. Although Dutch and English are both Germanic languages, it is not necessarily the case that they are equally easy to listen to. The second reason is that different tests were used to measure the constructs. As presented in the Methodology chapter, the forms of the tests used in the two studies are different. For example, in their study, a written receptive test was used to measure learners' vocabulary knowledge, but in the current study an aural receptive test was used for the same purpose. In their study four digit span tasks comprised of two in visual form and two in auditory form, and one non-word recognition task were used to measure learners' working memory capacity. However, in the current study, four digit span tasks comprised of two in an L1 version and two in L2 version were used to measure learners' working memory.

The third reason is that frequency of the use of English in daily life was added to the models of LC in the current study. A particular contrast that highlights the differences between the two studies is that the current study found the significant contribution of learners' ability to recognise words from speech and the frequency of the use of English in daily life to an explanation for the success of L2 learners in LC. Andringa et al. commented on why the native speakers' model of LC differs from the L2 learners' model of LC. They claimed that because native speakers use Dutch more than L2 learners the frequency of the use of L2 could explain some variance in L2 learners' LC. The current study in this respect is suggested by that comment and is why it looked in more detail at the contribution of usage of L2 to success in LC. The current study found that in the two hierarchical regression models of LC as measured with the two different listening tasks, frequency of the use of English in daily life explained overall 34.9% of the variance and 5.1% in LC, respectively. As presented in the Methodology Chapter, learners in China did not have study abroad experience, but learners in the UK had studied for a duration ranging from 12 to 36 months. I did not put the length of residence as a variable but used an index of English usage in the present study. The results concerning the frequency of English use in everyday activities, investigated in a personal background questionnaire, showed that learners in China spent 4.33% of their time each day using English, significantly less than that of learners in the UK who spent 51.12% of their time using English. The differences are particularly noticeable in their use of English in the contexts of education, transportation, trips, shopping, studies, health, and politics. This means that there are very different patterns of English usage between learners in an L2 environment and learners who are not immersed in an L2 environment. Learners of English in China mainly use English in classes, nowhere else, but learners in the UK use English when they go shopping, see films with friends, and travel. This measure of English use reflects learners' residence experiences because it correlates well with length of residence ($r = .84$, $p < .01$). This finding means that the longer learners are immersed in an English language environment, the more frequently they use English. The findings suggest that Andringa et al.'s comment on the usage of L2 is right and L2 use plays an important role in explaining success in LC. The fourth reason is that in the current study L2 learners are mainly from one language background and in Andringa et al. L2 learners are from 35 L1s. Therefore, there is much more variance in their data because of all these different L1s the L2 learners brought compared to those in the current study with just one L1. First languages have a great influence on a person's capacity to learn another language.

Although different components were found to explain individual differences in LC of L2 learners in both studies, there was a common component which had significant associations with LC: linguistic knowledge. In both studies, linguistic knowledge comprised three subcategories: phonological knowledge; vocabulary knowledge; and grammar knowledge. The factor loading of L2 learners' linguistic knowledge onto LC in Andringa et al.'s study is higher than the two factor loadings in the current study for SEM models. In the present study, hierarchical regression models showed that learners' linguistic knowledge explained 36.3% of the variance out of 44.3% in LC as measured with the PET listening. Meanwhile, this variable explained 17.9% of the variance out of 28.2% in LC as measured with the CET4 listening. The results mean that learners' linguistic knowledge contributes the most to explaining variance in LC among the other variables. Results of the current study also suggest that Chinese learners' ability to recognise words from speech contributes more to predicting LC than does their ability to segment words from speech.

Other studies also built models of L2 LC. Vandergrift and Baker (2015) built a provisional model of LC based on an investigation among L2 learners in French Immersion classrooms. Their study found that the key variables were auditory discrimination, working memory, L1 and L2 vocabulary knowledge, and metacognition. Their model suggests that auditory discrimination has a positive influence on working memory and both are important to the development of L1 and L2 vocabulary knowledge. In addition, L1 vocabulary and metacognition have a positive impact on L2 LC through L2 vocabulary knowledge. Although L2 vocabulary knowledge was found to lead to L2 LC directly, the contribution of L2 vocabulary knowledge to LC was the result of impact from other learner variables in that study. The findings in the current study differ from their findings for several reasons. Firstly, it might be that different learner variables were explored in my study. Although both linguistic knowledge and working memory were explored in both studies, in the current study sentence processing speed, reasoning ability, and frequency of the use of English in daily life were also explored. The contribution of sentence processing speed reflects how efficiency of processing speech input can influence L2 LC. In addition, the findings on the contribution of frequency of the use of English in daily life throw new light on the understanding of factors which influence L2 LC. Since the current study mainly explored the contribution of learners' lower-level processes in LC, metacognition was not included. Such focus was chosen because of the phonological differences between English and Chinese (see Section 2.2) and because phonology has not been studied as widely as vocabulary. In future studies, learners' ability to use both linguistic and non-linguistic knowledge alongside listening strategies in processing L2 speech input can be explored.

The second reason is that the L2 was different in the two studies: English in this study and French in Vandergrift and Baker (2015). Although French and English are both Indo-European languages, it is not necessarily the case that they are equally easy to listen to. The third reason is that the language learning backgrounds of L2 learners in the two studies differed. In this study, although learners in the UK (*N* = 40) had an L2 immersion environment, learners in China (*N* = 147) had no L2 immersion environment. They mainly received instructions in L2 and used L2 in English classes which amounted to 200 minutes each week during the academic term. Meanwhile, in Vandergrift and Baker (2015) learners were from late French immersion classes in which all instruction was in French, except for mathematics, which means those learners were immersed in an L2 environment longer than the learners in China.

The fourth reason is that the importance of phonology for learners differed between the two studies. Phonology is a big barrier for Chinese learners. Y. Wang (2008), Goh (2000) and F. Wang (2008) found that Chinese learners had listening problems mainly in the perception stage. In Vandergrift and Baker (2015), phonology is initially important, but L2 vocabulary knowledge is the most important predictor of LC. In their study, sound discrimination which was input in an unknown language and a sound-system association task in which nonsense words were heard were used to measure learners' auditory discrimination. However, the two tasks might not have reflected learners' experiences in L2 because one task required learners to process an unknown language and the other to process nonsense words. Although auditory discrimination was found to be initially important in their model, if tasks in French were used to measure learners' auditory discrimination the results might have been different. Vandergrift and Baker (2015) argue that perception can be affected by learners' ability to discriminate sounds which are not included in their L1; the current study measured such ability. As mentioned in Chapter 2, interdentals and consonant clusters are unknown in Chinese. The fifth reason is that different analysis methods were used in the two studies. Vandergrift and Baker (2015) used path analyses to explore a potential causal relationship between learner variables and L2 LC and to explore the relationship among the learner variables. In the current study, the SEM analysis method was adopted to test against the SEM models and the hierarchical regression method was adopted to build the regression models of LC. The former is able to deal with multiple dependent and multiple independent variables and can be used to uncover very complicated relationships between variables (Schoonen, 2015). The hierarchical regression method is more suitable to the analysis of the changes in contribution of each added variable with partialing out impact from other added variables. Findings from the current study fill some gaps in the literature concerning the contributions of learner variables to an explanation of unique variance in L2 listening, e.g., word recognition from speech, word segmentation from the speech stream, phonological knowledge, and frequency of the use of English in daily life.

There are also some important similarities and differences between results from the current study and those of other studies which built multiple regression models of LC. Mecartty (2000) investigated relationships between L2 vocabulary knowledge and grammatical knowledge in reading and LC. She found that both vocabulary knowledge and grammar knowledge are significantly correlated with LC. However, her regression model of LC indicated that only vocabulary knowledge was the key variable as it explained 14% of the variance in LC; meanwhile grammar knowledge did not explain any variance in LC. Findings in the current study differ from Mecartty (2000) for several reasons. Firstly, additional learner variables were explored. While Mecartty (2000) investigated learners' linguistic knowledge which comprised vocabulary knowledge and grammar knowledge, in the current study phonological knowledge was also explored. This variable was found to be the most important predictor for Chinese learners' LC, but it was not explored in Mecartty (2000). Secondly, the learners' L1s are different. The current study investigated Chinese learners and Mecartty (2000) investigated Spanish learners of English. Chinese is a Tibetan language and Spanish is an Indo-European language, thus they are completely different. Thirdly, different tasks were adopted to measure learners' vocabulary knowledge and grammar knowledge. The current study adopted an aural vocabulary test to measure learners' vocabulary knowledge and written and aural grammar tests to measure learners' grammar knowledge. In Mecartty (2000), learners' vocabulary and grammar knowledge were measured using written tasks only. Although the written vocabulary test worked well in Mecartty's (2000) study, Vandergrift and Baker (2015) argue that a measurement of aural word recognition is more valid for an assessment of learners' word recognition in processing speech input. The current study used an aural vocabulary task. A written grammar task measures learners' ability to comprehend texts with grammar knowledge embedded in written form. The fact that learners comprehend texts with grammar knowledge embedded in written form does not necessarily mean that learners can comprehend such texts from speech. If an aural vocabulary task and an aural grammar task had been adopted in Mecartty (2000) the results might have been different.

Wang and Treffers-Daller (2017) built another hierarchical regression model of LC based on their study of Chinese learners of English. In their study the key variables were learners' vocabulary knowledge, general L2 language proficiency, and metacognition. Vocabulary knowledge explained the variance in LC over and above the contribution of L2 language proficiency. The current study confirms these findings concerning the contribution of linguistic knowledge to LC, but the current study provided additional evidence to show that for Chinese learners of English, phonological knowledge is more important than vocabulary and grammar knowledge. Findings of the current study differ from those of Wang and Treffers-Daller (2017) mainly because different variables were explored. As mentioned above, the current study investigated the contribution of learners' linguistic knowledge, sentence processing speed, cognitive ability, and frequency of the use of English in daily life. Although Wang and Treffers-Daller (2017) also explored the contribution of learners' linguistic knowledge, their study comprised only learners' vocabulary knowledge and grammar knowledge. So phonological knowledge was not included. If the two variables had been included their findings may have been different as phonological knowledge was found to be an important predictor in the current study. Although the current study explored the contribution of cognitive ability, it did not explore the contribution of metacognitive awareness as did Wang and Treffers-Daller (2017). Field (2013) built a model of meaning construction in listening and a model of discourse construction in listening where learners' listening strategies, e.g., inferencing and monitoring, were assumed to play roles in LC (see Chapter 1 for details). Future studies could include constructs from those higher-order processes in listening. Findings on the roles of linguistic knowledge in the two studies are similar in that such linguistic knowledge was found to be the strongest predictor of L2 LC.

Findings of the current study indicate that different amounts of variance are explained in either the SEM models or the hierarchical regression models, but the greatest variance is explained in one of the SEM models. The findings mean that a SEM model can explain more variance than a hierarchical regression model. A SEM model can include many more variables than a hierarchical regression model and is a more powerful way to analyse data (Schoonen, 2015). This is why the current study used this method of analysis. Since 65.7% of the variance was explained in one of the SEM models of LC, the residual 34.3% must be explained by other variables. What then would be the other variables which have not been measured? It is possible that remaining variance occurs in higher-level processes of LC, as assumed in Field (2013), but this was not explored in the present study. If non-linguistic knowledge, metacognitive strategies, and inferencing had been measured, more variance might have been explained and the SEM models of LC might have been different.

6.3 Lower-level processes in L2 listening comprehension

The operation of the present study was based on the model of lower-level processes in listening (Field, 2013, see Chapter 1 for details). In this model, learners' LC is the result of the interaction of three types of knowledge because according to Field (2013) learners have to update and constantly revise their provisional hypotheses of word, phrase, and clause with the continuous speech input. There is evidence for Field's (2013) model based on the results of my study. As can be seen in the Results Chapter, phonological knowledge, vocabulary knowledge, and grammar knowledge had high factor loadings onto linguistic knowledge (factor loadings >.70) which means that these three types of knowledge have strong associations with linguistic knowledge. The results of analyses indicated that linguistic knowledge is a key predictor of L2 LC and phonological knowledge is the strongest predictor when compared with vocabulary knowledge and grammar knowledge. When LC was measured with the PET listening, linguistic knowledge explained 36.3% of the variance in LC; while overall variance explained by phonological knowledge was 35.3%. When LC was measured with the CET4 listening, linguistic knowledge explained 17.9% of the variance in LC; while overall variance explained by phonological knowledge was 13.5%. Therefore, the assumption of the positive effect from phonological knowledge on success in Chinese learners' LC is confirmed.

According to Poelmans (2003), when heard words are recognised, the words are assigned to categories of grammar and then structural and semantic relationships of the words are established. Field's (2013) model of lower-level processes in LC is tested against evidence. The results of analyses show that vocabulary knowledge is a significant predictor on its own. The overall variance explained by vocabulary knowledge was 22.5% when LC was measured with the PET listening; while it was 13% when measured with the CET4 listening. Therefore, the important role of vocabulary knowledge in the listening processes in Field (2013) among Chinese learners of English is confirmed. Grammar knowledge is also a key predictor of LC among Chinese learners. It explained 2.2% of the variance in LC when measured with the PET listening and 3.3% of the variance in LC when measured with the CET4 listening. Therefore, the results of the current study regarding the contribution of L2 learners' linguistic knowledge to predicting LC support Field's (2013) theory about the lower-level processes in LC. It is confirmed that the sources of phonological knowledge, vocabulary knowledge, and grammar knowledge help Chinese learners comprehend speech input.

Although my data provides some evidence for Field's view of lower-level processes in listening I think the three processes can be adjusted. Our results show that phonological knowledge plays a key role at the input, lexical search and parsing stages, and therefore separating these out as per Field's model does not reflect ongoing speech processing accurately. Overlapping modules might better reflect the ways in which the different processes interact.

6.4 Evidence of Chinese learners' listening problems

Y. Wang (2008), Goh (2000), F. Wang (2008) and Xu (2014) all revealed that Chinese learners of English have many listening problems related to decoding sounds and words in the processes of LC. The current study made similar findings in that individual differences in phonological knowledge emerged with the strongest predicting power for LC. The findings help explain why learners found listening comprehension a source of frustration and found it hard to make progress with listening (Graham, 2011). The main reason is that many learners do not have the necessary phonological knowledge to segment words from the stream of speech, recognise words from speech, or distinguish similar sounds or phonemes.

The studies mentioned above show what learners or teachers perceived learners' listening problems to be, but they did not measure any listening processes. Nor did they measure learners' LC proficiency. Therefore, results of the studies did not reflect listening processes or the statistical associations between listening problems and listening proficiency. Although Sun and Li (2008) measured learners' listening proficiency, their test was a self-designed one which comprised two longer monologues. They used a retrospective questionnaire to investigate learners' listening problems and listening strategies. Those instruments differ from those used in the current study. Sun and Li (2008) found that between high-level and low-level language learners, learners had similar listening problems, but with significant differences in three particular types of problems out of a total of 13. The findings of the current study differed from Y. Wang (2008), Goh (2000), F. Wang (2008), Xu (2014) and Sun and Li (2008) because the current study used tests to explore Chinese learners' LC problems. The current study adopted online tests to assess learners' LC proficiency and learners' knowledge of segmenting words from the speech stream, knowledge of recognising words from speech, and knowledge of similar sounds or phonemes, and thus provides empirical evidence for Chinese learners' decoding problems in LC. It also showed how Chinese learners process speech input and what their listening problems are in the processes of listening. If learners do not have sufficient phonological knowledge they will find it difficult to understand input from speech in the L2. In addition, the difficulties caused by insufficient phonological knowledge are more than those caused by insufficient vocabulary or grammar knowledge. The reason why the lower-order phonological knowledge is more important than the vocabulary or the grammar knowledge might have been due to the fact that the Chinese language is so different from English, as presented in Chapter 2. The findings cast new insights into L2 learners' listening problems by providing empirical evidence regarding students' ability to process the incoming speech signal.

In addition to the issue of phonological knowledge, Su (2003) and Hu (2009) found other factors which also impacted Chinese learners' listening performance. Su (2003) found that learners' listening problems were related to a lack of linguistic knowledge, a limited working memory capacity, and learners' psychological state. The current study made similar findings about the role of learners' linguistic knowledge, but it revealed different findings for several reasons. Firstly, Su explored learners' listening problems with a self-designed questionnaire and the current study used well-established tests. Secondly, Su's data gathered from the questionnaire were not directly connected with learners' listening scores measured with preand post- tests. The data were only used to evaluate learners' choice of listening strategies. Hu (2009) found that 11 out of 13 listening problems were related to the three stages of processing put forward by Anderson (1995); while the other two problems were related to emotional and psychological factors. Hu (2009) explored learners' listening problems through introspection and self-reports which indicated what learners perceived their listening problems to be. However, it is possible there is a discrepancy between self-reported problems in listening and real listening problems during the online listening processes. The current study revealed different findings because learners' listening problems in listening processes were measured. If Hu (2009) and Su (2003) had used different tools to explore learners' listening problems, their findings would have been different. Zhang et al. (2010) provided empirical evidence on Chinese learners' listening problems. They found that Chinese learners' listening problems mainly lay in the parsing stage. They also found that learners' scores concerning their perceptions of listening problems explained 15.9% of the variance in LC. The current study did not support Zhang et al. (2010) for several reasons. Firstly, the current study measured learners' listening problems with tests, but Zhang et al. (2010) investigated learners' listening problems using a self-designed questionnaire in which learners self-rated their perceptions of listening problems. It is quite possible that learners' perceptions of their problems do not respond to their real problems in online listening processes. Secondly, in the current study additional tasks were adopted to explore learners' listening problems, e.g., tasks to measure learners' decoding problems and a questionnaire to measure learners' frequency of use of the L2 in daily activities. These were not used in Zhang et al.'s study. If they had used additional tasks, they may have found different listening problems faced by Chinese learners.

6.5 Learner variables in success of L2 listening comprehension

The current study found that the key learner variables to predict L2 LC were linguistic knowledge, frequency of the use of English in daily life, and sentence processing speed.

Matthews and Cheng (2015) found that learners' ability to recognise words from speech explained 54% of the variance in LC. The current study supports the finding that recognition from speech is the most important ability in predicting LC of Chinese learners. Although both studies shared the test of word recognition from speech and investigated adult Chinese learners of English, the two studies differ mainly because additional learner variables were explored in the current study. The current study not only found other learner variables, e.g., frequency of the use of English in daily life and processing speed, to be predictors, but also found that Chinese learners' ability to recognise words from speech is more important than their ability to segment words from the speech stream and to distinguish similar sound or phonemes. These findings have important pedagogical implications for teaching English listening to adult Chinese learners: teachers of English should pay more attention to improving Chinese learners' ability to recognise words from speech, a point elaborated on in the concluding chapter.

Bonk (2000) found that learners' linguistic knowledge explained 23% of the variance in L2 LC. Findings of the present study are similar to Bonk's finding on the roles of the linguistic knowledge in predicting LC. However, the present study also differs from that of Bonk in that additional variables were explored and linguistic knowledge was found to be the strongest predictor amongst all others. However, since the contribution of only one variable was explored in Bonk (2000), there was no way to investigate whether linguistic knowledge had the strongest predicting power for LC. In addition, results from Bonk (2000) indicated that L2 learners' individual differences in linguistic knowledge were not enough to explain variation in LC, so variance in LC must be explained by other factors (Macaro et al., 2016). The current study might have filled a gap in this respect by exploring the contribution of the frequency of the use of English in daily life, sentence processing speed, reasoning ability, and working memory to LC and found that L2 use and sentence processing speed are also predictors of LC. Comparatively, the variance explained by all the learner variables in the current study is more than that explained by Bonk (2000). The importance of linguistic knowledge in LC was also explored in Staehr (2008). The researcher found that participants' linguistic knowledge contributed 39% of the variance in L2 LC. The current study and Staehr (2008) share similarities about the contribution of linguistic knowledge to predicting LC, however, there are differences in the findings between the two studies for which there may be several reasons. Firstly, learners' L1s differed. The current study investigated Chinese learners of English, but Staehr (2008) investigated Danish learners of English. Chinese is a Tibetan language and Danish is an Indo-European language, so they are two completely different languages. Secondly, learner variables explored in the current study differed from those in Staehr, as they differed from Bonk (2000). Thirdly, different tools were adopted to measure learners' linguistic knowledge. In the current study, learners' linguistic knowledge comprised three sources of knowledge, but in Staehr (2008) linguistic knowledge only included vocabulary knowledge which was measured with a written receptive vocabulary test. In the current study, learners' vocabulary knowledge was measured with an aural vocabulary test. These differences help explain why the correlations between the vocabulary test and LC between the studies differed. The correlation (r = .69, p < .001) found in Staehr is stronger than that found in the current study (r = .47, p < .01 when LC was measured with the PET listening; r = .36, p < .01 when LC was measured with the CET4 listening). Staehr suggested that if he had used a measure of phonological vocabulary size or aural vocabulary, stronger correlations would have been found. Although the correlations between vocabulary knowledge and LC found in the current study were not high, an aural vocabulary test required learners to process words from speech – the same pattern as with processing listening texts. Therefore, findings of the current study make a new contribution to research on L2 listening.

Very few studies have used aural vocabulary tests to explore individual difference in explaining success in L2 LC. Milton et al. (2010) found that the correlation between the scores of an aural vocabulary test and LC (r = .67) was stronger than that between the scores of a written vocabulary test and LC (r = .48), but scores of the aural vocabulary test explained less variance in L2 LC (44%) than that explained by the written vocabulary test (51%). Their findings suggest that although an aural vocabulary test can have stronger correlation with LC than a written vocabulary test, it is not necessarily the case that this explains more variance in listening than a written vocabulary test (Wang & Treffers-Daller, 2017). Although different aural vocabulary tasks were used in the current study and in Milton et al. (2010), both studies found a contribution of vocabulary knowledge to predicting L2 LC, which means that aural vocabulary tasks are effective tools to measure L2 learners' vocabulary knowledge in L2 LC. However, the correlations between scores for the aural vocabulary task and LC in the current study were lower than those in Milton et al.. The main reason for the differences might be that different aural vocabulary tasks and listening tasks were used. In the current study, LVLT (McLean et al., 2015) was used to measure learners' vocabulary knowledge in which learners heard each short carrier sentence with a target word and chose the word meaning in

Chinese from four alternatives. In Milton et al. (2010) A_Lex (Milton & Hopkins 2005) was used to measure learners' vocabulary knowledge. Learners heard each target word and decided whether they knew it or not. In that study, LC was measured with an IELTS listening and in the current study, LC was measured with the CET4 listening and the PET listening. Matthews (2018) also used an aural vocabulary task to measure Chinese learners' vocabulary knowledge and explore its contribution to predicting LC. He found that vocabulary of the high frequency levels (0 - 2000, 2001- 3000) and the mid-frequency level (3001 - 5000) explained 52.8% of the variance in LC. Results in Matthews (2018) suggest that aural vocabulary knowledge throughout both the high frequency levels and the mid-frequency level is an important predictor of success in L2 LC. Although a receptive vocabulary task was used in the current study and a productive task was used in Matthews (2018), findings from the two studies provide further empirical evidence that much of the variance in LC can be attributed to phonological forms of words.

Since vocabulary size is not the only component of vocabulary knowledge, some studies have explored contribution of both vocabulary size and vocabulary depth to explaining the variance in L2 LC. Staehr (2009) found that linguistic knowledge explained 51% of the variance in L2 LC. Findings in the current study revealed similarities about the strength of linguistic knowledge in predicting LC. More than half of the variance in LC was explained by linguistic knowledge in Staehr (2009) which means that it has the strongest predicting power. However, findings from the current study differed from those in Staehr's (2009) study mainly because linguistic knowledge has different components. As mentioned above, linguistic knowledge (comprised of three types of knowledge in the current study) and phonological knowledge had the strongest predicting power to LC compared to the others. In Staehr (2009) linguistic knowledge comprised vocabulary size and vocabulary depth. Staehr found that vocabulary size accounted for 49% of the 51% of the variance and vocabulary depth accounted for 2% unique variance in LC. Another reason for the differences in the findings between the two studies might be that different tests were adopted to measure learners' LC and vocabulary knowledge. The PET listening and the CET4 listening tests were used to measure learners' listening levels in the current study and in Staehr listening was measured through the Cambridge certificate of proficiency in English. In Staehr both vocabulary size and vocabulary depth were assessed using a written vocabulary task. If aural vocabulary tasks had been used in Staehr the findings might have been different. Zhang (2011) made similar findings about the stronger role of vocabulary size than that of vocabulary depth in predicting LC. Zhang found that the former explained 27% of the variance in LC and the latter explained 2% of the variance in LC. Although the current study and Zhang both investigated Chinese learners of English, findings differed because different tasks were used to explore learners' linguistic knowledge and additional learner variables were explored. In Zhang (2011), learners' vocabulary size was assessed through a written receptive task which required learners to process word meaning in written form. Learners' vocabulary depth was assessed through a productive vocabulary task which required learners to produce words using correct spelling, singular or plural form, and tense. However, the current study required learners to process English words aurally and provide empirical evidence that much variance in LC can be explained by the phonological forms of words.

As mentioned in Section 6.2, one of the reasons that the SEM models of LC built in the current study differ from that in Andringa et al. (2012) is that the frequency of L2 use was added to the model that they developed. The current study found frequency of the use of English to be a key predictor of LC. The variable uniquely explained 5.8% of the variance in LC as measured with the PET listening and it explained 5.1% of the variance in LC as measured with the CET4 listening. The results mean that those who use English frequently in daily activities are likely to have better LC performance than those who do not use English frequently. More evidence for the fact that language use is important can be found in Allen and Herron (2003). Their study showed that after a summer study abroad programme L2 learners' listening proficiency was improved significantly. This was reflected in a comparison between pre- and post-test listening scores and from learners' self-reports in questionnaires and interviews. Findings of the current study are in line with Allen and Herron (2003) concerning the positive roles of an L2 immersion environment in impacting learners' listening. Llanes and Botana (2015) found that adult L2 learners' LC was improved as a result of a six-week study abroad experience in which learners had frequent interactions with native monologue speakers and L2 was required for the negotiation of meaning. The findings of the current study are in line with those studies concerning the important role of living in an immersion language environment in influencing LC performance.

Cubillos et al. (2008) found that, in general, study abroad groups do not show significant differences in terms of higher gains in L2 LC than their peers who study on a home campus. The researchers argued that to consider the impact on LC as a result of foreign language immersion the following factors should be included: longer (a semester or a year) study abroad courses; and the use of different LC measurements and listening strategy techniques. The current study included the first two factors and made different findings about the roles of foreign language immersion in impacting L2 LC. In the current study, Chinese learners in the UK had a study abroad experience of between 12 to 36 months and their LC proficiency was significantly higher than that of learners in China. Comparatively, L2 learners in Cubillos et al. (2008) had shorter, i.e., five week, duration of study abroad experience. There are other reasons which might have caused the differences in the findings between the two studies. The first is that learners in China received instruction in English mainly in English classes which they attended for less than 200 minutes in each week in each academic term, but learners in the UK received instruction through the medium of English in all their courses, in addition to living in an immersion language environment. In Cubillos et al. (2008), learners on the home campus and learners abroad shared the same syllabus, textbooks, and ancillaries. They had the same number of instructional hours and had been instructed with the same pedagogical approach. Although learners abroad experienced an immersion language environment, the differences in the effect of that environment on improving LC might have been reduced by receiving the same instructions as the home campus learners. The second reason is that the variable of L2 use had components of frequency of English use in daily activities in the current study. Learners in the UK had significantly higher frequencies of the use of English than did learners in China. In Cubillos et al. (2008), the impact of an immersion language environment was assessed by comparing listening scores between the study abroad group and the home campus group. They found that both groups had significant gains on two out of three sections of the LC test, but the study abroad group made equal progress in listening development with the home campus group. However, learners' questionnaire responses to some items about the impact of their study abroad experience suggested that study abroad groups demonstrate higher confidence levels in L2 listening skills and had higher self-perceived ability than home campus students after the short course. Although different questionnaires were used in Cubillos et al. (2008) and the current study, the findings on this point were similar. In addition, the current study provides empirical evidence that studying in an immersion language environment has a positive effect on LC performance.

In addition to the linguistic knowledge and the frequency of L2 language use, the current study found a third predictor of LC: sentence processing speed. The variable explained 2.5% of the variance in LC. Andringa et al. (2012) found that although processing speed was significantly correlated with LC (r = -.67) for non-native speakers, it did not explain any unique variance in success of LC. The researchers estimated that this was because the knowledge variable or the cognitive factor, or both, mediated the role of processing speed. The findings of the current study do not support Andringa et al. on the roles of processing speed in predicting L2 LC. This might have been due do the fact that the tasks used to assess learners' processing speed and LC differed between the two studies, as seen in Chapter 3. In their study, Andringa et al. found processing speed to be correlated with linguistic knowledge (r = -.68) for L2 learners. Therefore, they concluded that the non-native listeners who could process input language information quickly and efficiently also had greater linguistic knowledge. What was found in the current study supports that conclusion. The results showed that sentence processing speed is correlated with linguistic knowledge although the correlation was weaker (r = -.10, see Appendices 149 & 176) compared to Andringa et al.'s findings. This result suggests that learners' processing speed was not as strong a predictor as the phonological variables.

The role of processing speed in contributing to explaining the variance in L2 LC was also explored by Oh (2016). The researcher found that processing speed is not a predictor of LC. The findings of the current study do not support Oh (2016). There may be two reasons for this. Firstly, learners' L1 differed between the two studies. The current study looked at Chinese learners of English, but Oh (2016) looked at Korean learners of English. Secondly, Oh's research only took accuracy measures, not RTs. However, learners' accuracy rates to items might not have reflected how fast they process sentences heard. Differing from Oh (2016), in the current study learners' accuracy and RTs were recorded automatically by computer software and only RTs from items learners answered correctly were used.

The current study found neither learners' working memory capacity nor reasoning ability to be a key variable in L2 LC. The findings suggest that learners who have larger working memory capacity do not necessarily have higher LC proficiency. Although different working memory tasks were used in the current study and in Andringa et al. (2012), the findings on the impact of working memory on L2 learners' LC confirmed the latter's findings. They estimated that the reason working memory was not found to be a predictor of LC might be that learners' working memory scores did not reflect their experience with L2. The visual working memory tasks allowed for the use of L1 and in the auditory digit span tasks it could not be assured that learners had not translated numbers into their L1. In addition, the non-word recognition task did not consist of correct L2 words which makes it likely that the scores on this task did not reflect their L2 experience. These might be the same reasons why working memory capacity did not contribute to explaining the variance in LC in the current study. When working memory capacity was measured with digit span tasks in Chinese, scores on the tasks did not reflect learners' experience with L2; when working memory capacity was measured with tasks in L2, one cannot be certain that learners had not translated numbers into their L1. Therefore, in future the use of a different working memory task that does not rely on numbers should be considered. Another reason might be that the role of working memory capacity is eclipsed by learners' language competence, which means that a lack of understanding of the language played such an important role in LC that some of the other variables did not have the chance to become significant, e.g., working memory capacity. Vandergrift and Baker (2015) found that working memory was related to L2 LC, but a significant correlation was found in only one out of three cohorts ($r = .37$, $p < .05$). For the combined cohorts, working memory was not found to be significantly correlated with LC. The two researchers concluded that, as a general skill, working memory is initially important in leading to more specific language skills in success of L2 LC. However, they are not certain whether working memory impacts L2 LC only through L1 and L2 vocabulary knowledge. Results of the current study show that although when LC was measured with the PET listening, the four digit span tasks were significantly correlated with LC (which was comparable with that found in one of three cohorts in Vandergrift and Baker's study where WM was not found to predict the success in L2 LC). Findings on the significant correlations between digit span tasks and LC are in accordance with Brunfaut and Révész (2015). Similar results on the correlation between backward digit span and LC were found in Kormos and Safar (2008). Although Kormos and Safar (2008) found that working memory explained

30.25% of the variance in learners' performance on a language test comprising reading, listening, speaking, and writing sections, the current study found that working memory is not a key predictor of LC. The above discussion on the correlation between working memory capacity and L2 LC and the role of working memory in explaining variance in LC means that working memory is weakly but significantly correlated with L2 LC, but working memory is not a key predictor of LC as found in the models built in the current study and in the SEM model of LC for native speakers and L2 learners in Andringa et al. (2012).

In addition to working memory, reasoning ability is the other subcategory of learners' cognitive ability in the current study. Reasoning ability was not found to be a key variable in L2 LC. The findings about this variable differed from those in Andringa et al. (2012). They found that learners' reasoning ability is a key variable and it could predict L2 LC. Reasoning ability, together with linguistic knowledge, explained 96% of the variance in LC in their study. Since the roles of reasoning ability in L2 learners' LC performance have not received substantial attention, the different findings between the two studies have provided empirical evidence and cast new light in the field of L2 listening.

6.6 Between group differences in LC

A strength of the current study lies with the findings about the between group differences in LC. The study found that LC of Group 1 was significantly lower than that of Group 2, as presented in the Results Chapter. For Group 1, the key variables were learners' word recognition from speech, distinguishing phonological sounds and sentence processing speed. Phonological knowledge which comprised of word recognition from speech and distinguishing phonological sounds was much more important than sentence processing speed. For Group 2, the key variables were learners' grammar knowledge and reasoning ability. Grammar knowledge was much more important than reasoning ability. The results mean that learners in China who have greater ability to recognise words from speech and distinguish phonological sounds are likely to have higher L2 LC. The findings also reflect the importance of automaticity in processing speech input because L2 listeners have gaps in L2 language knowledge and such gaps have significant impact on L2 LC (Buck, 2001). Field (2008) holds that novice and expert listeners majorly differ in whether they command a set of highly automatic decoding routines, or not. Therefore, the findings about the individual differences in automaticity among learners in China indicate that processing speed had an impact on their listening performance. The findings that learners' word recognition from speech and distinguishing phonological sounds had a stronger predicting power in LC indicate that the lower-order phonological knowledge was more important than vocabulary or grammar knowledge for learners in China. This was not the case for Group 2. Word recognition from speech was not a predictor of LC for learners in the UK. The reason for the different roles of recognising words from speech between the two groups might have been due to the fact that those learners in the UK are immersed in an L2 environment, thus they use L2 more frequently than do learners in China. The latter generally lack an L2 environment, which means that they are not used to segmenting speech streams, distinguishing similar phonological sounds, or recognising words from speech. The findings mean that the ability to recognise words from speech is more important for learners in China than that for learners in the UK. Although grammar knowledge is a predictor of LC for learners in the UK, it is not for learners in China. The findings mean that grammar knowledge is more important for learners in the UK than for learners in China. Since Group 2 had higher LC proficiency than Group 1 and reasoning ability was found to explain some variance in LC in Group 2, the findings suggest that in addition to linguistic knowledge, reasoning ability can predict LC of higher-level learners. Reasoning ability was not found to be a key variable to predict LC for the entire group, but it was a key predictor of LC for Group 2. Comparatively, sentence processing speed, in addition to linguistic knowledge, can predict LC of lower-level learners. The findings confirm the important roles of processing speed and linguistic knowledge in explaining success in L2 LC.

6.7 Comparing the CET4 listening and the PET listening

Since one of the aims of the present study is to provide suggestions to Chinese learners of English, teachers of English, and CET4 listening test designers in China, it is necessary to analyse the task in detail. By comparing the CET4 listening and the PET listening, the current study found several differences between the two tests. Firstly, the PET listening was found to better explain variance than the CET4 listening, as seen in Chapter 5. Secondly, the PET listening measured more of the abilities learners required for listening than did the CET4 listening. The PET listening measured learners' abilities to apply linguistic knowledge to processing speech input and to apply world knowledge and strategies, e.g., using inferencing to identify explicit and implicit information when processing information heard. It actually measured the learners' abilities which were given in the test specifications, as presented in Chapter 3. However, in the CET4 listening, all 25 items required learners' lower-order processing; none required learners' higher-order or interactive processing while the questions were answered. For lower-order processing, listeners can answer the questions correctly as long as they have heard the original words or sentences from the aural input. The findings mean that only abilities mentioned in A in the test specifications were really measured, i.e., the ability to comprehend explicit information. However, the abilities to distinguish sound differences or to comprehend sound stress and causal or contrasting relations between sentences, to comprehend implicit information and to adopt listening strategies were not measured. Field (2013) holds that listening comprehension is a process where the lower-level processing and the higher-level processing function interactively. Therefore, the failure to measure learners' abilities to distinguish sound differences, to comprehend sound stress and causal or contrasting relations between sentences, to process implicit information and to adopt strategies when processing speech input could be major problems of the CET4 listening. The findings suggest that CET4 test designers reform the listening section, in terms of its form and texts, and design items which require learners to activate abilities in lower- and higher-order processing in LC. Thus, the current study makes a new contribution to the L2 testing field.

6.8 Implications of the results

I am aware that if I had used different tasks to measure the variables, the results would have been different. But the tasks are very similar to what other researchers used in the field of L2 listening and similar results were obtained with these tasks.

The group of participants in China and the group in the UK differ from each other in that the latter had to have been in the UK for at least one year prior to data collection. We have to acknowledge the fact that the one-year residence is an arbitrary cut off point and that participants in the UK vary greatly in the time they have spent in the UK. This group was included to allow for a wide range of contact with English, which would not have been possible if only Chinese students in China had been included: their contact with English is necessarily much more limited. The UK-based learners had the opportunity to listen to and use L2 frequently which will have helped them to practice recognising words from speech and to increase their ability to process speech input quickly.

The L1 of participants is really important. As mentioned in Chapter 2, Chinese and English are very different languages. For example, Chinese sounds have more limited coda inventory, simpler syllabic structures (CGVX) and no consonant clusters (Lin & Wang, 2018). English polysyllabic words are common and these can be based on a variety of syllable structures. It is possible that Chinese learners of English do not always realize that stress provides important information for speech segmentation. Some sounds in English do not exist in Chinese. Such sounds include /θ/ /ð/ /ʒ/ or /v/ (Zhang et al., 2005). Therefore, Chinese learners struggle in processing words containing those sounds. The differences between the two languages will truly affect participants' ability to process speech because the learners have to learn a very different phonetic system and different intonation patterns.

Proficiency of participants generally impacts on their performance in tasks. In this study, the participants in China were at approximately B1 level and the participants in the UK were at approximately B2 level on the CEFR scale of English proficiency. We have found that the listening proficiency of the participants in China was significantly lower than that of the participants in the UK. The differences in listening proficiency between the two groups can be attributed partially to the differences in English proficiency as the results in Chapter 5 indicate that the mean scores of all tasks except the task of processing speed for the group in China were lower than those for the group in the UK.

Since the study focuses on the lower-order processing of L2 LC, predictors that represent higher-order processing, e.g., world knowledge and metacognitive strategies, are not explored. The absence of such predictors is a limitation of the study. Although we have found that 65.7% of the variance in LC can be explained by the five learner variables investigated, still 34.3% of the variance in LC is not explained. Therefore, if other learner variables which tap into higher-order processing had been included, more variance could have been explained.

Although it is common to subdivide proficiency into subcomponents (vocabulary, grammar and phonology), it is becoming increasingly clear that separating these out is problematic. For example, when we use a segmentation task to measure learners' phonological knowledge, we have to use words for the task. Many researchers, including Wang and Treffers-Daller (2017), have found that vocabulary is a key determinant of listening. But as soon as phonetic knowledge is included in a test battery measuring the effects of different variables on listening, vocabulary knowledge might turn out to be less important than phonetic and phonological knowledge. However these two types of knowledge are of course linked: vocabulary knowledge covers knowledge about the form of words and therefore covers phonetic and phonological information. According to Halliday (1994), grammar and vocabulary cannot be separated completely because "grammar and vocabulary are merely different ends of the same continuum – they are the same phenomenon as seen from opposite perspectives" (p. 15). In a similar vein, one could argue that separating vocabulary and phonetic/phonological knowledge is difficult. Evidence for this comes from studies investigating the link between vocabulary size and phonological reorganization in adult L2 learners. In a study among adult Japanese L2 learners of English who had been learning English for about twelve weeks in Australia, Bundgaard-Nielsen, Best and Tyler (2011) show, for example, that students with a larger vocabulary were better at identifying L2 vowels than those with smaller vocabularies. The authors therefore call for further studies of the relationship between L2 vocabulary growth and L2 phonological acquisition. The current thesis has the potential to contribute further to the contribution of vocabulary knowledge on the perception of L2 vowels. These insights can also be used to reach more in-depth analyses of L1 Chinese learners' problems with L2 listening.

Chapter 7 Conclusion

This study built two structural equation models and two regression models of L2 LC among Chinese learners of English as the entire group. The findings on which learner variables can explain the variance in LC and to what degree the learner variables can predict it fill some gaps in our understanding of the contribution of individual differences to predicting success in L2 LC. The study also found which learner variables can explain the difference in LC between Chinese learners in China and in the UK. The findings have important implications for pedagogy in China and in the UK, and have implications for test development in HE in China.

In this chapter, the main findings of the study are summarised in Section 7.1, followed by the limitations explained in Section 7.2. The third section presents the implications for pedagogy in China and in the UK and the final section presents the implications for test development in HE in China.

7.1 Main findings

Findings for the entire group

The study found that for one SEM model, when LC was measured with the PET listening, together the five groups of factors (linguistic knowledge, aural sentence processing speed, working memory capacity, reasoning ability, and frequency of the use of English) explained 65.7% of the variance in LC among adult Chinese learners of English. Linguistic knowledge and frequency of the use of English were key predictors of LC.

The study found that for the other SEM model, when LC was measured with the CET4 listening, together the five groups of factors explained 35.4% of the variance in LC. Linguistic knowledge, frequency of the use of English, and sentence processing speed were key predictors of LC. Linguistic knowledge was the most important predictor of L2 LC for both dependent variables.

The study found that among the different aspects of linguistic knowledge that were tested as part of the study, phonological knowledge was the most important predictor of LC and explained more variance in LC than vocabulary or grammar knowledge. The findings in this respect are quite different from many studies which found that vocabulary knowledge was more important as a predictor of L2 learners' LC. The study also found that a key aspect of lexical knowledge is knowledge about the form of words. The findings suggest that when teaching listening to Chinese learners, the focus should be on improving learners' phonological knowledge.

Chinese learners' greatest listening problem is recognizing words from speech, with word recognition explaining the variance in LC over and above the contribution of segmenting words from speech. The findings in this respect suggest that in the teaching of English listening to Chinese learners, more focus should be put on improving learners' ability to recognize words from speech than on segmenting words from the speech stream, although in order to recognize a word learners first have to segment words.

The study found that in one of the regression models, when LC was measured with the PET listening, all observed variables explained 44.3% of the variance in LC. Linguistic knowledge, phonological knowledge, frequency of the use of English, vocabulary knowledge and grammar knowledge were key predictors of L2 LC.

The study found that in the other regression model, when LC was measured with the CET4 listening, together all observed variables explained 28.2% of the variance in LC. The key predictors were linguistic knowledge, phonological knowledge, frequency of the use of English, sentence processing speed, grammar knowledge, and vocabulary knowledge. For both dependent variables, linguistic knowledge was more important in predicting success in L2 LC than the other variables.

The findings of the SEM models and the regression models of LC for the entire group help us understand which individual differences can contribute to explaining success in L2 learners' LC. The findings contribute to our understanding of which learner variables are more important than others.

The findings on the positive impact of frequency of the use of L2 in daily life on LC provides empirical evidence that immersion in an L2 environment plays a very important role in influencing learners' listening proficiency. The findings suggest that learners who use L2 frequently in daily activities are more likely to have better LC performance than those who do not use L2 frequently.

Findings for subgroups

Important differences were also found when the informants were split in two groups: a group of learners in China (Group 1) and a group of learners in the UK (Group 2). As might have been expected, Group 1 obtained significantly lower scores on the listening comprehension than Group 2. The findings indicate that for Group 1, the key variables which explained LC were learners' phonological knowledge which comprised of word recognition from speech and distinguishing phonological sounds, and sentence processing speed; phonological knowledge was much more important than sentence processing speed. For Group 2, the key variables ware learners' grammar knowledge and reasoning ability; the former was much more important than the latter.

Findings on analysis methods

The study found that different amounts of variance are explained in either the SEM models or the hierarchical regression models, but the biggest amount of variance is explained in one of the SEM models. The findings suggest that a SEM model can explain more variance than a hierarchical regression model. However, since 65.7% of the variance was explained in one of the SEM models of LC, the 34.3% residuals would need to be explained by other variables. It is possible that the remaining variance is related to higher-order processes involved in LC, as assumed in Field (2013), but these were not included in the present study.

Findings on the similarities and differences between the current study and Andringa et al. (2012)

The findings of the current study and those of Andringa et al. (2012) on L2 LC share both similarities and differences. Both studies built one or two structural equation models of LC and found learner variables which could explain success in L2 LC. Findings of both studies provide empirical evidence for the role of L2 learners' linguistic knowledge in predicting L2 LC. Linguistic knowledge was found to have the strongest predicting power among the other learner variables in L2 LC. Differing from Andringa et al. who found that the explained variance was 96%, the current study found it was 65.7%. This difference means that for Chinese learners of English, there remain some other learner variables which can explain the variance in LC and it is worth exploring what these may be. Another difference between the two studies is that learners' aural sentence processing speed was found to be a key predictor of L2 LC in the current study, but not in Andringa et al.. They found that this learner variable was a predictor of L1 LC, but not of L2 LC. The third difference between the two studies is that learners' frequency of the use of L2 was added to the SEM models of LC in the current study and was found to be a key predictor of L2 LC.

7.2 Limitations of the study

In the present study some limitations could not be avoided, which is not uncommon among other related studies. Firstly, learners' linguistic knowledge, aural sentence processing speed, and working memory capacity were measured with listening tasks. The variables used to explain the dependent variable (listening comprehension) are, therefore, very close to the dependent variable itself, which might give the impression of circular reasoning. However, in order to obtain further insights into the component processes involved in listening, using some aural tasks was unavoidable. In addition, tasks in which listening is not involved to measure learners' linguistic knowledge were included. For example, I used the OPT grammar section, which is a written task, to measure learners' grammar knowledge, together with the aural task of TROG-2. I used listening tasks to measure learners' phonological knowledge as learners' ability to distinguish similar syllables, to segment the speech stream, and to recognize words from speech can only be measured with listening tasks. Although it worked well to use written tasks to measure learners' vocabulary knowledge in other studies (e.g., Andringa et al., 2012; Mecartty, 2000; Staehr, 2008, 2009; Wang & Treffers-Daller, 2017), not so many studies tried aural tasks to measure learners' vocabulary knowledge. Therefore, I tried this in the current study and I found that it worked well because the findings indicate that a considerable amount of variance in LC is attributed to phonological factors. As for the sentence processing speed, this can only be measured during online processing. Therefore, listening tasks were unavoidable. In the current study, I also used written tasks to measure learners' reasoning ability and to investigate learners' personal language learning background. I know that comparatively the number of written tasks used in the study is fewer than the number of aural tasks; in future studies, researchers should try to avoid this problem and find a way to balance the number of written and aural tasks for listening studies.

Secondly, in the current study the numbers of participants in Group 1 (N = 147) and Group 2 (N = 40) are not balanced, clearly Group 2 contained considerably fewer participants than Group 1. This problem is attributed to the limited research time. The data from Group 2 in the UK were collected by me alone with one or two participants taking part during each session. Comparatively, it took much more time to collect data among participants in Group 2 than among participants in Group 1. It was easier for the participants in each individual group in Group 1 to find a suitable time to complete all the tasks, which helped to shorten the processes of data collection in China. Although it would have been better to have equal numbers of participants in each group this was not feasible in the current study.

Thirdly, this study focuses on the lower-order processes in LC because the previous literature shows that Chinese learners have many listening problems in the perception stage of comprehension and these problems might have been caused by the typological differences between Chinese and English. However, LC involves the interaction of learners' abilities to use linguistic knowledge and world knowledge in processing speech input. Although I have found that the variables included in the study can explain more than half of the variance in L2 learners' LC, still 34.3% of the variance in LC is not explained by the five learner variables. Therefore, future studies should focus on both lower- and higher-order processes of LC. For the lower-order processes future studies should focus on measuring learners' phonological knowledge and more focus should be put on word recognition from speech than on word segmentation. When vocabulary knowledge is included, studies should measure learners' knowledge of phonological forms of words. For the higher-order processes, the variables regarding non-linguistic knowledge (Chen, Lee, & Lin, 2010), metacognitive strategies (Vendergrift & Baker, 2015), and inferencing (Valentini & Serratrice, 2018), which have been found to be important in predicting L2 LC, should be explored. Adding a qualitative perspective could bring additional insights – e.g., since the study found that word recognition is the key difficulty, a qualitative exploration of how learners are trying to make sense of words that they can segment from the speech stream, but cannot recognize, would give a richer perspective.

7.3 Implications for pedagogy in China and the UK

In the introduction to the thesis it was mentioned that, as a teacher of Chinese learners of English, my students often asked why they found it hard to improve their English LC and how this problem could be ameliorated. The findings of the study suggest that phonological knowledge has the greatest impact on Chinese learners' LC. If a learner lacks linguistic knowledge, or has slow sentence processing speed, or does not use the English language frequently, LC proficiency is affected. The findings of the study provide answers to the question on why it is hard for L2 learners to improve their LC. As for the question on how these learners can improve their LC, suggestions can be provided based on the findings of the study. For Chinese learners of English in China, it is important they are provided with opportunities to improve their phonological knowledge and to improve their speed to process aural input, e.g., fluency training. Teachers should strengthen their input of phonological knowledge in English classes and develop phonetics/phonology exercises so that students have opportunities to practice recognizing sounds which belong to the English sound system, but which are not included in the Chinese sound system. Therefore, textbooks need to be changed and teachers need to be trained so that they can teach phonetics. Teachers should help students to improve their abilities in distinguishing similar sounds/syllables, segmenting the connected speech stream, and recognizing words from speech. Since, generally, learners of English in China live in a Chinese context, this means that they lack an English context and this might result in the lower processing speed of aural input information. There are practical barriers to an immersion in the English language, but learners are encouraged to use English in daily activities after class as much as possible. For Chinese learners of English in the UK, learners' grammar knowledge is found to significantly contribute to explaining the variance in LC. Therefore, the study suggests that language teachers teach Chinese learners grammar knowledge and develop grammar exercises so that learners have opportunities to practice grammar knowledge exercises in written and aural forms.

7.4 Implications for test developers in HE in China

Since one of the purposes of the study is to formulate implications for teachers of English, learners of English, and test developers in China, the CET4 listening was analyzed and compared with the PET listening. The current study finds that the CET4 listening explains much less variance in LC than the PET listening. In addition, the CET4 listening measures fewer of the abilities learners require for listening comprehension than does the PET listening. The findings suggest that only abilities mentioned in A in the test specifications were really measured. The findings mean that the CET4 listening measures learners' ability to comprehend explicit information (themes, key or specific information, speakers' viewpoints or attitudes which are conveyed explicitly). However, it does not measure learners' ability to distinguish sound differences or to comprehend sound stress and causal or contrasting relations between sentences. It does not measure learners' ability to comprehend implicit information (inferencing implicit information, deciding communicative functions of speech, inferencing speakers' opinions and attitudes), or the ability to adopt listening strategies to help with comprehension. The three types of ability are also specified by the CET-4/6 Testing Committee (2016). Comparatively, the PET listening measures learners' ability to identify key information and detailed meaning, the ability to interpret information, and the ability to identify attitudes and opinions of speakers, as required in test specifications. On the one hand, the findings on what both listening tests have really measured can help test developers in China identify what requirements for Chinese learners' listening ability are not included in the present CET4 listening. On the other hand, the findings suggest that test designers of the CET4 listening need to reform the listening section in its forms and texts and to design items which target learners' abilities for lower- and higher-order processing in LC with reference to other international listening tests, e.g., the PET listening. Thus, the current study has implications for listening test development in HE in China.

The findings suggest a need to design the listening texts and the listening questions in the CET4 so that they measure learners' ability to comprehend both explicit and implicit messages, including inferencing implicit meaning and speakers' opinions and attitudes. Listening questions should include items which require learners to distinguish similar phonemes, to segment words or phrases from the speech stream, and to recognize words from speech. Listening questions should also include items which require learners to infer implicit information.

References

Adolphs, S., & Schmitt, N. (2003). Lexical coverage in spoken discourse. *Applied Linguistics, 24*, 425-438.

Akker, E., & Cutler, A. (2003). Prosodic cues to semantic structure in native and non-native listening. *Bilingualism: Language and Cognition, 6*, 81 – 96.

Allan, D. (2004). *Oxford Placement Test 1*. Oxford: Oxford University Press.

Allen, H., & Herron, C. (2003). A mixed-methodology investigation of the linguistic and affective outcomes of summer study abroad. *Foreign Language Annals, 36*, 370 – 385.

Altenberg, E. (2005). The perception of word boundaries in a second language. *Second Language Research, 21*, 325-358.

Anderson, J. R. (1995). *Cognitive psychology and its implications (2nd edition)*, New York: Freeman.

Anderson, J. R. (2000). *Cognitive psychology and its implications (5nd edition)*, New York: Freeman.

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77-117). Newark, DE: International Reading Association.

Andringa, S., Olsthoorn, N., Beuningen, C., Schoonen, R., & Hulstijn, J. (2012). Determinants of success in native and non-native listening comprehension: An individual differences approach. *Language Learning, 62*, 49-78.

Awang, Z. (2012). *A handbook on structural equation modelling using AMOS*. Malisia: Universiti Technologi MARA Press.

Bachman, L. F., & Palmer, A. (2010). *Language assessment in practice*. Oxford: Oxford University Press.

Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*, 189-208.

Baddeley, A. D. (2007). *Working memory, thought, and action*. Oxford: Oxford University Press.

Bentler, P. M. (2005). *EQS 6 Structural equations program manual*. Encino, CA: Multivariate Software Inc.

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

Bishop, D. V. M. (2003). *The Test for Reception of Grammar version 2*. London: Pearson.

Bonk, W. (2000). Second language lexical knowledge and listening comprehension. *International Journal of Listening*, *14*, 14-31.

Boomsma, A. (1987). The robustness of maximum likelihood estimation in structural equation models. In P. Cutance & R. Ecob (eds.), *Structural modelling by example* (pp. 160-188). NY: Cambridge University Press.

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

Brown, G. (1990). *Listening to spoken English (2nd edition)*. London & NY: Longman.

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

Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.

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

Byrne, B. M. (2010). *Structural equation modeling with Amos: Basic concepts, applications, and programming (2nd edition)*. New York: Routledge.

Byrne, B. M. (2016). *Structural equation modeling with Amos: Basic concepts, applications, and programming (3rd edition)*. New York: Routledge.

Cambridge English Language Assessment (2014). *The Preliminary English Test (PET)*. Cambridge: Cambridge University Press.

CET Committee (2006). *College English Test Syllabus*. Shanghai: Shanghai Foreign Language Education Press.

CET-4/6 Testing Committee (2016). 全国大学英语四、六级考试大纲. www.cet.edu.cn.

Chen, Z. G. (2005). 英语听力:从制约因素看教学策略 (English listening comprehension: the teaching strategy based on an analysis of its influencing factors). *韶关学院学报（社会科学版）(Journal of Shaoguan University (Social Science))*, *26*, 127-131.

Chen, K. T., Lee, I. Y., & Lin, C. Y. (2010). EFL learners' uses of listening comprehension strategies and learning style preferences. *The International Journal of Learning, 17*, 245 – 256.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27,* 3-42.

Colantoni, L., Steele, J., & Escudero, P. (2015). *Second language speech.: Theory and practice*. Cambridge: Cambridge University Press.

Cubillos, J. H., Chieffo, L., & Fan, C. (2008). The impact of short-term study abroad programs on L2 listening comprehension skills. *Foreign Language Annals, 41*, 157 – 185.

Cutler, A. (1997). The comparative perspective on spoken-language processing. S*peech Communication, 21*, 3 – 15.

Cutler, A. (2001). Listening to a second language through the ears of a first. *Interpreting*, *5,* 1 – 23.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. The MIT Press.

Cutler, A., & Clifton, C. (1999). Comprehending spoken language: A blueprint of the listener. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 123-166). Oxford, UK: Oxford University Press.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 113-121.

Dunn, A. L., & Tree, J. F. (2009). A quick, gradient bilingual dominance scale. *Bilingualism: Language and Cognition*, *12,* 273-289.

Felser, C., Marinis, T, & Clahsen, H. (2003a). Children's processing of ambiguous sentences: A study of relative clause attachment. *Language Acquisition, 11*, 127-163.

Felser, C., Roberts, l., Gross, R., & Marinis, T. (2003b). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics, 24*, 453 – 489.

Field, J. (2003). Promoting perception: Lexical segmentation in L2 listening. *ELT Journal*. *57*, 325-334.

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

Field, J. (2013). Cognitive validity. In A. Geranpayeh, & L. Taylor (Eds.), *Examining listening: Research and practice in assessing second language listening* (pp. 77 - 151). Cambridge: Cambridge University Press.

Field, A. (2017). Discovering statistics using IBM SPSS statistics (5th ed.). Los Angeles: London: SAGE.

Flowerdew, J., & Miller, L. (2005). *Second language listening: theory and practice*. Cambridge: Cambridge University Press.

Fodor, J. D. (1998). Learning to parse?. *Journal of Psycholinguistic Research, 27*, 285-319.

Garner, M., & Huang, D. (2014). *Testing a nation: The social and educational impact of the College English Test in China*. Bern, Switzerland: Peter Lang AG.

Gibson, E., & Pearlmutter, N. J. (1998). Constraints on sentence comprehension. *Trends in Cognitive Sciences, 2*, 262 – 268.

Goh, C.C.M. (2000). A cognitive perspective on language learners' comprehension problems. *System, 28*, 55-75.

Grabe, W. (2009). Reading in a second language. Cambridge: UK: Cambridge University Press.

Graham, S. (2006). Listening comprehension: The learners' perspective. *System*, *34*, 165-182.

Graham, S. (2011). Self-efficacy and academic listening. *Journal of English for Academic Purposes*, *10*, 113-117.

Graham, S., & Macaro, E. (2008). Strategy instruction in listening for lower-intermediate learners of French. *Language Learning*, *58*, 747-783.

Graham, S., Santos, D., & Vanderplank, R. (2008). Strategy clusters and sources of knowledge in French L2 listening comprehension. *Innovation in Language Learning and Teaching*, *4*, 1-20.

Graham, S., Santos, D., & Vanderplank, R. (2010). Strategy clusters and sources of knowledge in French L2 listening comprehension. *Innovation in Language Learning and Teaching*, *4*, 1-20.

Grosjean, F. (2015). The complementarity principles and its impact on processing, acquisition, and dominance. In C. Silva-Corvalan & J. Treffers-Daller (Eds.), *Language* *dominance in billinguals: Issues of measurement and operationalization* (pp. 66 - 84). Cambridge: Cambridge University press.

Gu, S., & Wang, T. (2007). Study on the relationship between working memory and EFL listening comprehension. *中国英语教学, 30, Issue 6,* 46 – 56.

Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). London: Arnold.

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

Hu, X. Y. (2009). 二语听力理解困难的调查及认知心理分析. *云梦学刊, 30, Issue 1,* 142 – 145.

Ito, K., & Strange, W. (2009). Perception of allophonic cues to English word boundaries by Japanese second language learners of English. *Journal of the Acoustical Society of America, 125,* 2348-2360.

Institute voor Toegepaste Sociologie (ITS). (1994). *Nonverbal intelligence test: Adapted from testing series grade 4 developed by RION*. Mijmegen, the Netherlands.

Jia, M. F. (2003). 非语言性因素对英语听力理解的影响及对策. *清华大学教育研究, 24,* 33 – 36.

Juffs, A. (1998). Main verb versus reduced relative clause ambiguity resolution in L2 sentence processing. *Language Learning, 48,* 107-147.

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

Kline, R. B. (2016). *Principles and practice of structural equation modelling* (4th Ed.). NY: The Guilford Press.

Lapkin, S., Hart, D., & Swain, M. (1995). A Canadian international exchange: Evaluating the linguistic impact of a three-month stay in Quebec. In B. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 67 - 94). Amsterdam: John Benjamins.

Lee, J., & Schallert, D. L. (1997). The relative contribution of L2 language proficiency and L1 reading ability to L2 reading performance: A test of the threshold hypothesis in an EFL context. *TESOL Quarterly, 31,* 713-739.

Li, J., Zhang, N., Wang, Q., & Tang, Y. (2010). 英语听力的影响因素及应对策略(On affect factors of English listening and its counter measure), *安徽警官职业学院学报 (Journal of Anhui Vocational College of Police Officers)*, *9*, 72-74.

Li, R. (2012). 基于因子分析的英语听力理解影响因素研究(Influencing factor analysis of English listening comprehension). *渤海大学学报（哲学社会科学版）(Journal of Bohai Universit y (Social Science)), Issue 5*, 116-120.

Li, W., Zhang, H., & Shu, H. (2009). *教育与心理定量研究方法与统计方法——SPSS 实用指导 (Jiaoyu yu xinli dingliang yanjiu fangfa yu tongji fangfa ---- SPSS shiyong zhidao).* 北京 (Beijing): 北京师范大学出版社 (Beijing Normal University Publishing Group).

Lin, C. Y., & Wang, M. (2018). The use of lexical and sublexical cues in speech segmentation by second language learners of English. *Journal of Second Language Studies*, 166-198.

Liu, M. L. (2011). 英语词汇知识和听力理解的关系研究 (The relationship between English vocabulary knowledge and listening comprehension). *外语教学与研究 (China Electric Power Education), (17)*, 193-196.

Liu, S. (1995). 英语词汇量与听力理解关系的研究 (The relationship between English vocabulary size and listening comprehension). *海南大学学报 (Journal of Hainan University (Social Science)*), *(2)*, 84-89.

Llanes, A. & Botana, G. P. (2015). Does listening comprehension improve as a result of a short study abroad experience? *Spanish Journal of Applied Linguistics, 28*, 199 – 212.

Macaro, E., Graham, S., & Woore, R. (2016). *Improving foreign language teaching: towards a research-based curriculum and pedagogy*. Oxon: Routldege.

Macaro, E., Vanderplank, R., & Graham, S. (2005). *A systematic review of the role of prior knowledge in unidirectional listening comprehension*. EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Matthews, J. (2018). Vocabulary for listening: Emerging evidence for high and mid-frequency vocabulary knowledge. *System*, *72*, 23 – 36.

Matthews, J., & Cheng, J. (2015). Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System*, *52*, 1-13.

Matthews, J., Cheng, J., & O'Toole, J. M. (2015). Computer-mediated input, output and feedback in the development of L2 word recognition from speech. *ReCALL*, *27*, 321-339.

Matthews, J., O'Toole, J. M., & Chen, S. (2016). The impact of word recognition from speech (WRS) proficiency level on interaction, task success and word learning: design implications for CALL to develop L2 WRS. *Computer Assisted Language Learning*,

Marinis, T. (2010). Using on-line processing methods in language acquisition research. In E. Blom & S. Unsworth (Eds.), Experimental methods in language acquisition research (139 - 162). Amsterdam: John Benjamins Publishing Company.

McLaughlin, B., & Heredia, R. (1996). Information-processing approaches to research on second language acquisition and use. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 213-228). San Diego, CA: Academic Press.

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

Meara, P., & Milton, J. (2003). *X_Lex. The Swansea Levels Test*. Newbury, UK: Express.

Mecartty, F. (2000). Lexical and grammatical knowledge in reading and listening comprehension by foreign language learners of Spanish. *Applied Language Learning*, *11*, 323-348.

Mills, C. J., Ablard, K. E., & Brody, L. E. (1993). The Raven's Progressive Matrices: Its usefulness for identifying gifted/talented students. *Roeper Review: A Journal on Gifted education*, *15*, 183 – 186.

Milton, J., & Hopkins, N. (2005). *Aural Lex*. Swansea, UK: Swansea University.

Milton,J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In, R. Chacón-Beltrán, C. Abello-Contesse, C., & M. Torreblanca-López, *Further Insights into Non-native Vocabulary Teaching and Learning* (pp. 83-98). Bristol: Multilingual Matters.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*. *63*, 59-81.

Nation, I. S. P. (2012). Vocabulary Size Test Instructions and description. Retrieved April 30, 2014, from http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Vocabulary-Size-Test-information-and-specifications.pdf.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*, 9-13.

Netten, A., Droop, M., & Verhoeven, L. (2011). Predictors of reading literacy for first and second language learners. *Read Writ, 24*, 413 – 425.

Ockey, G. J., Papageorgiou, S., & French, R. (2016). Effects of Strength of Accent on an L2 Interactive Lecture Listening Comprehension Test. *International Journal of Listening*, *30*, 84-98.

Oh, E. (2016). Comparative studies on the roles of linguistic knowledge and sentence processing speed in L2 listening and reading comprehension in an EFL tertiary setting. *Reading Psychology*, *37*, 257-285.

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition, 24*, 501 – 528.

Poelmans, P. (2003). *Developing second-language listening comprehension: Effects of training lower-order skills versus higher-order strategy.* Doctoral dissertation, University of Amsterdam, the Netherlands.

Raven, J. C., Court, J. H., & Raven, J. (1992). *Standard Progressive Matrices. 1992 Edition*. Oxford: Oxford Psychologists Press.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, *10*, 355 –371.

Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. Kunnan (Ed.), *Validation in language assessment* (pp. 41 –60). Mahwah, NJ: Erlbaum.

Révész, A., & Brunfaut, T. (2013). Text characteristics of task input and difficulty in second language listening comprehension. *Studies in Second Language Acquisitio*n, *35*, 31–65.

Roberts, L. (2012). Individual differences in second language sentence processing. *Language Learning*, *62*, 172-188.

Rost, M. (1990). *Listening in language learning*. NY: Longman.

Rost, M. (2016). *Teaching and research listening (3rd edition.)*. NY: Routledge.

Rubin, J. (1994). A review of second language listening comprehension research. *The Modern Language Journal*, *78*, 199-221.

Salamoura, A., & Saville, N. (2010) Exemplifying the CEFR criterial features of written learner English from the English Profile Programme. In Bartning, I., Martin, M., & Vedder, I. (Eds.) *Second language acquisition and language testing in Europe*. EUROSLA Monograph Series Pallotti, G.

Schoonen, R. (2015). Structural equation modelling in L2 research. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp 213 -242). NY: Routledge.

Segalowitz, N. (1991). Does advanced skill in a second language reduce automaticity in the first language? *Language Learning, 41*, 59 – 83.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. NY: Routledge.

Shoemaker, E. (2014). The exploitation of subphonemic acoustic detail in L2 speech segmentation. *Studies in Second Language Acquisition*, *36*, 709-731.

Staehr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *The Language Learning Journal*, *36*, 139-152.

Staehr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *SSLA*, *31*, 577-607.

Steele, J., & Ambady, N. (2006). “Math is Hard!” The effect of gender priming on women’s attitudes. *Journal of Experimental Social Psychology, 42*, 428 – 436.

Sternberg, R. J. (2009). *Cognitive psychology* (5th edition). Wadsworth: Cengage Learning.

Su, Y. L. (2003). 论听力学习策略的可教性—一项基于中国外语初学者的实验研究. *现代外语, 26, Issue 1*, 48 – 58.

Sun, L., & Li, J. Q. (2008). 大学英语学习者中介语中的听力理解策略使用模型. *解放军外国语学院学报, 31, Issue 6*, 49 – 53.

Tauroza, S., & Allison, D. (1990). Speech rates in British English. *Applied Linguistics, 11*, 90 – 105.

Traxler, M. (2002). Plausibility and subcategorization preference in children’s processing of temporarily ambiguous sentences: Evidence from self-paced reading. *Quarterly Journal of Experimental Psychology, 55*, 75-96.

Trenkic, D., & Warmington, M. (2018). Language and literacy skills of home and international university students: How different are they, and does it matter?. *Bilingualism: Language and Cognition*, *21*, 1 – 17.

Valentini, A., & Serratrice, L. (2018). Listening comprehension in bilingual children: Local and global inferences. A poster presentation on the conference *The protective role of language in global migration and mobility: language policies and practices in refugee setting*, September, University of Reading.

van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, *34*, 457-479.

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*, 6-18.

Vandergrift, L., & Baker, S. (2015). Learner variables in second language listening comprehension: An exploratory path analysis. *Language Learning*, *65*, 390-416.

Vandergrift, L., Goh, C. C. M., Mareschal, C. J., & Tafaghodtari, M. H. (2006). The metacognitive awareness listening questionnaire: Development and validation. *Language Learning*, *56*, 431-462.

Wall, M. (2012). How to analyse reaction time (RT) data: Part 1. Retrived from https://computingforpsychologists.wordpress.com/2012/12/11/how-to-analyse-reaction-time-rt-data-part-1/.

Wang, F. (2008). 从英语听力理解本质透视听力学习障碍. *西北大学学报(哲学社会科学版)*, *38*, *Issue 2*, 144 – 147.

Wang, Y. (2008). 基于认知框架的二语学生听力理解困难分析. *外语教学理论与实践*, *Issue 1*, 53 – 58.

Wang, Y. (2014). *Explaining listening comprehension among non-English major university students: the impact of general language proficiency, vocabulary knowledge, motivation and learning strategies*. MA thesis, University of Reading, the UK.

Wang, R. (2014). 英汉语法差异对大学生英语写作的影响. *海外英语*, *Issue 6*, 92 – 93.

Wang, Y., & Treffers-Daller, J. (2017). Explaining listening comprehension among L2 learners of English: the contribution of general language proficiency, vocabulary knowledge and metacognitive awareness. *System, 65*, 139 – 150.

Webb, S., & Rodgers, M. H. (2009). The lexical coverage of movies. *Applied Linguistics, 30,* 407-427.

Wechsler, D. (1997). *Wechsler Memory Scale* (3rd Ed). San Antonio, TX: The Psychological Corporation.

Weir, C. (2005). *Language testing and validation: An evidence-based approach (research and practice in applied linguistics)*. NY: Palgrave Macmillan.

Winne, P., & Perry, N. (2000). Measuring self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). San Diego: Academic Press.

Wu, Y. (1998). What do tests of listening comprehension test? – A retrospection study of EFL test-takers performing multiple-choice task. *Language Learning, 15*, 21 – 44.

Wu, M. L. (2010). *结构方程模型----AMOS 的操作与应用 (Structural equation modeling----the operation and application of AMOS)*. Chongqing, China: Chongqing University Press.

Xu, Y. (2014). 大学英语听力障碍及对策. *科技创业家, Issue 2*, 195.

Zampini, L., Suttora, C., D'Odorico L., & Zanchi, P. (2013). Sequential reasoning and listening text comprehension in preschool children. *European Journal of Developmental Psychology, 5*, 563 – 579.

Zhang, X. D. (2011). 词汇知识与二语听力理解关系研究. *外语界, Issue 2*, 36 – 42.

Zhang, J. J., & Lin, Z. H. (2002). 汉语拼音水平与英语语音意识、英语拼写能力的关系. *心理科学, 25, Issue 5*, 601 – 603.

Zhang, Y. (2012). 浅谈英汉语法对比. *海外英语, Issue 1*, 261, 271.

Zhang, Y., Li, Y. J., & Liu, J. (2005). 试论汉语拼音在英语语音教学中的正迁移作用. *基础教育外语教学研究. Issue 1*, 34 – 36.

Zhang, Z. H., Lv, L., & Tan, X. (2010). 语言水平与听力理解困难关联性的认知心理分析. *外语教学理论与实践, Issue 1*, 33 – 38.

Zhu, S., & Zhu, L. X. (2007). 改革后的大学英语四级考试的测试质量分析与评价 (Testing analysis and evaluation of new CET4). *绥化学院学报*(*Journal of Suihua University)*, *27*(4), 119-121.

Zoghlami, N. (2015). *Foreign-language listening: bottom-up and top-down processes*----*issues for EFL teaching and research*. Unpublished doctoral thesis, University Paris.

Appendices

Appendix 1 – Appendix 198

Appendix 1 Mean percentage (%) of rates of English use in daily activities in two subgroups

Activities	***Mean* (Group 1, *N* = 147)**	***Mean* (Group 2, *N* = 40)**
Work	5.41	48.25
studies (in general)	18.93	68.75
immediate family (with whom you live)	0.49	20.55
distant family (or ancestors)	0.62	7.33
cooking	0.82	26.18
cleaning	0.66	26.3
shopping	2.52	75.75
administrative matters	1.93	75.10
leisure (indicate which activities)	2.29	37.75
trips	4.0	66.25
evening out	1.48	46.75
clothes	3.99	64.75
sports (indicate which sports)	3.01	52.75
transportation	4.65	79.88
health	3.29	70.50
education	17.72	84.63
politics	3.73	56.88

religion	1.84	37.38
love/affection	4.95	25.50
other topics, namely…..	Not complete	Not complete
Total	4.33	51.12

Appendix 2

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Enguse	1	147	4.3333	3.99436	.32945
	2	40	51.1158	16.52792	2.61329

Appendix 3

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Enguse	Equal variances assumed	68.393	0.000	-31.315	185	0.000	-46.78242	1.49395	-49.72979	-43.83505
	Equal variances not assumed			-17.761	40.247	0.000	-46.78242	2.63398	-52.10487	-41.45996

Appendix 4

Group Statistics					
	Group	N	Mean	Std. Deviation	Std. Error Mean
Speaking	1	147	4.24	1.981	.163
	2	40	5.10	1.661	.263
Writing	1	147	4.18	1.911	.158
	2	40	5.25	1.410	.223
Listening	1	147	3.93	1.940	.160
	2	40	5.68	1.966	.311
Reading	1	147	4.75	2.119	.175
	2	40	6.08	1.845	.292

Appendix 5

Group Statistics					
	Group	N	Mean	Std. Deviation	Std. Error Mean
selfrating	1	147	17.1020	7.16607	.59105
	2	40	22.1000	5.70110	.90142

Appendix 6

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
selfrating	Equal variances assumed	4.973	.027	-4.072	185	.000	-4.99796	1.22751	-7.41967	-2.57625
	Equal variances not assumed			-4.637	75.990	.000	-4.99796	1.07792	-7.14482	-2.85110

Appendix 7

Group Statistics					
	Group	N	Mean	Std. Deviation	Std. Error Mean
OPT	1	147	135.5918	12.78363	1.05438
	2	40	146.3000	15.01999	2.37487

Appendix 8

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
OPT	Equal variances assumed	1.840	.177	-4.519	185	.000	-10.70816	2.36941	-15.38270	-6.03362
	Equal variances not assumed			-4.121	55.316	.000	-10.70816	2.59841	-15.91482	-5.50151

Appendix 9 Numbers of outliers in each Task (N=187)

Task	Number of Outliers
The test of word segmentation (WST)	9
The OPT listening section (OPT listening)	8
Word Recognition from Speech (WRS)	3
LVLT	6
TROG-2	7
The OPT grammar section (OPT grammar)	8
WM forward task in Chinese (WMCF)	14
WM backward task in Chinese (WMCB)	4
WM forward task in English (WMEF)	7
WM backward task in English (WMEB)	11
Raven's Standard Progressive Matrices (Raven's SPM)	8
PET listening section (PET listening)	11

Appendix 10 The number of outliers in the six sections of LVLT (N=187)

Section	Number of outliers
1	6
2	7
3	5
4	11
5	9
6	5

Appendix 11 The number of outliers in five sections of Raven's Standard Progressive Matrices (N=187)

Section	Number of outliers
A	2
B	5
C	6
D	7
E	9

Appendix 12 The number of outliers in experimental items in the task of SPS (N=187)

Item	Number of outliers
4	4
6	4
7	3
9	3
12	5
14	4
16	3
19	2
22	5
23	7
25	5
27	7
29	9
31	6
32	5
34	4
36	4
38	7
41	3
43	7
44	4
45	6
47	2
48	5

Appendix 13 The mean score of the PET listening section in the pilot study 3

	N	Range	Minimum	Maximum	Mean	Std. Deviation
PETpilot	13	14	8	22	12.85	4.413
Valid N (listwise)	13					

Appendix 14 Reliability statistics of the PET listening in the pilot study 3

Cronbach's Alpha	N of Items
.789	25

Appendix 15 Reliability of WST

Cronbach's Alpha	N of Items
.451	84

Appendix 16

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
WSTterm1	57.59	27.297	-.024	.457
WSTterm2	57.68	26.767	.076	.447
WSTterm3	57.61	26.561	.127	.442
WSTterm4	57.54	27.626	-.093	.463
WSTterm5	57.45	27.001	.068	.448
WSTterm6	57.75	27.705	-.107	.467
WSTterm7	57.44	26.850	.114	.445
WSTterm8	57.45	27.410	-.040	.457
WSTterm9	57.43	26.559	.199	.438
WSTterm10	57.66	26.440	.144	.440
WSTterm11	57.63	26.739	.087	.446
WSTterm12	57.87	26.547	.117	.443
WSTterm13	57.71	26.370	.154	.439
WSTterm14	57.49	26.972	.063	.449
WSTterm15	57.42	26.493	.228	.437
WSTterm16	57.95	27.664	-.100	.465
WSTterm17	57.52	27.359	-.032	.457
WSTterm18	58.05	26.734	.105	.445

WSTterm19	57.54	26.207	.226	.433
WSTterm20	57.52	27.380	-.037	.458
WSTterm21	57.66	26.398	.153	.439
WSTterm22	57.40	27.047	.080	.448
WSTterm23	57.33	26.943	.209	.443
WSTterm24	57.40	27.316	-.005	.453
WSTterm25	57.35	27.262	.030	.451
WSTterm26	57.40	27.198	.030	.451
WSTterm27	58.02	27.516	-.069	.461
WSTterm28	57.51	27.294	-.017	.456
WSTterm29	57.64	27.039	.025	.453
WSTterm30	57.51	26.746	.110	.444
WSTterm31	57.61	26.261	.191	.436
WSTterm32	57.76	27.020	.023	.453
WSTterm33	57.56	26.592	.132	.442
WSTterm34	58.04	26.982	.048	.450
WSTterm35	57.63	26.698	.097	.445
WSTterm36	57.75	27.512	-.071	.463
WSTterm37	57.63	25.881	.268	.427
WSTterm38	57.40	26.778	.166	.442
WSTterm39	57.56	26.958	.052	.450
WSTterm40	57.81	27.414	-.052	.461
WSTterm41	57.48	26.670	.143	.442
WSTterm42	57.34	26.635	.321	.437
WSTterm43	57.37	27.008	.114	.446
WSTterm44	57.39	27.012	.099	.447
WSTterm45	57.81	26.648	.095	.445
WSTterm46	57.58	26.621	.120	.443
WSTterm47	57.51	27.498	-.064	.460
WSTterm48	57.52	26.552	.154	.440
WSTterm49	57.42	26.449	.246	.436
WSTterm50	57.63	27.148	.004	.455
WSTterm51	57.81	27.307	-.032	.459
WSTterm52	57.84	27.286	-.028	.458
WSTterm53	57.79	26.965	.034	.452
WSTterm54	57.63	26.051	.231	.431
WSTterm55	57.57	26.569	.135	.442
WSTterm56	57.56	26.656	.118	.443
WSTterm57	57.52	25.616	.374	.420
WSTterm58	57.57	27.666	-.101	.464
WSTterm59	57.84	27.680	-.102	.466
WSTterm60	57.34	27.364	-.007	.452

WSTterm61	57.79	26.169	.189	.435
WSTterm62	58.01	27.495	-.065	.461
WSTterm63	57.61	26.315	.180	.437
WSTterm64	57.38	26.602	.246	.438
WSTterm65	57.66	27.377	-.043	.460
WSTterm66	57.81	27.855	-.135	.470
WSTterm67	57.72	27.150	-.001	.456
WSTterm68	57.49	26.757	.115	.444
WSTterm69	57.65	27.142	.004	.455
WSTterm70	57.43	27.483	-.058	.458
WSTterm71	57.32	27.163	.137	.447
WSTterm72	57.34	27.001	.155	.445
WSTterm73	57.52	26.649	.131	.442
WSTterm74	57.43	26.655	.172	.440
WSTterm75	57.66	27.106	.010	.454
WSTterm76	57.58	26.966	.047	.450
WSTterm77	57.50	26.219	.241	.433
WSTterm78	57.80	26.375	.149	.439
WSTterm79	57.73	26.995	.029	.452
WSTterm80	57.50	26.810	.100	.445
WSTterm81	57.99	27.086	.019	.453
WSTterm82	57.42	26.900	.112	.445
WSTterm83	57.67	26.782	.074	.447
WSTterm84	57.65	27.293	-.026	.458

In order to improve the reliability, the following items were deleted from the task based on the results of statistics: item66, item59, item6, item36, item16, item4, item27, item65, item40, item62, item51, item20, item76, item50, item70, item58, item52, item47, item8, item1, item24, item28, item17. In total, 23 items were deleted and 61 items were kept for later analysis. The reliability of the new WST was .66.

Appendix 17 The reliability of the new WST

Cronbach's Alpha	N of Items
.66	61

Appendix 18 The reliability of the OPT listening task

Cronbach's Alpha	N of Items
.68	100

In order to improve the reliability of the test, based on the results of the statistics, items 5, 20, 31, 43, 52, 54, 75, 76, 89, 95 were deleted. The new OPT listening section composed of 90 items and the reliability of the new OPT listening task was .73.

Appendix 19 The reliability of the new OPT listening task

Cronbach's Alpha	N of Items
.73	90

Appendix 20

Descriptive Statistics

	N	Minimum	Maximum	Mean		Std. Deviation
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic
aspipositivepercen	187	38.89	100.00	77.6292	1.06444	14.55606
aspinegativepercen	187	22.22	100.00	70.6477	1.22241	16.71627
glotpositivepercen	187	27.78	88.89	58.8532	.80435	10.99938
glotnegativepercen	187	33.33	100.00	75.6982	.84874	11.60632
double1percen	187	.00	100.00	51.6934	1.73072	23.66727
double2percen	187	16.67	100.00	71.3904	1.62336	22.19909
aspirationpercen	187	50.00	97.22	74.1384	.77278	10.56767
glottalpercen	187	47.22	83.33	67.2757	.50553	6.91304
doublepercen	187	25.00	91.67	61.5419	.93544	12.79196
totalstim	187	41.00	73.00	58.2941	.38273	5.23369
totalstimpercen	187	48.81	86.90	69.3978	.45563	6.23058
Valid N (listwise)	187					

Appendix 21

Correlation between duration of studying in the UK and frequency of L2 language use (N = 187)

Correlations

		MonthinUK	Enguse
MonthinUK	Pearson Correlation	1	.841**
	Sig. (2-tailed)		.000
	N	187	187
Enguse	Pearson Correlation	.841**	1
	Sig. (2-tailed)	.000	
	N	187	187

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

Appendix 22

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-1.098	5.019		-.219	.827		
	WSEGTTL	-.011	.051	-.014	-.211	.833	.767	1.304
	WRSTL	.033	.027	.173	1.222	.223	.161	6.212
	OPTLSNTL	.054	.047	.084	1.135	.258	.591	1.692
	TROG2TL	.018	.034	.048	.548	.584	.426	2.346
	OPTGTL	.061	.033	.127	1.832	.069	.670	1.492
	LVLTTL	-.003	.025	-.013	-.120	.905	.294	3.397
	SPSAVER	.000	.000	.039	.650	.516	.906	1.103
	WMCF	.101	.187	.032	.540	.590	.905	1.104
	WMCB	-.050	.093	-.037	-.536	.593	.686	1.458
	WMEF	-.054	.147	-.025	-.371	.711	.724	1.381
	WMEB	.244	.200	.094	1.218	.225	.540	1.852
	RAVENTL	.049	.052	.057	.947	.345	.895	1.118
	Enguse	.069	.017	.350	3.942	.000	.408	2.448

a. Dependent Variable: PETTL

Appendix 23

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1091.177	1	1091.177	98.127	.000[b]
	Residual	2057.208	185	11.120		
	Total	3148.385	186			
2	Regression	1091.664	2	545.832	48.832	.000[c]
	Residual	2056.721	184	11.178		
	Total	3148.385	186			

a. Dependent Variable: PETTL

b. Predictors: (Constant), WRSTL

c. Predictors: (Constant), WRSTL, WSEGTTL

Appendix 24

Model Summary

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.589[a]	.347	.343	3.335	.347	98.127	1	185	.000
2	.589[b]	.347	.340	3.343	.000	.044	1	184	.835

a. Predictors: (Constant), WRSTL

b. Predictors: (Constant), WRSTL, WSEGTTL

Appendix 25

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	8.373	.555		15.078	.000					
	WRSTL	.111	.011	.589	9.906	.000	.589	.589	.589	1.000	1.000
2	(Constant)	7.951	2.097		3.791	.000					
	WRSTL	.110	.012	.584	9.185	.000	.589	.561	.547	.878	1.139
	WSEGTTL	.010	.050	.013	.209	.835	.217	.015	.012	.878	1.139

a. Dependent Variable: PETTL

Appendix 26

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	148.733	1	148.733	9.173	.003[b]
	Residual	2999.652	185	16.214		
	Total	3148.385	186			
2	Regression	1091.664	2	545.832	48.832	.000[c]
	Residual	2056.721	184	11.178		
	Total	3148.385	186			

a. Dependent Variable: PETTL

b. Predictors: (Constant), WSEGTTL

c. Predictors: (Constant), WSEGTTL, WRSTL

Appendix 27

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.217[a]	.047	.042	4.027	.047	9.173	1	185	.003
2	.589[b]	.347	.340	3.343	.299	84.357	1	184	.000

a. Predictors: (Constant), WSEGTTL

b. Predictors: (Constant), WSEGTTL, WRSTL

Appendix 28

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	5.766	2.510		2.298	.023					
	WSEGTTL	.171	.057	.217	3.029	.003	.217	.217	.217	1.000	1.000
2	(Constant)	7.951	2.097		3.791	.000					
	WSEGTTL	.010	.050	.013	.209	.835	.217	.015	.012	.878	1.139
	WRSTL	.110	.012	.584	9.185	.000	.589	.561	.547	.878	1.139

a. Dependent Variable: PETTL

Appendix 29

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1111.234	3	370.411	33.275	.000[b]
	Residual	2037.151	183	11.132		
	Total	3148.385	186			
2	Regression	1111.719	4	277.930	24.836	.000[c]
	Residual	2036.666	182	11.190		
	Total	3148.385	186			
3	Regression	1141.597	6	190.266	17.066	.000[d]
	Residual	2006.789	180	11.149		
	Total	3148.385	186			

a. Dependent Variable: PETTL

b. Predictors: (Constant), OPTLSNTL, WSEGTTL, WRSTL

c. Predictors: (Constant), OPTLSNTL, WSEGTTL, WRSTL, LVLTTL

d. Predictors: (Constant), OPTLSNTL, WSEGTTL, WRSTL, LVLTTL, OPTGTL, TROG2TL

Appendix 30

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.594[a]	.353	.342	3.336	.353	33.275	3	183	.000
2	.594[b]	.353	.339	3.345	.000	.043	1	182	.835
3	.602[c]	.363	.341	3.339	.009	1.340	2	180	.264

a. Predictors: (Constant), OPTLSNTL, WSEGTTL, WRSTL

b. Predictors: (Constant), OPTLSNTL, WSEGTTL, WRSTL, LVLTTL

c. Predictors: (Constant), OPTLSNTL, WSEGTTL, WRSTL, LVLTTL, OPTGTL, TROG2TL

Appendix 31

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Correlations Zero-order	Correlations Partial	Correlations Part	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	4.266	3.479		1.226	.222					
	WSEGTTL	.003	.050	.004	.069	.945	.217	.005	.004	.868	1.152
	WRSTL	.100	.015	.527	6.872	.000	.589	.453	.409	.601	1.663
	OPTLSNTL	.064	.048	.100	1.326	.187	.419	.098	.079	.627	1.594
2	(Constant)	4.628	3.899		1.187	.237					
	WSEGTTL	.005	.051	.007	.103	.918	.217	.008	.006	.844	1.185
	WRSTL	.103	.022	.545	4.692	.000	.589	.329	.280	.263	3.797
	OPTLSNTL	.063	.048	.098	1.302	.194	.419	.096	.078	.624	1.604
	LVLTTL	-.005	.026	-.022	-.208	.835	.474	-.015	-.012	.315	3.179
3	(Constant)	2.933	4.049		.724	.470					
	WSEGTTL	-.013	.052	-.016	-.242	.809	.217	-.018	-.014	.807	1.240
	WRSTL	.093	.024	.490	3.873	.000	.589	.277	.230	.222	4.514
	OPTLSNTL	.054	.049	.085	1.106	.270	.419	.082	.066	.605	1.652
	LVLTTL	-.004	.026	-.017	-.155	.877	.474	-.012	-.009	.312	3.204
	TROG2TL	.010	.034	.025	.288	.774	.449	.021	.017	.464	2.157
	OPTGTL	.051	.033	.108	1.547	.124	.361	.115	.092	.731	1.368

a. Dependent Variable: PETTL

Appendix 32

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	708.768	1	708.768	53.747	.000[b]
	Residual	2439.617	185	13.187		
	Total	3148.385	186			
2	Regression	1111.719	4	277.930	24.836	.000[c]
	Residual	2036.666	182	11.190		
	Total	3148.385	186			
3	Regression	1141.597	6	190.266	17.066	.000[d]
	Residual	2006.789	180	11.149		
	Total	3148.385	186			

a. Dependent Variable: PETTL

b. Predictors: (Constant), LVLTTL

c. Predictors: (Constant), LVLTTL, WSEGTTL, OPTLSNTL, WRSTL

d. Predictors: (Constant), LVLTTL, WSEGTTL, OPTLSNTL, WRSTL, OPTGTL, TROG2TL

Appendix 33

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.474[a]	.225	.221	3.631	.225	53.747	1	185	.000
2	.594[b]	.353	.339	3.345	.128	12.003	3	182	.000
3	.602[c]	.363	.341	3.339	.009	1.340	2	180	.264

a. Predictors: (Constant), LVLTTL

b. Predictors: (Constant), LVLTTL, WSEGTTL, OPTLSNTL, WRSTL

c. Predictors: (Constant), LVLTTL, WSEGTTL, OPTLSNTL, WRSTL, OPTGTL, TROG2TL

Appendix 34

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Correlations Zero-order	Correlations Partial	Correlations Part	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	1.765	1.598		1.104	.271					
	LVLTTL	.115	.016	.474	7.331	.000	.474	.474	.474	1.000	1.000
2	(Constant)	4.628	3.899		1.187	.237					
	LVLTTL	-.005	.026	-.022	-.208	.835	.474	-.015	-.012	.315	3.179
	WSEGTTL	.005	.051	.007	.103	.918	.217	.008	.006	.844	1.185
	WRSTL	.103	.022	.545	4.692	.000	.589	.329	.280	.263	3.797
	OPTLSNTL	.063	.048	.098	1.302	.194	.419	.096	.078	.624	1.604
3	(Constant)	2.933	4.049		.724	.470					
	LVLTTL	-.004	.026	-.017	-.155	.877	.474	-.012	-.009	.312	3.204
	WSEGTTL	-.013	.052	-.016	-.242	.809	.217	-.018	-.014	.807	1.240
	WRSTL	.093	.024	.490	3.873	.000	.589	.277	.230	.222	4.514
	OPTLSNTL	.054	.049	.085	1.106	.270	.419	.082	.066	.605	1.652
	TROG2TL	.010	.034	.025	.288	.774	.449	.021	.017	.464	2.157
	OPTGTL	.051	.033	.108	1.547	.124	.361	.115	.092	.731	1.368

a. Dependent Variable: PETTL

Appendix 35

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1098.263	1	1098.263	99.106	.000[b]
	Residual	2050.122	185	11.082		
	Total	3148.385	186			
2	Regression	1293.418	4	323.355	31.726	.000[c]
	Residual	1854.967	182	10.192		
	Total	3148.385	186			
3	Regression	1361.161	6	226.860	22.848	.000[d]
	Residual	1787.224	180	9.929		
	Total	3148.385	186			
4	Regression	1362.993	7	194.713	19.522	.000[e]
	Residual	1785.392	179	9.974		
	Total	3148.385	186			
5	Regression	1380.836	11	125.531	12.428	.000[f]
	Residual	1767.549	175	10.100		
	Total	3148.385	186			
6	Regression	1391.371	12	115.948	11.482	.000[g]
	Residual	1757.014	174	10.098		
	Total	3148.385	186			

a. Dependent Variable: PETTL

b. Predictors: (Constant), Enguse

c. Predictors: (Constant), Enguse, WSEGTTL, OPTLSNTL, WRSTL

d. Predictors: (Constant), Enguse, WSEGTTL, OPTLSNTL, WRSTL, OPTGTL, TROG2TL

e. Predictors: (Constant), Enguse, WSEGTTL, OPTLSNTL, WRSTL, OPTGTL, TROG2TL, LVLTTL

f. Predictors: (Constant), Enguse, WSEGTTL, OPTLSNTL, WRSTL, OPTGTL, TROG2TL, LVLTTL, WMCF, WMCB, WMEF, WMEB

g. Predictors: (Constant), Enguse, WSEGTTL, OPTLSNTL, WRSTL, OPTGTL, TROG2TL, LVLTTL, WMCF, WMCB, WMEF, WMEB, RAVENTL

Appendix 36

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.591[a]	.349	.345	3.329	.349	99.106	1	185	.000
2	.641[b]	.411	.398	3.193	.062	6.383	3	182	.000
3	.658[c]	.432	.413	3.151	.022	3.411	2	180	.035
4	.658[d]	.433	.411	3.158	.001	.184	1	179	.669
5	.662[e]	.439	.403	3.178	.006	.442	4	175	.778
6	.665[f]	.442	.403	3.178	.003	1.043	1	174	.308

a. Predictors: (Constant), Enguse

b. Predictors: (Constant), Enguse, WSEGTTL, OPTLSNTL, WRSTL

c. Predictors: (Constant), Enguse, WSEGTTL, OPTLSNTL, WRSTL, OPTGTL, TROG2TL

d. Predictors: (Constant), Enguse, WSEGTTL, OPTLSNTL, WRSTL, OPTGTL, TROG2TL, LVLTTL

e. Predictors: (Constant), Enguse, WSEGTTL, OPTLSNTL, WRSTL, OPTGTL, TROG2TL, LVLTTL, WMCF, WMCB, WMEF, WMEB

f. Predictors: (Constant), Enguse, WSEGTTL, OPTLSNTL, WRSTL, OPTGTL, TROG2TL, LVLTTL, WMCF, WMCB, WMEF, WMEB, RAVENTL

Appendix 37

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Correlations Zero-order	Correlations Partial	Correlations Part	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	11.654	.295		39.483	.000					
	Enguse	.116	.012	.591	9.955	.000	.591	.591	.591	1.000	1.000
2	(Constant)	4.714	3.331		1.415	.159					
	Enguse	.069	.016	.350	4.228	.000	.591	.299	.241	.473	2.115
	WSEGTTL	.015	.048	.019	.318	.750	.217	.024	.018	.865	1.156
	OPTLSNTL	.067	.046	.105	1.461	.146	.419	.108	.083	.627	1.594
	WRSTL	.050	.018	.265	2.756	.006	.589	.200	.157	.351	2.851
3	(Constant)	2.132	3.456		.617	.538					
	Enguse	.077	.016	.393	4.705	.000	.591	.331	.264	.453	2.207
	WSEGTTL	-.010	.049	-.013	-.209	.835	.217	-.016	-.012	.830	1.204
	OPTLSNTL	.052	.046	.081	1.126	.262	.419	.084	.063	.610	1.640
	WRSTL	.026	.022	.140	1.227	.221	.589	.091	.069	.243	4.116
	TROG2TL	.026	.032	.069	.830	.408	.449	.062	.047	.460	2.173
	OPTGTL	.074	.032	.156	2.346	.020	.361	.172	.132	.717	1.396
4	(Constant)	2.833	3.830		.740	.461					
	Enguse	.077	.016	.395	4.711	.000	.591	.332	.265	.452	2.214
	WSEGTTL	-.007	.049	-.008	-.132	.895	.217	-.010	-.007	.806	1.240
	OPTLSNTL	.050	.046	.078	1.083	.280	.419	.081	.061	.605	1.652

	WRSTL	.033	.026	.172	1.257	.211	.589	.094	.071	.168	5.945
	TROG2TL	.028	.032	.072	.860	.391	.449	.064	.048	.457	2.188
	OPTGTL	.074	.032	.154	2.317	.022	.361	.171	.130	.715	1.399
	LVLTTL	-.010	.024	-.043	-.429	.669	.474	-.032	-.024	.311	3.214
5	(Constant)	1.375	4.563		.301	.763					
	Enguse	.073	.017	.372	4.306	.000	.591	.310	.244	.429	2.332
	WSEGTTL	-.008	.051	-.010	-.152	.879	.217	-.011	-.009	.770	1.299
	OPTLSNTL	.052	.047	.082	1.113	.267	.419	.084	.063	.595	1.680
	WRSTL	.031	.027	.164	1.168	.244	.589	.088	.066	.163	6.153
	TROG2TL	.022	.033	.057	.666	.506	.449	.050	.038	.442	2.265
	OPTGTL	.066	.033	.138	2.022	.045	.361	.151	.115	.685	1.459
	LVLTTL	-.007	.025	-.027	-.263	.793	.474	-.020	-.015	.300	3.334
	WMCF	.090	.186	.029	.480	.632	.162	.036	.027	.908	1.101
	WMCB	-.045	.093	-.033	-.485	.628	.207	-.037	-.027	.688	1.454
	WMEF	-.056	.146	-.025	-.382	.703	.247	-.029	-.022	.729	1.372
	WMEB	.251	.199	.097	1.262	.209	.370	.095	.071	.545	1.835
6	(Constant)	-.580	4.947		-.117	.907					
	Enguse	.070	.017	.359	4.109	.000	.591	.297	.233	.420	2.383
	WSEGTTL	-.009	.051	-.011	-.173	.863	.217	-.013	-.010	.770	1.299
	OPTLSNTL	.051	.047	.080	1.088	.278	.419	.082	.062	.595	1.681
	WRSTL	.033	.027	.177	1.253	.212	.589	.095	.071	.161	6.200
	TROG2TL	.016	.033	.042	.491	.624	.449	.037	.028	.430	2.326
	OPTGTL	.061	.033	.128	1.851	.066	.361	.139	.105	.671	1.491
	LVLTTL	-.005	.025	-.021	-.201	.841	.474	-.015	-.011	.299	3.347

WMCF	.100	.187	.032	.536	.593	.162	.041	.030	.906	1.104
WMCB	-.051	.093	-.037	-.541	.589	.207	-.041	-.031	.686	1.458
WMEF	-.061	.146	-.028	-.419	.675	.247	-.032	-.024	.728	1.374
WMEB	.256	.199	.099	1.286	.200	.370	.097	.073	.545	1.836
RAVENTL	.053	.052	.061	1.021	.308	.198	.077	.058	.905	1.106

a. Dependent Variable: PETTL

Appendix 38

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1111.234	3	370.411	33.275	.000[b]
	Residual	2037.151	183	11.132		
	Total	3148.385	186			
2	Regression	1293.418	4	323.355	31.726	.000[c]
	Residual	1854.967	182	10.192		
	Total	3148.385	186			
3	Regression	1361.161	6	226.860	22.848	.000[d]
	Residual	1787.224	180	9.929		
	Total	3148.385	186			
4	Regression	1362.993	7	194.713	19.522	.000[e]
	Residual	1785.392	179	9.974		
	Total	3148.385	186			
5	Regression	1380.836	11	125.531	12.428	.000[f]
	Residual	1767.549	175	10.100		
	Total	3148.385	186			
6	Regression	1391.371	12	115.948	11.482	.000[g]
	Residual	1757.014	174	10.098		
	Total	3148.385	186			

a. Dependent Variable: PETTL

b. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL

c. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, Enguse

d. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, Enguse, OPTGTL, TROG2TL

e. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, Enguse, OPTGTL, TROG2TL, LVLTTL

f. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, Enguse, OPTGTL, TROG2TL, LVLTTL, WMCF, WMCB, WMEF, WMEB

g. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, Enguse, OPTGTL, TROG2TL, LVLTTL, WMCF, WMCB, WMEF, WMEB, RAVENTL

Appendix 39

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.594[a]	.353	.342	3.336	.353	33.275	3	183	.000
2	.641[b]	.411	.398	3.193	.058	17.875	1	182	.000
3	.658[c]	.432	.413	3.151	.022	3.411	2	180	.035
4	.658[d]	.433	.411	3.158	.001	.184	1	179	.669
5	.662[e]	.439	.403	3.178	.006	.442	4	175	.778
6	.665[f]	.442	.403	3.178	.003	1.043	1	174	.308

a. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL

b. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, Enguse

c. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, Enguse, OPTGTL, TROG2TL

d. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, Enguse, OPTGTL, TROG2TL, LVLTTL

e. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, Enguse, OPTGTL, TROG2TL, LVLTTL, WMCF, WMCB, WMEF, WMEB

f. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, Enguse, OPTGTL, TROG2TL, LVLTTL, WMCF, WMCB, WMEF, WMEB, RAVENTL

Appendix 40

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1111.234	3	370.411	33.275	.000[b]
	Residual	2037.151	183	11.132		
	Total	3148.385	186			
2	Regression	1293.418	4	323.355	31.726	.000[c]
	Residual	1854.967	182	10.192		
	Total	3148.385	186			
3	Regression	1361.161	6	226.860	22.848	.000[d]
	Residual	1787.224	180	9.929		
	Total	3148.385	186			
4	Regression	1362.993	7	194.713	19.522	.000[e]
	Residual	1785.392	179	9.974		
	Total	3148.385	186			
5	Regression	1380.836	11	125.531	12.428	.000[f]
	Residual	1767.549	175	10.100		
	Total	3148.385	186			
6	Regression	1391.371	12	115.948	11.482	.000[g]
	Residual	1757.014	174	10.098		
	Total	3148.385	186			
7	Regression	1395.655	13	107.358	10.597	.000[h]
	Residual	1752.730	173	10.131		
	Total	3148.385	186			

a. Dependent Variable: PETTL

b. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL

c. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, Enguse

d. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, Enguse, OPTGTL, TROG2TL

e. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, Enguse, OPTGTL, TROG2TL, LVLTTL

f. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, Enguse, OPTGTL, TROG2TL, LVLTTL, WMCF, WMCB, WMEF, WMEB

g. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, Enguse, OPTGTL, TROG2TL, LVLTTL, WMCF, WMCB, WMEF, WMEB, RAVENTL

h. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, Enguse, OPTGTL, TROG2TL, LVLTTL, WMCF, WMCB, WMEF, WMEB, RAVENTL, SPSAVER

Appendix 41

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.594[a]	.353	.342	3.336	.353	33.275	3	183	.000
2	.641[b]	.411	.398	3.193	.058	17.875	1	182	.000
3	.658[c]	.432	.413	3.151	.022	3.411	2	180	.035
4	.658[d]	.433	.411	3.158	.001	.184	1	179	.669
5	.662[e]	.439	.403	3.178	.006	.442	4	175	.778
6	.665[f]	.442	.403	3.178	.003	1.043	1	174	.308
7	.666[g]	.443	.401	3.183	.001	.423	1	173	.516

a. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL

b. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, Enguse

c. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, Enguse, OPTGTL, TROG2TL

d. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, Enguse, OPTGTL, TROG2TL, LVLTTL

e. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, Enguse, OPTGTL, TROG2TL, LVLTTL, WMCF, WMCB, WMEF, WMEB

f. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, Enguse, OPTGTL, TROG2TL, LVLTTL, WMCF, WMCB, WMEF, WMEB, RAVENTL

g. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, Enguse, OPTGTL, TROG2TL, LVLTTL, WMCF, WMCB, WMEF, WMEB, RAVENTL, SPSAVER

Appendix 42

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Correlations Zero-order	Correlations Partial	Correlations Part	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	4.266	3.479		1.226	.222					
	WSEGTTL	.003	.050	.004	.069	.945	.217	.005	.004	.868	1.152
	OPTLSNTL	.064	.048	.100	1.326	.187	.419	.098	.079	.627	1.594
	WRSTL	.100	.015	.527	6.872	.000	.589	.453	.409	.601	1.663
2	(Constant)	4.714	3.331		1.415	.159					
	WSEGTTL	.015	.048	.019	.318	.750	.217	.024	.018	.865	1.156
	OPTLSNTL	.067	.046	.105	1.461	.146	.419	.108	.083	.627	1.594
	WRSTL	.050	.018	.265	2.756	.006	.589	.200	.157	.351	2.851
	Enguse	.069	.016	.350	4.228	.000	.591	.299	.241	.473	2.115
3	(Constant)	2.132	3.456		.617	.538					
	WSEGTTL	-.010	.049	-.013	-.209	.835	.217	-.016	-.012	.830	1.204
	OPTLSNTL	.052	.046	.081	1.126	.262	.419	.084	.063	.610	1.640
	WRSTL	.026	.022	.140	1.227	.221	.589	.091	.069	.243	4.116
	Enguse	.077	.016	.393	4.705	.000	.591	.331	.264	.453	2.207
	TROG2TL	.026	.032	.069	.830	.408	.449	.062	.047	.460	2.173
	OPTGTL	.074	.032	.156	2.346	.020	.361	.172	.132	.717	1.396
4	(Constant)	2.833	3.830		.740	.461					
	WSEGTTL	-.007	.049	-.008	-.132	.895	.217	-.010	-.007	.806	1.240
	OPTLSNTL	.050	.046	.078	1.083	.280	.419	.081	.061	.605	1.652

	WRSTL	.033	.026	.172	1.257	.211	.589	.094	.071	.168	5.945
	Enguse	.077	.016	.395	4.711	.000	.591	.332	.265	.452	2.214
	TROG2TL	.028	.032	.072	.860	.391	.449	.064	.048	.457	2.188
	OPTGTL	.074	.032	.154	2.317	.022	.361	.171	.130	.715	1.399
	LVLTTL	-.010	.024	-.043	-.429	.669	.474	-.032	-.024	.311	3.214
5	(Constant)	1.375	4.563		.301	.763					
	WSEGTTL	-.008	.051	-.010	-.152	.879	.217	-.011	-.009	.770	1.299
	OPTLSNTL	.052	.047	.082	1.113	.267	.419	.084	.063	.595	1.680
	WRSTL	.031	.027	.164	1.168	.244	.589	.088	.066	.163	6.153
	Enguse	.073	.017	.372	4.306	.000	.591	.310	.244	.429	2.332
	TROG2TL	.022	.033	.057	.666	.506	.449	.050	.038	.442	2.265
	OPTGTL	.066	.033	.138	2.022	.045	.361	.151	.115	.685	1.459
	LVLTTL	-.007	.025	-.027	-.263	.793	.474	-.020	-.015	.300	3.334
	WMCF	.090	.186	.029	.480	.632	.162	.036	.027	.908	1.101
	WMCB	-.045	.093	-.033	-.485	.628	.207	-.037	-.027	.688	1.454
	WMEF	-.056	.146	-.025	-.382	.703	.247	-.029	-.022	.729	1.372
	WMEB	.251	.199	.097	1.262	.209	.370	.095	.071	.545	1.835
6	(Constant)	-.580	4.947		-.117	.907					
	WSEGTTL	-.009	.051	-.011	-.173	.863	.217	-.013	-.010	.770	1.299
	OPTLSNTL	.051	.047	.080	1.088	.278	.419	.082	.062	.595	1.681
	WRSTL	.033	.027	.177	1.253	.212	.589	.095	.071	.161	6.200
	Enguse	.070	.017	.359	4.109	.000	.591	.297	.233	.420	2.383
	TROG2TL	.016	.033	.042	.491	.624	.449	.037	.028	.430	2.326
	OPTGTL	.061	.033	.128	1.851	.066	.361	.139	.105	.671	1.491

	LVLTTL	-.005	.025	-.021	-.201	.841	.474	-.015	-.011	.299	3.347
	WMCF	.100	.187	.032	.536	.593	.162	.041	.030	.906	1.104
	WMCB	-.051	.093	-.037	-.541	.589	.207	-.041	-.031	.686	1.458
	WMEF	-.061	.146	-.028	-.419	.675	.247	-.032	-.024	.728	1.374
	WMEB	.256	.199	.099	1.286	.200	.370	.097	.073	.545	1.836
	RAVENTL	.053	.052	.061	1.021	.308	.198	.077	.058	.905	1.106
7	(Constant)	-1.098	5.019		-.219	.827					
	WSEGTTL	-.011	.051	-.014	-.211	.833	.217	-.016	-.012	.767	1.304
	OPTLSNTL	.054	.047	.084	1.135	.258	.419	.086	.064	.591	1.692
	WRSTL	.033	.027	.173	1.222	.223	.589	.092	.069	.161	6.212
	Enguse	.069	.017	.350	3.942	.000	.591	.287	.224	.408	2.448
	TROG2TL	.018	.034	.048	.548	.584	.449	.042	.031	.426	2.346
	OPTGTL	.061	.033	.127	1.832	.069	.361	.138	.104	.670	1.492
	LVLTTL	-.003	.025	-.013	-.120	.905	.474	-.009	-.007	.294	3.397
	WMCF	.101	.187	.032	.540	.590	.162	.041	.031	.905	1.104
	WMCB	-.050	.093	-.037	-.536	.593	.207	-.041	-.030	.686	1.458
	WMEF	-.054	.147	-.025	-.371	.711	.247	-.028	-.021	.724	1.381
	WMEB	.244	.200	.094	1.218	.225	.370	.092	.069	.540	1.852
	RAVENTL	.049	.052	.057	.947	.345	.198	.072	.054	.895	1.118
	SPSAVER	.000	.000	.039	.650	.516	.102	.049	.037	.906	1.103

a. Dependent Variable: PETTL

Appendix 43

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	22.000	1	22.000	27.897	.000[b]
	Residual	145.891	185	.789		
	Total	167.891	186			
2	Regression	22.679	2	11.339	14.368	.000[c]
	Residual	145.212	184	.789		
	Total	167.891	186			

a. Dependent Variable: ZCET4TL

b. Predictors: (Constant), WRSTL

c. Predictors: (Constant), WRSTL, WSEGTTL

Appendix 44

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.362[a]	.131	.126	.88803069	.131	27.897	1	185	.000
2	.368[b]	.135	.126	.88836581	.004	.860	1	184	.355

a. Predictors: (Constant), WRSTL

b. Predictors: (Constant), WRSTL, WSEGTTL

Appendix 45

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-.714	.148		-4.825	.000					
	WRSTL	.016	.003	.362	5.282	.000	.362	.362	.362	1.000	1.000
2	(Constant)	-1.212	.557		-2.175	.031					
	WRSTL	.015	.003	.338	4.623	.000	.362	.323	.317	.878	1.139
	WSEGTTL	.012	.013	.068	.928	.355	.186	.068	.064	.878	1.139

a. Dependent Variable: ZCET4TL

Appendix 46

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5.813	1	5.813	6.635	.011[b]
	Residual	162.078	185	.876		
	Total	167.891	186			
2	Regression	22.679	2	11.339	14.368	.000[c]
	Residual	145.212	184	.789		
	Total	167.891	186			

a. Dependent Variable: ZCET4TL

b. Predictors: (Constant), WSEGTTL

c. Predictors: (Constant), WSEGTTL, WRSTL

Appendix 47

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.186[a]	.035	.029	.93600028	.035	6.635	1	185	.011
2	.368[b]	.135	.126	.88836581	.100	21.371	1	184	.000

a. Predictors: (Constant), WSEGTTL

b. Predictors: (Constant), WSEGTTL, WRSTL

Appendix 48

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-1.504	.583		-2.578	.011					
	WSEGTTL	.034	.013	.186	2.576	.011	.186	.186	.186	1.000	1.000
2	(Constant)	-1.212	.557		-2.175	.031					
	WSEGTTL	.012	.013	.068	.928	.355	.186	.068	.064	.878	1.139
	WRSTL	.015	.003	.338	4.623	.000	.362	.323	.317	.878	1.139

a. Dependent Variable: ZCET4TL

Appendix 49

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	22.691	3	7.564	9.533	.000[b]
	Residual	145.200	183	.793		
	Total	167.891	186			
2	Regression	24.406	4	6.101	7.739	.000[c]
	Residual	143.485	182	.788		
	Total	167.891	186			
3	Regression	30.028	6	5.005	6.534	.000[d]
	Residual	137.862	180	.766		
	Total	167.891	186			

a. Dependent Variable: ZCET4TL

b. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL

c. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, LVLTTL

d. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, LVLTTL, OPTGTL, TROG2TL

Appendix 50

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.368[a]	.135	.121	.89075363	.135	9.533	3	183	.000
2	.381[b]	.145	.127	.88790707	.010	2.175	1	182	.142
3	.423[c]	.179	.151	.87515851	.033	3.671	2	180	.027

a. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL

b. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, LVLTTL

c. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, LVLTTL, OPTGTL, TROG2TL

Appendix 51

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-1.122	.929		-1.207	.229					
	WSEGTTL	.013	.013	.069	.933	.352	.186	.069	.064	.868	1.152
	OPTLSNTL	-.002	.013	-.011	-.122	.903	.218	-.009	-.008	.627	1.594
	WRSTL	.015	.004	.344	3.884	.000	.362	.276	.267	.601	1.663
2	(Constant)	-1.804	1.035		-1.743	.083					
	WSEGTTL	.009	.014	.050	.675	.501	.186	.050	.046	.844	1.185
	OPTLSNTL	-9.246E-5	.013	-.001	-.007	.994	.218	-.001	.000	.624	1.604
	WRSTL	.009	.006	.197	1.473	.142	.362	.109	.101	.263	3.797
	LVLTTL	.010	.007	.180	1.475	.142	.360	.109	.101	.315	3.179
3	(Constant)	-2.556	1.061		-2.408	.017					
	WSEGTTL	.001	.014	.008	.102	.919	.186	.008	.007	.807	1.240
	OPTLSNTL	-.004	.013	-.028	-.325	.746	.218	-.024	-.022	.605	1.652
	WRSTL	.004	.006	.086	.598	.551	.362	.045	.040	.222	4.514
	LVLTTL	.011	.007	.189	1.564	.120	.360	.116	.106	.312	3.204
	TROG2TL	.006	.009	.063	.637	.525	.314	.047	.043	.464	2.157
	OPTGTL	.022	.009	.198	2.502	.013	.326	.183	.169	.731	1.368

a. Dependent Variable: ZCET4TL

Appendix 52

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	22.691	3	7.564	9.533	.000[b]
	Residual	145.200	183	.793		
	Total	167.891	186			
2	Regression	28.155	5	5.631	7.294	.000[c]
	Residual	139.735	181	.772		
	Total	167.891	186			
3	Regression	30.028	6	5.005	6.534	.000[d]
	Residual	137.862	180	.766		
	Total	167.891	186			

a. Dependent Variable: ZCET4TL

b. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL

c. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL

d. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL, LVLTTL

Appendix 53

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.368[a]	.135	.121	.89075363	.135	9.533	3	183	.000
2	.410[b]	.168	.145	.87864554	.033	3.539	2	181	.031
3	.423[c]	.179	.151	.87515851	.011	2.445	1	180	.120

a. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL

b. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL

c. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL, LVLTTL

Appendix 54

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-1.122	.929		-1.207	.229					
	WSEGTTL	.013	.013	.069	.933	.352	.186	.069	.064	.868	1.152
	OPTLSNTL	-.002	.013	-.011	-.122	.903	.218	-.009	-.008	.627	1.594
	WRSTL	.015	.004	.344	3.884	.000	.362	.276	.267	.601	1.663
2	(Constant)	-1.846	.963		-1.916	.057					
	WSEGTTL	.005	.014	.028	.370	.712	.186	.028	.025	.831	1.204
	OPTLSNTL	-.006	.013	-.040	-.455	.650	.218	-.034	-.031	.610	1.640
	WRSTL	.010	.005	.236	2.204	.029	.362	.162	.149	.401	2.492
	TROG2TL	.007	.009	.075	.754	.452	.314	.056	.051	.466	2.145
	OPTGTL	.021	.009	.191	2.407	.017	.326	.176	.163	.733	1.364
3	(Constant)	-2.556	1.061		-2.408	.017					
	WSEGTTL	.001	.014	.008	.102	.919	.186	.008	.007	.807	1.240
	OPTLSNTL	-.004	.013	-.028	-.325	.746	.218	-.024	-.022	.605	1.652
	WRSTL	.004	.006	.086	.598	.551	.362	.045	.040	.222	4.514
	TROG2TL	.006	.009	.063	.637	.525	.314	.047	.043	.464	2.157
	OPTGTL	.022	.009	.198	2.502	.013	.326	.183	.169	.731	1.368
	LVLTTL	.011	.007	.189	1.564	.120	.360	.116	.106	.312	3.204

a. Dependent Variable: ZCET4TL

Appendix 55

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	21.796	1	21.796	27.600	.000[b]
	Residual	146.095	185	.790		
	Total	167.891	186			
2	Regression	24.406	4	6.101	7.739	.000[c]
	Residual	143.485	182	.788		
	Total	167.891	186			
3	Regression	30.028	6	5.005	6.534	.000[d]
	Residual	137.862	180	.766		
	Total	167.891	186			

a. Dependent Variable: ZCET4TL

b. Predictors: (Constant), LVLTTL

c. Predictors: (Constant), LVLTTL, WSEGTTL, OPTLSNTL, WRSTL

d. Predictors: (Constant), LVLTTL, WSEGTTL, OPTLSNTL, WRSTL, OPTGTL, TROG2TL

Appendix 56

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.360[a]	.130	.125	.88865222	.130	27.600	1	185	.000
2	.381[b]	.145	.127	.88790707	.016	1.104	3	182	.349
3	.423[c]	.179	.151	.87515851	.033	3.671	2	180	.027

a. Predictors: (Constant), LVLTTL

b. Predictors: (Constant), LVLTTL, WSEGTTL, OPTLSNTL, WRSTL

c. Predictors: (Constant), LVLTTL, WSEGTTL, OPTLSNTL, WRSTL, OPTGTL, TROG2TL

Appendix 57

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-2.037	.391		-5.211	.000					
	LVLTTL	.020	.004	.360	5.254	.000	.360	.360	.360	1.000	1.000
2	(Constant)	-1.804	1.035		-1.743	.083					
	LVLTTL	.010	.007	.180	1.475	.142	.360	.109	.101	.315	3.179
	WSEGTTL	.009	.014	.050	.675	.501	.186	.050	.046	.844	1.185
	OPTLSNTL	-9.246E-5	.013	-.001	-.007	.994	.218	-.001	.000	.624	1.604
	WRSTL	.009	.006	.197	1.473	.142	.362	.109	.101	.263	3.797
3	(Constant)	-2.556	1.061		-2.408	.017					
	LVLTTL	.011	.007	.189	1.564	.120	.360	.116	.106	.312	3.204

WSEGTTL	.001	.014	.008	.102	.919	.186	.008	.007	.807	1.240
OPTLSNTL	-.004	.013	-.028	-.325	.746	.218	-.024	-.022	.605	1.652
WRSTL	.004	.006	.086	.598	.551	.362	.045	.040	.222	4.514
TROG2TL	.006	.009	.063	.637	.525	.314	.047	.043	.464	2.157
OPTGTL	.022	.009	.198	2.502	.013	.326	.183	.169	.731	1.368

a. Dependent Variable: ZCET4TL

Appendix 58

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	21.796	1	21.796	27.600	.000[b]
	Residual	146.095	185	.790		
	Total	167.891	186			
2	Regression	29.735	3	9.912	13.129	.000[c]
	Residual	138.155	183	.755		
	Total	167.891	186			
3	Regression	30.028	6	5.005	6.534	.000[d]
	Residual	137.862	180	.766		
	Total	167.891	186			

a. Dependent Variable: ZCET4TL

b. Predictors: (Constant), LVLTTL

c. Predictors: (Constant), LVLTTL, OPTGTL, TROG2TL

d. Predictors: (Constant), LVLTTL, OPTGTL, TROG2TL, WSEGTTL, OPTLSNTL, WRSTL

Appendix 59

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.360[a]	.130	.125	.88865222	.130	27.600	1	185	.000
2	.421[b]	.177	.164	.86887624	.047	5.259	2	183	.006
3	.423[c]	.179	.151	.87515851	.002	.127	3	180	.944

a. Predictors: (Constant), LVLTTL

b. Predictors: (Constant), LVLTTL, OPTGTL, TROG2TL

c. Predictors: (Constant), LVLTTL, OPTGTL, TROG2TL, WSEGTTL, OPTLSNTL, WRSTL

Appendix 60

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-2.037	.391		-5.211	.000					
	LVLTTL	.020	.004	.360	5.254	.000	.360	.360	.360	1.000	1.000
2	(Constant)	-3.020	.488		-6.191	.000					
	LVLTTL	.013	.005	.236	2.760	.006	.360	.200	.185	.613	1.631
	TROG2TL	.007	.008	.080	.900	.369	.314	.066	.060	.572	1.747
	OPTGTL	.023	.008	.205	2.719	.007	.326	.197	.182	.793	1.262
3	(Constant)	-2.556	1.061		-2.408	.017					
	LVLTTL	.011	.007	.189	1.564	.120	.360	.116	.106	.312	3.204
	TROG2TL	.006	.009	.063	.637	.525	.314	.047	.043	.464	2.157

OPTGTL	.022	.009	.198	2.502	.013	.326	.183	.169	.731	1.368
WSEGTTL	.001	.014	.008	.102	.919	.186	.008	.007	.807	1.240
OPTLSNTL	-.004	.013	-.028	-.325	.746	.218	-.024	-.022	.605	1.652
WRSTL	.004	.006	.086	.598	.551	.362	.045	.040	.222	4.514

a. Dependent Variable: ZCET4TL

Appendix 61

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	22.691	3	7.564	9.533	.000[b]
	Residual	145.200	183	.793		
	Total	167.891	186			
2	Regression	28.155	5	5.631	7.294	.000[c]
	Residual	139.735	181	.772		
	Total	167.891	186			
3	Regression	30.028	6	5.005	6.534	.000[d]
	Residual	137.862	180	.766		
	Total	167.891	186			
4	Regression	32.775	10	3.277	4.269	.000[e]
	Residual	135.116	176	.768		
	Total	167.891	186			
5	Regression	33.410	11	3.037	3.952	.000[f]
	Residual	134.481	175	.768		
	Total	167.891	186			

6	Regression	37.160	12	3.097	4.122	.000[g]
	Residual	130.731	174	.751		
	Total	167.891	186			
7	Regression	47.399	13	3.646	5.235	.000[h]
	Residual	120.491	173	.696		
	Total	167.891	186			

a. Dependent Variable: ZCET4TL

b. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL

c. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL

d. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL, LVLTTL

e. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL, LVLTTL, WMCF, WMCB, WMEF, WMEB

f. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL, LVLTTL, WMCF, WMCB, WMEF, WMEB, RAVENTL

g. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL, LVLTTL, WMCF, WMCB, WMEF, WMEB, RAVENTL, SPSAVER

h. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL, LVLTTL, WMCF, WMCB, WMEF, WMEB, RAVENTL, SPSAVER, Enguse

Appendix 62

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.368[a]	.135	.121	.89075363	.135	9.533	3	183	.000
2	.410[b]	.168	.145	.87864554	.033	3.539	2	181	.031
3	.423[c]	.179	.151	.87515851	.011	2.445	1	180	.120
4	.442[d]	.195	.149	.87618746	.016	.894	4	176	.469
5	.446[e]	.199	.149	.87661949	.004	.827	1	175	.365
6	.470[f]	.221	.168	.86679013	.022	4.991	1	174	.027
7	.531[g]	.282	.228	.83455405	.061	14.702	1	173	.000

a. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL

b. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL

c. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL, LVLTTL

d. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL, LVLTTL, WMCF, WMCB, WMEF, WMEB

e. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL, LVLTTL, WMCF, WMCB, WMEF, WMEB, RAVENTL

f. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL, LVLTTL, WMCF, WMCB, WMEF, WMEB, RAVENTL, SPSAVER

g. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL, LVLTTL, WMCF, WMCB, WMEF, WMEB, RAVENTL, SPSAVER, Enguse

Appendix 63

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-1.122	.929		-1.207	.229					
	WSEGTTL	.013	.013	.069	.933	.352	.186	.069	.064	.868	1.152
	OPTLSNTL	-.002	.013	-.011	-.122	.903	.218	-.009	-.008	.627	1.594
	WRSTL	.015	.004	.344	3.884	.000	.362	.276	.267	.601	1.663
2	(Constant)	-1.846	.963		-1.916	.057					
	WSEGTTL	.005	.014	.028	.370	.712	.186	.028	.025	.831	1.204
	OPTLSNTL	-.006	.013	-.040	-.455	.650	.218	-.034	-.031	.610	1.640
	WRSTL	.010	.005	.236	2.204	.029	.362	.162	.149	.401	2.492
	TROG2TL	.007	.009	.075	.754	.452	.314	.056	.051	.466	2.145
	OPTGTL	.021	.009	.191	2.407	.017	.326	.176	.163	.733	1.364
3	(Constant)	-2.556	1.061		-2.408	.017					
	WSEGTTL	.001	.014	.008	.102	.919	.186	.008	.007	.807	1.240
	OPTLSNTL	-.004	.013	-.028	-.325	.746	.218	-.024	-.022	.605	1.652
	WRSTL	.004	.006	.086	.598	.551	.362	.045	.040	.222	4.514
	TROG2TL	.006	.009	.063	.637	.525	.314	.047	.043	.464	2.157
	OPTGTL	.022	.009	.198	2.502	.013	.326	.183	.169	.731	1.368
	LVLTTL	.011	.007	.189	1.564	.120	.360	.116	.106	.312	3.204
4	(Constant)	-1.578	1.255		-1.257	.210					
	WSEGTTL	.000	.014	.001	.014	.989	.186	.001	.001	.772	1.295
	OPTLSNTL	-.007	.013	-.046	-.526	.599	.218	-.040	-.036	.595	1.679

	WRSTL	.004	.007	.097	.648	.518	.362	.049	.044	.204	4.906
	TROG2TL	.006	.009	.072	.710	.479	.314	.053	.048	.449	2.225
	OPTGTL	.024	.009	.221	2.754	.007	.326	.203	.186	.708	1.412
	LVLTTL	.010	.007	.180	1.459	.146	.360	.109	.099	.302	3.316
	WMCF	-.074	.051	-.102	-1.451	.149	.016	-.109	-.098	.916	1.091
	WMCB	.007	.026	.021	.258	.797	.096	.019	.017	.688	1.454
	WMEF	.051	.040	.101	1.276	.204	.204	.096	.086	.729	1.372
	WMEB	-.045	.054	-.074	-.823	.412	.142	-.062	-.056	.560	1.785
5	(Constant)	-1.114	1.355		-.822	.412					
	WSEGTTL	.000	.014	.002	.025	.980	.186	.002	.002	.772	1.295
	OPTLSNTL	-.006	.013	-.044	-.501	.617	.218	-.038	-.034	.595	1.681
	WRSTL	.004	.007	.094	.627	.532	.362	.047	.042	.204	4.909
	TROG2TL	.008	.009	.085	.832	.407	.314	.063	.056	.440	2.272
	OPTGTL	.025	.009	.230	2.842	.005	.326	.210	.192	.698	1.432
	LVLTTL	.010	.007	.174	1.412	.160	.360	.106	.096	.301	3.324
	WMCF	-.076	.051	-.105	-1.486	.139	.016	-.112	-.101	.915	1.093
	WMCB	.008	.026	.025	.308	.758	.096	.023	.021	.686	1.458
	WMEF	.053	.040	.104	1.311	.192	.204	.099	.089	.728	1.374
	WMEB	-.045	.054	-.074	-.822	.412	.142	-.062	-.056	.560	1.785
	RAVENTL	-.013	.014	-.064	-.909	.365	.027	-.069	-.062	.925	1.081
6	(Constant)	-1.535	1.353		-1.134	.258					
	WSEGTTL	-.001	.014	-.007	-.087	.930	.186	-.007	-.006	.770	1.299
	OPTLSNTL	-.004	.013	-.029	-.333	.739	.218	-.025	-.022	.592	1.691
	WRSTL	.002	.007	.050	.337	.737	.362	.026	.023	.200	4.994

	TROG2TL	.010	.009	.111	1.097	.274	.314	.083	.073	.434	2.303
	OPTGTL	.026	.009	.232	2.895	.004	.326	.214	.194	.698	1.433
	LVLTTL	.011	.007	.204	1.660	.099	.360	.125	.111	.297	3.363
	WMCF	-.077	.051	-.106	-1.523	.130	.016	-.115	-.102	.915	1.093
	WMCB	.008	.025	.027	.329	.743	.096	.025	.022	.686	1.458
	WMEF	.059	.040	.116	1.474	.142	.204	.111	.099	.724	1.380
	WMEB	-.059	.054	-.098	-1.093	.276	.142	-.083	-.073	.552	1.811
	RAVENTL	-.017	.014	-.084	-1.196	.233	.027	-.090	-.080	.910	1.099
	SPSAVER	.000	.000	.155	2.234	.027	.119	.167	.149	.931	1.074
7	(Constant)	-2.242	1.316		-1.704	.090					
	WSEGTTL	-.004	.013	-.025	-.336	.737	.186	-.026	-.022	.767	1.304
	OPTLSNTL	-.003	.012	-.019	-.231	.817	.218	-.018	-.015	.591	1.692
	WRSTL	.014	.007	.323	2.011	.046	.362	.151	.130	.161	6.212
	TROG2TL	.005	.009	.060	.612	.541	.314	.046	.039	.426	2.346
	OPTGTL	.019	.009	.171	2.179	.031	.326	.163	.140	.670	1.492
	LVLTTL	.014	.007	.249	2.101	.037	.360	.158	.135	.294	3.397
	WMCF	-.058	.049	-.081	-1.189	.236	.016	-.090	-.077	.905	1.104
	WMCB	.008	.025	.026	.331	.741	.096	.025	.021	.686	1.458
	WMEF	.063	.038	.124	1.639	.103	.204	.124	.106	.724	1.381
	WMEB	-.029	.053	-.049	-.555	.579	.142	-.042	-.036	.540	1.852
	RAVENTL	-.010	.014	-.051	-.742	.459	.027	-.056	-.048	.895	1.118
	SPSAVER	.000	.000	.197	2.914	.004	.119	.216	.188	.906	1.103
	Enguse	-.018	.005	-.386	-3.834	.000	.083	-.280	-.247	.408	2.448

a. Dependent Variable: ZCET4TL

Appendix 64

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	22.691	3	7.564	9.533	.000[b]
	Residual	145.200	183	.793		
	Total	167.891	186			
2	Regression	28.155	5	5.631	7.294	.000[c]
	Residual	139.735	181	.772		
	Total	167.891	186			
3	Regression	36.772	6	6.129	8.414	.000[d]
	Residual	131.118	180	.728		
	Total	167.891	186			
4	Regression	41.052	7	5.865	8.276	.000[e]
	Residual	126.838	179	.709		
	Total	167.891	186			
5	Regression	44.530	8	5.566	8.032	.000[f]
	Residual	123.360	178	.693		
	Total	167.891	186			
6	Regression	47.016	12	3.918	5.640	.000[g]
	Residual	120.875	174	.695		
	Total	167.891	186			
7	Regression	47.399	13	3.646	5.235	.000[h]
	Residual	120.491	173	.696		
	Total	167.891	186			

a. Dependent Variable: ZCET4TL

b. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL

c. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL

d. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL, Enguse

e. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL, Enguse, SPSAVER

f. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL, Enguse, SPSAVER, LVLTTL

g. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL, Enguse, SPSAVER, LVLTTL, WMCF, WMCB, WMEF, WMEB

h. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL, Enguse, SPSAVER, LVLTTL, WMCF, WMCB, WMEF, WMEB, RAVENTL

Appendix 65

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.368[a]	.135	.121	.89075363	.135	9.533	3	183	.000
2	.410[b]	.168	.145	.87864554	.033	3.539	2	181	.031
3	.468[c]	.219	.193	.85348321	.051	11.830	1	180	.001
4	.494[d]	.245	.215	.84178015	.025	6.040	1	179	.015
5	.515[e]	.265	.232	.83248708	.021	5.019	1	178	.026
6	.529[f]	.280	.230	.83347621	.015	.894	4	174	.469
7	.531[g]	.282	.228	.83455405	.002	.551	1	173	.459

a. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL

b. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL

c. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL, Enguse

d. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL, Enguse, SPSAVER

e. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL, Enguse, SPSAVER, LVLTTL

f. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL, Enguse, SPSAVER, LVLTTL, WMCF, WMCB, WMEF, WMEB

g. Predictors: (Constant), WRSTL, WSEGTTL, OPTLSNTL, OPTGTL, TROG2TL, Enguse, SPSAVER, LVLTTL, WMCF, WMCB, WMEF, WMEB, RAVENTL

Appendix 66

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Correlations			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-1.122	.929		-1.207	.229					
	WSEGTTL	.013	.013	.069	.933	.352	.186	.069	.064	.868	1.152
	OPTLSNTL	-.002	.013	-.011	-.122	.903	.218	-.009	-.008	.627	1.594
	WRSTL	.015	.004	.344	3.884	.000	.362	.276	.267	.601	1.663
2	(Constant)	-1.846	.963		-1.916	.057					
	WSEGTTL	.005	.014	.028	.370	.712	.186	.028	.025	.831	1.204
	OPTLSNTL	-.006	.013	-.040	-.455	.650	.218	-.034	-.031	.610	1.640
	WRSTL	.010	.005	.236	2.204	.029	.362	.162	.149	.401	2.492
	TROG2TL	.007	.009	.075	.754	.452	.314	.056	.051	.466	2.145
	OPTGTL	.021	.009	.191	2.407	.017	.326	.176	.163	.733	1.364
3	(Constant)	-1.741	.936		-1.859	.065					
	WSEGTTL	.004	.013	.023	.323	.747	.186	.024	.021	.830	1.204
	OPTLSNTL	-.005	.012	-.036	-.421	.674	.218	-.031	-.028	.610	1.640
	WRSTL	.023	.006	.525	3.926	.000	.362	.281	.259	.243	4.116
	TROG2TL	.003	.009	.037	.377	.706	.314	.028	.025	.460	2.173
	OPTGTL	.017	.009	.150	1.928	.055	.326	.142	.127	.717	1.396
	Enguse	-.015	.004	-.337	-3.439	.001	.083	-.248	-.227	.453	2.207
4	(Constant)	-2.221	.944		-2.353	.020					
	WSEGTTL	.003	.013	.018	.247	.805	.186	.018	.016	.829	1.206
	OPTLSNTL	-.003	.012	-.021	-.252	.801	.218	-.019	-.016	.607	1.649

	WRSTL	.024	.006	.539	4.084	.000	.362	.292	.265	.242	4.124
	TROG2TL	.005	.009	.056	.586	.558	.314	.044	.038	.457	2.189
	OPTGTL	.015	.008	.139	1.805	.073	.326	.134	.117	.714	1.401
	Enguse	-.017	.004	-.381	-3.882	.000	.083	-.279	-.252	.438	2.285
	SPSAVER	.000	.000	.164	2.458	.015	.119	.181	.160	.948	1.055
5	(Constant)	-3.261	1.043		-3.128	.002					
	WSEGTTL	-.002	.013	-.011	-.151	.880	.186	-.011	-.010	.803	1.245
	OPTLSNTL	.000	.012	-.003	-.039	.969	.218	-.003	-.003	.601	1.664
	WRSTL	.015	.007	.344	2.192	.030	.362	.162	.141	.168	5.958
	TROG2TL	.004	.008	.042	.436	.663	.314	.033	.028	.455	2.199
	OPTGTL	.016	.008	.146	1.912	.057	.326	.142	.123	.713	1.403
	Enguse	-.018	.004	-.400	-4.099	.000	.083	-.294	-.263	.435	2.301
	SPSAVER	.000	.000	.186	2.791	.006	.119	.205	.179	.927	1.079
	LVLTTL	.015	.007	.261	2.240	.026	.360	.166	.144	.304	3.287
6	(Constant)	-2.597	1.224		-2.121	.035					
	WSEGTTL	-.005	.013	-.025	-.346	.730	.186	-.026	-.022	.767	1.304
	OPTLSNTL	-.003	.012	-.021	-.256	.798	.218	-.019	-.016	.592	1.690
	WRSTL	.015	.007	.334	2.091	.038	.362	.157	.134	.162	6.160
	TROG2TL	.004	.009	.048	.494	.622	.314	.037	.032	.439	2.278
	OPTGTL	.018	.009	.163	2.097	.037	.326	.157	.135	.684	1.462
	Enguse	-.018	.005	-.396	-3.967	.000	.083	-.288	-.255	.415	2.408
	SPSAVER	.000	.000	.192	2.856	.005	.119	.212	.184	.916	1.092
	LVLTTL	.014	.007	.254	2.141	.034	.360	.160	.138	.295	3.390
	WMCF	-.056	.049	-.078	-1.152	.251	.016	-.087	-.074	.908	1.101

	WMCB	.007	.024	.023	.290	.772	.096	.022	.019	.688	1.454
	WMEF	.062	.038	.122	1.609	.109	.204	.121	.104	.725	1.379
	WMEB	-.028	.052	-.046	-.531	.596	.142	-.040	-.034	.541	1.850
7	(Constant)	-2.242	1.316		-1.704	.090					
	WSEGTTL	-.004	.013	-.025	-.336	.737	.186	-.026	-.022	.767	1.304
	OPTLSNTL	-.003	.012	-.019	-.231	.817	.218	-.018	-.015	.591	1.692
	WRSTL	.014	.007	.323	2.011	.046	.362	.151	.130	.161	6.212
	TROG2TL	.005	.009	.060	.612	.541	.314	.046	.039	.426	2.346
	OPTGTL	.019	.009	.171	2.179	.031	.326	.163	.140	.670	1.492
	Enguse	-.018	.005	-.386	-3.834	.000	.083	-.280	-.247	.408	2.448
	SPSAVER	.000	.000	.197	2.914	.004	.119	.216	.188	.906	1.103
	LVLTTL	.014	.007	.249	2.101	.037	.360	.158	.135	.294	3.397
	WMCF	-.058	.049	-.081	-1.189	.236	.016	-.090	-.077	.905	1.104
	WMCB	.008	.025	.026	.331	.741	.096	.025	.021	.686	1.458
	WMEF	.063	.038	.124	1.639	.103	.204	.124	.106	.724	1.381
	WMEB	-.029	.053	-.049	-.555	.579	.142	-.042	-.036	.540	1.852
	RAVENTL	-.010	.014	-.051	-.742	.459	.027	-.056	-.048	.895	1.118

a. Dependent Variable: ZCET4TL

Appendix 67

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	33.401	2	16.700	24.966	.000[b]
	Residual	96.325	144	.669		
	Total	129.726	146			
2	Regression	34.350	3	11.450	17.168	.000[c]
	Residual	95.375	143	.667		
	Total	129.726	146			
3	Regression	34.450	5	6.890	10.197	.000[d]
	Residual	95.276	141	.676		
	Total	129.726	146			
4	Regression	36.735	6	6.122	9.218	.000[e]
	Residual	92.991	140	.664		
	Total	129.726	146			
5	Regression	36.947	7	5.278	7.908	.000[f]
	Residual	92.778	139	.667		
	Total	129.726	146			
6	Regression	42.546	8	5.318	8.418	.000[g]
	Residual	87.179	138	.632		
	Total	129.726	146			

a. Dependent Variable: ZCET4TL

b. Predictors: (Constant), OPTLSNTL, WRSTL

c. Predictors: (Constant), OPTLSNTL, WRSTL, LVLTTL

d. Predictors: (Constant), OPTLSNTL, WRSTL, LVLTTL, OPTGTL, TROG2TL

e. Predictors: (Constant), OPTLSNTL, WRSTL, LVLTTL, OPTGTL, TROG2TL, WMEF

f. Predictors: (Constant), OPTLSNTL, WRSTL, LVLTTL, OPTGTL, TROG2TL, WMEF, Enguse

g. Predictors: (Constant), OPTLSNTL, WRSTL, LVLTTL, OPTGTL, TROG2TL, WMEF, Enguse, SPSAVER

Appendix 68

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.507[a]	.257	.247	.81787760	.257	24.966	2	144	.000
2	.515[b]	.265	.249	.81667539	.007	1.424	1	143	.235
3	.515[c]	.266	.240	.82201877	.001	.073	2	141	.929
4	.532[d]	.283	.252	.81499688	.018	3.440	1	140	.066
5	.534[e]	.285	.249	.81698850	.002	.318	1	139	.574
6	.573[f]	.328	.289	.79481777	.043	8.863	1	138	.003

a. Predictors: (Constant), OPTLSNTL, WRSTL

b. Predictors: (Constant), OPTLSNTL, WRSTL, LVLTTL

c. Predictors: (Constant), OPTLSNTL, WRSTL, LVLTTL, OPTGTL, TROG2TL

d. Predictors: (Constant), OPTLSNTL, WRSTL, LVLTTL, OPTGTL, TROG2TL, WMEF

e. Predictors: (Constant), OPTLSNTL, WRSTL, LVLTTL, OPTGTL, TROG2TL, WMEF, Enguse

f. Predictors: (Constant), OPTLSNTL, WRSTL, LVLTTL, OPTGTL, TROG2TL, WMEF, Enguse, SPSAVER

Appendix 69

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	Correlations Zero-order	Partial	Part	Collinearity Statistics Tolerance	VIF
1	(Constant)	-1.227	.868		-1.413	.160					
	WRSTL	.030	.005	.502	6.036	.000	.507	.449	.433	.745	1.342
	OPTLSNTL	.002	.014	.011	.127	.899	.264	.011	.009	.745	1.342
2	(Constant)	-1.916	1.042		-1.840	.068					
	WRSTL	.025	.007	.410	3.627	.000	.507	.290	.260	.402	2.490
	OPTLSNTL	.003	.014	.019	.228	.820	.264	.019	.016	.740	1.351
	LVLTTL	.008	.007	.123	1.193	.235	.421	.099	.086	.488	2.051
3	(Constant)	-2.021	1.145		-1.765	.080					
	WRSTL	.024	.008	.406	3.192	.002	.507	.260	.230	.322	3.105
	OPTLSNTL	.003	.014	.021	.248	.805	.264	.021	.018	.721	1.387
	LVLTTL	.009	.007	.129	1.233	.220	.421	.103	.089	.473	2.112
	TROG2TL	-.002	.009	-.022	-.238	.812	.307	-.020	-.017	.593	1.685
	OPTGTL	.003	.009	.027	.331	.741	.237	.028	.024	.771	1.297
4	(Constant)	-2.277	1.144		-1.991	.048					
	WRSTL	.024	.008	.406	3.223	.002	.507	.263	.231	.322	3.105
	OPTLSNTL	.002	.014	.009	.112	.911	.264	.009	.008	.717	1.395
	LVLTTL	.008	.007	.117	1.121	.264	.421	.094	.080	.471	2.121
	TROG2TL	-.003	.009	-.032	-.344	.732	.307	-.029	-.025	.592	1.690
	OPTGTL	.003	.009	.028	.338	.736	.237	.029	.024	.771	1.297

	WMEF	.069	.037	.135	1.855	.066	.211	.155	.133	.964	1.037
5	(Constant)	-2.309	1.148		-2.011	.046					
	WRSTL	.024	.008	.404	3.199	.002	.507	.262	.229	.322	3.107
	OPTLSNTL	.001	.014	.008	.094	.926	.264	.008	.007	.716	1.396
	LVLTTL	.008	.007	.113	1.080	.282	.421	.091	.077	.470	2.130
	TROG2TL	-.003	.009	-.029	-.309	.758	.307	-.026	-.022	.590	1.696
	OPTGTL	.003	.009	.028	.347	.729	.237	.029	.025	.771	1.298
	WMEF	.070	.037	.139	1.895	.060	.211	.159	.136	.956	1.046
	Enguse	.010	.017	.041	.564	.574	.065	.048	.040	.980	1.021
6	(Constant)	-3.106	1.149		-2.704	.008					
	WRSTL	.023	.007	.389	3.156	.002	.507	.259	.220	.321	3.113
	OPTLSNTL	.005	.014	.027	.330	.742	.264	.028	.023	.712	1.405
	LVLTTL	.010	.007	.145	1.420	.158	.421	.120	.099	.464	2.154
	TROG2TL	-2.013E-5	.008	.000	-.002	.998	.307	.000	.000	.583	1.715
	OPTGTL	.000	.009	-.002	-.026	.980	.237	-.002	-.002	.758	1.319
	WMEF	.070	.036	.139	1.941	.054	.211	.163	.135	.956	1.046
	Enguse	.007	.017	.031	.440	.661	.065	.037	.031	.978	1.023
	SPSAVER	.000	.000	.214	2.977	.003	.148	.246	.208	.943	1.061

a. Dependent Variable: ZCET4TL

Appendix 70

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	33.401	2	16.700	24.966	.000[b]
	Residual	96.325	144	.669		
	Total	129.726	146			
2	Regression	38.485	3	12.828	20.105	.000[c]
	Residual	91.241	143	.638		
	Total	129.726	146			
3	Regression	40.120	4	10.030	15.895	.000[d]
	Residual	89.606	142	.631		
	Total	129.726	146			
4	Regression	40.124	6	6.687	10.449	.000[e]
	Residual	89.601	140	.640		
	Total	129.726	146			
5	Regression	42.424	7	6.061	9.649	.000[f]
	Residual	87.302	139	.628		
	Total	129.726	146			
6	Regression	42.546	8	5.318	8.418	.000[g]
	Residual	87.179	138	.632		
	Total	129.726	146			

a. Dependent Variable: ZCET4TL

b. Predictors: (Constant), OPTLSNTL, WRSTL

c. Predictors: (Constant), OPTLSNTL, WRSTL, SPSAVER

d. Predictors: (Constant), OPTLSNTL, WRSTL, SPSAVER, LVLTTL

e. Predictors: (Constant), OPTLSNTL, WRSTL, SPSAVER, LVLTTL, OPTGTL, TROG2TL

f. Predictors: (Constant), OPTLSNTL, WRSTL, SPSAVER, LVLTTL, OPTGTL, TROG2TL, WMEF

g. Predictors: (Constant), OPTLSNTL, WRSTL, SPSAVER, LVLTTL, OPTGTL, TROG2TL, WMEF, Enguse

Appendix 71

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.507[a]	.257	.247	.81787760	.257	24.966	2	144	.000
2	.545[b]	.297	.282	.79877940	.039	7.968	1	143	.005
3	.556[c]	.309	.290	.79437179	.013	2.591	1	142	.110
4	.556[d]	.309	.280	.80000544	.000	.004	2	140	.996
5	.572[e]	.327	.293	.79250929	.018	3.661	1	139	.058
6	.573[f]	.328	.289	.79481777	.001	.194	1	138	.661

a. Predictors: (Constant), OPTLSNTL, WRSTL

b. Predictors: (Constant), OPTLSNTL, WRSTL, SPSAVER

c. Predictors: (Constant), OPTLSNTL, WRSTL, SPSAVER, LVLTTL

d. Predictors: (Constant), OPTLSNTL, WRSTL, SPSAVER, LVLTTL, OPTGTL, TROG2TL

e. Predictors: (Constant), OPTLSNTL, WRSTL, SPSAVER, LVLTTL, OPTGTL, TROG2TL, WMEF

f. Predictors: (Constant), OPTLSNTL, WRSTL, SPSAVER, LVLTTL, OPTGTL, TROG2TL, WMEF, Enguse

Appendix 72

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Correlations Zero-order	Correlations Partial	Correlations Part	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	-1.227	.868		-1.413	.160					
	WRSTL	.030	.005	.502	6.036	.000	.507	.449	.433	.745	1.342
	OPTLSNTL	.002	.014	.011	.127	.899	.264	.011	.009	.745	1.342
2	(Constant)	-1.869	.878		-2.129	.035					
	WRSTL	.031	.005	.512	6.294	.000	.507	.466	.441	.744	1.344
	OPTLSNTL	.005	.013	.029	.351	.726	.264	.029	.025	.741	1.350
	SPSAVER	.000	.000	.199	2.823	.005	.148	.230	.198	.985	1.015
3	(Constant)	-2.829	1.057		-2.676	.008					
	WRSTL	.023	.007	.391	3.549	.001	.507	.285	.247	.400	2.498
	OPTLSNTL	.007	.013	.041	.505	.615	.264	.042	.035	.734	1.362
	SPSAVER	.000	.000	.214	3.024	.003	.148	.246	.211	.968	1.033
	LVLTTL	.011	.007	.162	1.610	.110	.421	.134	.112	.479	2.086
4	(Constant)	-2.830	1.147		-2.467	.015					
	WRSTL	.023	.007	.389	3.144	.002	.507	.257	.221	.321	3.111
	OPTLSNTL	.007	.014	.040	.485	.628	.264	.041	.034	.717	1.395
	SPSAVER	.000	.000	.215	2.978	.003	.148	.244	.209	.945	1.058
	LVLTTL	.011	.007	.161	1.568	.119	.421	.131	.110	.468	2.135
	TROG2TL	.001	.008	.007	.080	.937	.307	.007	.006	.586	1.705
	OPTGTL	.000	.009	-.003	-.040	.968	.237	-.003	-.003	.759	1.318

5	(Constant)	-3.087	1.144		-2.698	.008					
	WRSTL	.023	.007	.390	3.177	.002	.507	.260	.221	.321	3.111
	OPTLSNTL	.005	.014	.029	.347	.729	.264	.029	.024	.713	1.403
	SPSAVER	.000	.000	.215	3.010	.003	.148	.247	.209	.945	1.058
	LVLTTL	.010	.007	.149	1.458	.147	.421	.123	.101	.466	2.144
	TROG2TL	.000	.008	-.002	-.026	.979	.307	-.002	-.002	.585	1.710
	OPTGTL	.000	.009	-.003	-.036	.971	.237	-.003	-.003	.759	1.318
	WMEF	.069	.036	.136	1.913	.058	.211	.160	.133	.964	1.037
6	(Constant)	-3.106	1.149		-2.704	.008					
	WRSTL	.023	.007	.389	3.156	.002	.507	.259	.220	.321	3.113
	OPTLSNTL	.005	.014	.027	.330	.742	.264	.028	.023	.712	1.405
	SPSAVER	.000	.000	.214	2.977	.003	.148	.246	.208	.943	1.061
	LVLTTL	.010	.007	.145	1.420	.158	.421	.120	.099	.464	2.154
	TROG2TL	-2.013E-5	.008	.000	-.002	.998	.307	.000	.000	.583	1.715
	OPTGTL	.000	.009	-.002	-.026	.980	.237	-.002	-.002	.758	1.319
	WMEF	.070	.036	.139	1.941	.054	.211	.163	.135	.956	1.046
	Enguse	.007	.017	.031	.440	.661	.065	.037	.031	.978	1.023

a. Dependent Variable: ZCET4TL

Appendix 73

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	15.410	1	15.410	25.753	.000[b]
	Residual	22.739	38	.598		
	Total	38.150	39			
2	Regression	16.338	2	8.169	13.857	.000[c]
	Residual	21.812	37	.590		
	Total	38.150	39			
3	Regression	16.750	3	5.583	9.393	.000[d]
	Residual	21.399	36	.594		
	Total	38.150	39			
4	Regression	19.921	4	4.980	9.562	.000[e]
	Residual	18.229	35	.521		
	Total	38.150	39			

a. Dependent Variable: ZCET4TL

b. Predictors: (Constant), OPTGTL

c. Predictors: (Constant), OPTGTL, WRSTL

d. Predictors: (Constant), OPTGTL, WRSTL, LVLTTL

e. Predictors: (Constant), OPTGTL, WRSTL, LVLTTL, RAVENTL

Appendix 74

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.636[a]	.404	.388	.77356480	.404	25.753	1	38	.000
2	.654[b]	.428	.397	.76779659	.024	1.573	1	37	.218
3	.663[c]	.439	.392	.77098950	.011	.694	1	36	.410
4	.723[d]	.522	.468	.72167997	.083	6.088	1	35	.019

a. Predictors: (Constant), OPTGTL

b. Predictors: (Constant), OPTGTL, WRSTL

c. Predictors: (Constant), OPTGTL, WRSTL, LVLTTL

d. Predictors: (Constant), OPTGTL, WRSTL, LVLTTL, RAVENTL

Appendix 75

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-4.041	.807		-5.009	.000					
	OPTGTL	.070	.014	.636	5.075	.000	.636	.636	.636	1.000	1.000
2	(Constant)	-4.355	.839		-5.191	.000					
	OPTGTL	.053	.019	.480	2.740	.009	.636	.411	.341	.502	1.990
	WRSTL	.017	.014	.220	1.254	.218	.559	.202	.156	.502	1.990
3	(Constant)	-5.090	1.220		-4.173	.000					
	OPTGTL	.043	.022	.395	1.937	.061	.636	.307	.242	.375	2.666
	WRSTL	.015	.014	.186	1.027	.311	.559	.169	.128	.477	2.098

	LVLTTL	.012	.015	.153	.833	.410	.551	.138	.104	.462	2.166
4	(Constant)	-7.297	1.450		-5.031	.000					
	OPTGTL	.034	.021	.312	1.613	.116	.636	.263	.188	.364	2.748
	WRSTL	.009	.014	.111	.647	.522	.559	.109	.076	.462	2.165
	LVLTTL	.015	.014	.182	1.054	.299	.551	.175	.123	.460	2.176
	RAVENTL	.063	.025	.315	2.467	.019	.513	.385	.288	.840	1.191

a. Dependent Variable: ZCET4TL

Appendix 76

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	15.410	1	15.410	25.753	.000[b]
	Residual	22.739	38	.598		
	Total	38.150	39			
2	Regression	18.902	2	9.451	18.168	.000[c]
	Residual	19.248	37	.520		
	Total	38.150	39			
3	Regression	19.343	3	6.448	12.342	.000[d]
	Residual	18.807	36	.522		
	Total	38.150	39			
4	Regression	19.921	4	4.980	9.562	.000[e]
	Residual	18.229	35	.521		
	Total	38.150	39			

a. Dependent Variable: ZCET4TL

b. Predictors: (Constant), OPTGTL

c. Predictors: (Constant), OPTGTL, RAVENTL

d. Predictors: (Constant), OPTGTL, RAVENTL, WRSTL

e. Predictors: (Constant), OPTGTL, RAVENTL, WRSTL, LVLTTL

Appendix 77

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.636[a]	.404	.388	.77356480	.404	25.753	1	38	.000
2	.704[b]	.495	.468	.72125451	.092	6.712	1	37	.014
3	.712[c]	.507	.466	.72278221	.012	.844	1	36	.364
4	.723[d]	.522	.468	.72167997	.015	1.110	1	35	.299

a. Predictors: (Constant), OPTGTL

b. Predictors: (Constant), OPTGTL, RAVENTL

c. Predictors: (Constant), OPTGTL, RAVENTL, WRSTL

d. Predictors: (Constant), OPTGTL, RAVENTL, WRSTL, LVLTTL

Appendix 78

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-4.041	.807		-5.009	.000					
	OPTGTL	.070	.014	.636	5.075	.000	.636	.636	.636	1.000	1.000
2	(Constant)	-6.280	1.146		-5.482	.000					
	OPTGTL	.057	.014	.517	4.125	.000	.636	.561	.482	.867	1.153
	RAVENTL	.065	.025	.325	2.591	.014	.513	.392	.303	.867	1.153
3	(Constant)	-6.366	1.152		-5.527	.000					
	OPTGTL	.046	.018	.416	2.486	.018	.636	.383	.291	.489	2.043
	RAVENTL	.061	.025	.306	2.398	.022	.513	.371	.281	.844	1.185
	WRSTL	.012	.013	.154	.919	.364	.559	.151	.107	.489	2.046
4	(Constant)	-7.297	1.450		-5.031	.000					
	OPTGTL	.034	.021	.312	1.613	.116	.636	.263	.188	.364	2.748
	RAVENTL	.063	.025	.315	2.467	.019	.513	.385	.288	.840	1.191
	WRSTL	.009	.014	.111	.647	.522	.559	.109	.076	.462	2.165
	LVLTTL	.015	.014	.182	1.054	.299	.551	.175	.123	.460	2.176

a. Dependent Variable: ZCET4TL

Appendix 79

Assessment of normality (Group number 1)

Variable	min	max	skew	c.r.	kurtosis	c.r.
LC4	.000	6.000	-.083	-.464	-.329	-.918
LC3	.000	6.000	-.004	-.021	-.415	-1.158
LC2	1.000	6.000	.387	2.159	-.615	-1.718
LC1	1.000	7.000	.302	1.684	-.478	-1.335
EU1	.000	90.000	2.094	11.692	3.721	10.387
EU2	.000	92.000	1.770	9.882	2.114	5.902
EU3	.000	98.000	1.855	10.358	2.508	7.001
EU4	.000	100.000	1.532	8.555	1.012	2.824
RA5	1.000	8.000	.011	.060	-.844	-2.355
RA4	4.000	12.000	.069	.387	-.732	-2.045
RA3	6.000	12.000	-.442	-2.469	-.512	-1.430
RA2	10.000	12.000	-1.668	-9.311	1.473	4.113
RA1	10.000	12.000	-1.373	-7.666	.800	2.232
WM4	2.000	8.000	-.189	-1.056	-.749	-2.089
WM3	4.000	12.000	.220	1.226	-.270	-.754
WM2	3.000	14.000	.082	.455	-.876	-2.446
WM1	12.000	16.000	-1.339	-7.476	.548	1.529
SPS4	183.500	4767.000	.826	4.610	.245	.685
SPS3	.000	5096.000	.798	4.453	-.001	-.002
SPS2	232.667	5342.000	1.117	6.237	.754	2.105
SPS1	.000	5859.667	.836	4.665	.399	1.114
GK5	17.000	33.000	-.245	-1.367	-.584	-1.629
GK4	19.000	41.000	-.204	-1.141	-.626	-1.747
GK3	11.000	24.000	-.542	-3.023	-.664	-1.853
GK2	11.000	28.000	-.437	-2.438	-.811	-2.263
GK1	15.000	28.000	-.326	-1.818	-.598	-1.669
VK6	7.000	28.000	.081	.452	-.860	-2.401
VK5	8.000	20.000	-.102	-.570	-.688	-1.919
VK4	6.000	19.000	.093	.518	-.616	-1.719
VK3	7.000	22.000	-.023	-.128	-.816	-2.277
VK2	13.000	24.000	-.310	-1.729	-.784	-2.188
VK1	19.000	24.000	-1.269	-7.086	.878	2.451
PK6	2.000	41.000	.535	2.987	-.543	-1.516
PK5	4.000	45.000	.173	.966	-1.048	-2.926
PK4	29.000	43.000	-.272	-1.521	-.606	-1.690
PK3	27.000	41.000	-.136	-.759	-.733	-2.046
PK2	14.000	24.000	-.244	-1.361	-.597	-1.667
PK1	18.000	32.000	-.098	-.549	-.603	-1.683
Multivariate					12.480	1.548

Appendix 80

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
WMCF	21.65	24.389	.180	.572
WMCB	28.28	11.460	.386	.489
WMEF	28.87	19.740	.313	.488
WMEB	31.40	18.091	.581	.314

According to the results of analysis, if the subset of the WMCF was deleted, the reliability of the latent construct would be improved from .545 to .572. Therefore, the subset of the WMCF was deleted from the construct.

Appendix 81 The reliability of the new construct of WM

Reliability Statistics

Cronbach's Alpha	N of Items
.572	3

Appendix 82 The reliability of the latent construct RA

Reliability Statistics

Cronbach's Alpha	N of Items
.585	5

Appendix 83

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
UPRAVENA	33.32	20.853	.233	.591
UPRAVENB	33.28	20.613	.261	.585
UPRAVENC	35.48	14.681	.408	.491
UPRAVEND	37.32	10.262	.556	.376
UPRAVENE	40.39	13.066	.386	.512

The results of analysis show that if the section A of Raven's SPM was deleted from the construct, the reliability of the construct would be improved from .585 to .591. Therefore, the subset was deleted from the construct.

Appendix 84 The reliability of the new latent construct RA

Reliability Statistics

Cronbach's Alpha	N of Items
.591	4

Appendix 85

Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
LC1	<---	LC	1.000				
LC2	<---	LC	.688	.106	6.524	***	par_1
LC3	<---	LC	1.037	.141	7.356	***	par_2
LC4	<---	LC	.715	.120	5.935	***	par_3

Appendix 86

Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
LC1	<---	LC	.964	.131	7.356	***	par_1
LC2	<---	LC	.664	.108	6.147	***	par_2
LC3	<---	LC	1.000				
LC4	<---	LC	.689	.107	6.429	***	par_3

Appendix 87

Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
PK1	<---	PK	1.000				
PK2	<---	PK	.808	.259	3.113	.002	par_1
PK3	<---	PK	1.887	.512	3.689	***	par_2
PK4	<---	PK	2.028	.541	3.751	***	par_3
PK5	<---	PK	10.945	2.696	4.060	***	par_4
PK6	<---	PK	9.616	2.363	4.069	***	par_5

Appendix 88
Regression Weights: (Group number 1 - Default model)

	Estimate	S.E.	C.R.	P	Label
PK1 <--- PK	.091	.023	4.060	***	par_1
PK2 <--- PK	.074	.016	4.752	***	par_2
PK3 <--- PK	.172	.021	8.341	***	par_3
PK4 <--- PK	.185	.020	9.302	***	par_4
PK5 <--- PK	1.000				
PK6 <--- PK	.879	.029	30.457	***	par_5

Appendix 89
CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	12	76.825	9	.000	8.536
Saturated model	21	.000	0		
Independence model	6	651.255	15	.000	43.417

Appendix 90
RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	1.137	.888	.739	.381
Saturated model	.000	1.000		
Independence model	26.508	.489	.285	.349

Appendix 91
Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.882	.803	.894	.822	.893
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Appendix 92
RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.201	.161	.244	.000
Independence model	.478	.447	.509	.000

Appendix 93
Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
PK3	<---	PK	.356	.046	7.669	***	par_1
PK4	<---	PK	.370	.047	7.881	***	par_2
PK6	<---	PK	1.000				
PK2	<---	PK	.110	.028	3.944	***	par_3

Appendix 94
CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	8	3.365	2	.186	1.683
Saturated model	10	.000	0		
Independence model	4	173.989	6	.000	28.998

Appendix 95
RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	.543	.991	.956	.198
Saturated model	.000	1.000		
Independence model	9.500	.652	.421	.391

Appendix 96
Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.981	.942	.992	.976	.992
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Appendix 97
RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.061	.000	.170	.325
Independence model	.388	.339	.439	.000

Appendix 98
Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
VK1	<---	VK	1.000				
VK2	<---	VK	3.700	.543	6.814	***	par_1
VK3	<---	VK	4.560	.667	6.834	***	par_2
VK4	<---	VK	3.865	.583	6.625	***	par_3
VK5	<---	VK	3.318	.518	6.403	***	par_4
VK6	<---	VK	6.378	.941	6.777	***	par_5

Appendix 99
Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
VK1	<---	VK	.157	.023	6.777	***	par_1
VK2	<---	VK	.580	.048	12.153	***	par_2
VK3	<---	VK	.715	.059	12.215	***	par_3
VK4	<---	VK	.606	.051	11.945	***	par_4
VK5	<---	VK	.520	.050	10.483	***	par_5
VK6	<---	VK	1.000				

Appendix100
CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	12	13.172	9	.155	1.464
Saturated model	21	.000	0		
Independence model	6	558.036	15	.000	37.202

Appendix 101
RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	.279	.977	.946	.419
Saturated model	.000	1.000		
Independence model	6.372	.392	.149	.280

Appendix 102
Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.976	.961	.992	.987	.992
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Appendix 103
RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.050	.000	.104	.444
Independence model	.441	.410	.473	.000

Appendix 104
Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
GK1	<---	GK	1.000				
GK2	<---	GK	1.472	.105	14.012	***	par_1
GK3	<---	GK	1.029	.084	12.180	***	par_2
GK4	<---	GK	1.101	.142	7.779	***	par_3
GK5	<---	GK	.449	.111	4.062	***	par_4

Appendix 105
Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
GK1	<---	GK	.680	.048	14.012	***	par_1
GK2	<---	GK	1.000				
GK3	<---	GK	.699	.055	12.825	***	par_2
GK4	<---	GK	.748	.093	8.028	***	par_3
GK5	<---	GK	.305	.075	4.082	***	par_4

Appendix 106
CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	10	69.415	5	.000	13.883
Saturated model	15	.000	0		
Independence model	5	443.811	10	.000	44.381

Appendix 107

RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	2.509	.886	.657	.295
Saturated model	.000	1.000		
Independence model	7.735	.485	.228	.323

Appendix 108

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.844	.687	.853	.703	.852
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Appendix 109

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.263	.210	.320	.000
Independence model	.483	.445	.522	.000

Appendix 110

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	8	1.962	2	.375	.981
Saturated model	10	.000	0		
Independence model	4	359.783	6	.000	59.964

Appendix 111

RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	.295	.995	.974	.199
Saturated model	.000	1.000		
Independence model	8.200	.482	.137	.289

Appendix 112
Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.995	.984	1.000	1.000	1.000
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Appendix 113
RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.000	.000	.144	.527
Independence model	.563	.514	.613	.000

Appendix 114
Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
SPS1	<---	SPS	1.000				
SPS2	<---	SPS	1.091	.111	9.824	***	par_1
SPS3	<---	SPS	.949	.106	8.925	***	par_2
SPS4	<---	SPS	.859	.088	9.772	***	par_3

Appendix 115
Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
SPS1	<---	SPS	.916	.093	9.824	***	par_1
SPS2	<---	SPS	1.000				
SPS3	<---	SPS	.869	.091	9.564	***	par_2
SPS4	<---	SPS	.787	.078	10.094	***	par_3

Appendix 116
CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	8	2.175	2	.337	1.088
Saturated model	10	.000	0		
Independence model	4	293.830	6	.000	48.972

Appendix 117
RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	16243.592	.994	.971	.199
Saturated model	.000	1.000		
Independence model	502849.488	.503	.172	.302

Appendix 118
Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.993	.978	.999	.998	.999
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Appendix 119
RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.022	.000	.149	.490
Independence model	.508	.459	.558	.000

Appendix 120
Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
WM1	<---	WM	1.000				
WM2	<---	WM	11.751	9.500	1.237	.216	par_1
WM3	<---	WM	5.849	4.774	1.225	.220	par_2
WM4	<---	WM	12.478	11.903	1.048	.295	par_3

Appendix 121
Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
WM1	<---	WM	.080	.076	1.048	.295	par_1
WM2	<---	WM	.942	.289	3.259	.001	par_2
WM3	<---	WM	.469	.152	3.091	.002	par_3
WM4	<---	WM	1.000				

Appendix 122
CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	8	6.168	2	.046	3.084
Saturated model	10	.000	0		
Independence model	4	100.944	6	.000	16.824

Appendix 123
RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	.157	.985	.923	.197
Saturated model	.000	1.000		
Independence model	.944	.792	.654	.475

Appendix 124
Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.939	.817	.958	.868	.956
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Appendix 125
RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.106	.013	.206	.119
Independence model	.292	.243	.343	.000

Appendix 126
Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
RA2	<---	RA	1.000				
RA3	<---	RA	5.300	1.932	2.743	.006	par_1
RA4	<---	RA	13.233	5.672	2.333	.020	par_2
RA5	<---	RA	6.469	2.416	2.677	.007	par_3

Appendix 127

Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
RA2	<---	RA	.076	.032	2.333	.020	par_1
RA3	<---	RA	.401	.107	3.759	***	par_2
RA4	<---	RA	1.000				
RA5	<---	RA	.489	.120	4.084	***	par_3

Appendix 128

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	8	5.333	2	.069	2.667
Saturated model	10	.000	0		
Independence model	4	92.027	6	.000	15.338

Appendix 129

RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	.060	.986	.929	.197
Saturated model	.000	1.000		
Independence model	.792	.793	.655	.476

Appendix 130

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.942	.826	.963	.884	.961
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Appendix 131

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.095	.000	.196	.161
Independence model	.278	.229	.329	.000

Appendix 132

Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
EU1	<---	EU	1.000				
EU2	<---	EU	1.177	.031	37.932	***	par_1
EU 3	<---	EU	1.071	.030	35.467	***	par_2
EU 4	<---	EU	1.168	.037	31.583	***	par_3

Appendix 133

Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
EU 1	<---	EU	.850	.022	37.932	***	par_1
EU 2	<---	EU	1.000				
EU 3	<---	EU	.910	.020	46.591	***	par_2
EU 4	<---	EU	.993	.026	38.839	***	par_3

Appendix 134

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	8	1.436	2	.488	.718
Saturated model	10	.000	0		
Independence model	4	1337.227	6	.000	222.871

Appendix 135

RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	.905	.996	.981	.199
Saturated model	.000	1.000		
Independence model	336.184	.275	-.209	.165

Appendix 136

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.999	.997	1.000	1.001	1.000
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Appendix 137
RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.000	.000	.132	.628
Independence model	1.092	1.043	1.142	.000

Appendix 138
CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	27	100.971	51	.000	1.980
Saturated model	78	.000	0		
Independence model	12	1235.815	66	.000	18.724

Appendix 139
RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	.698	.922	.881	.603
Saturated model	.000	1.000		
Independence model	6.403	.298	.171	.252

Appendix 140
Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.918	.894	.958	.945	.957
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Appendix 141
RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.073	.052	.093	.040
Independence model	.309	.294	.324	.000

Appendix 142
CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	61	380.042	239	.000	1.590
Saturated model	300	.000	0		
Independence model	24	3385.681	276	.000	12.267

Appendix 143
RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	3109.157	.867	.833	.691
Saturated model	.000	1.000		
Independence model	91810.061	.263	.199	.242

Appendix 144
Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.888	.870	.955	.948	.955
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Appendix 145
RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.056	.045	.067	.162
Independence model	.246	.239	.254	.000

Appendix 146

Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
LC	<---	EU	.016	.005	2.930	.003	par_30
LC	<---	RA	.062	.073	.850	.395	par_31
LC	<---	WM	.084	.056	1.516	.130	par_32
LC	<---	SPS	.000	.000	1.486	.137	par_33
LC	<---	LK	.141	.040	3.558	***	par_34
PK	<---	LK	.587	.081	7.222	***	par_35
VK	<---	LK	1.000				
GK	<---	LK	.875	.104	8.394	***	par_36
PK3	<---	PK	.903	.120	7.506	***	par_1
PK4	<---	PK	1.000				
VK1	<---	VK	.158	.023	6.936	***	par_2
VK2	<---	VK	.575	.046	12.467	***	par_3
VK3	<---	VK	.700	.057	12.332	***	par_4
VK4	<---	VK	.603	.050	12.186	***	par_5
VK5	<---	VK	.524	.048	10.806	***	par_6
VK6	<---	VK	1.000				
GK1	<---	GK	.688	.048	14.339	***	par_7
GK2	<---	GK	1.000				
GK3	<---	GK	.728	.054	13.361	***	par_8
GK4	<---	GK	.756	.094	8.061	***	par_9
SPS1	<---	SPS	.927	.094	9.857	***	par_10
SPS2	<---	SPS	1.000				
SPS3	<---	SPS	.874	.091	9.607	***	par_11
SPS4	<---	SPS	.776	.077	10.059	***	par_12
WM2	<---	WM	1.000				
WM4	<---	WM	.866	.197	4.389	***	par_13
EU1	<---	EU	.850	.022	37.830	***	par_14
EU2	<---	EU	1.000				
EU3	<---	EU	.911	.020	46.666	***	par_15
EU4	<---	EU	.993	.026	38.769	***	par_16
LC1	<---	LC	.996	.110	9.038	***	par_17
LC2	<---	LC	.617	.091	6.755	***	par_18
LC3	<---	LC	1.000				
LC4	<---	LC	.664	.096	6.890	***	par_19
RA5	<---	RA	1.000				
RA4	<---	RA	.954	.470	2.031	.042	par_37

Appendix 147

Standardized Regression Weights: (Group number 1 - Default model)

			Estimate
LC	<---	EU	.308
LC	<---	RA	.077
LC	<---	WM	.125
LC	<---	SPS	.114
LC	<---	LK	.464
PK	<---	LK	.765
VK	<---	LK	.854
GK	<---	LK	.797
PK3	<---	PK	.713
PK4	<---	PK	.792
VK1	<---	VK	.503
VK2	<---	VK	.814
VK3	<---	VK	.800
VK4	<---	VK	.785
VK5	<---	VK	.720
VK6	<---	VK	.822
GK1	<---	GK	.829
GK2	<---	GK	.882
GK3	<---	GK	.813
GK4	<---	GK	.560
SPS1	<---	SPS	.762
SPS2	<---	SPS	.805
SPS3	<---	SPS	.710
SPS4	<---	SPS	.751
WM2	<---	WM	.560
WM4	<---	WM	.922
EU1	<---	EU	.956
EU2	<---	EU	.984
EU3	<---	EU	.976
EU4	<---	EU	.959
LC1	<---	LC	.738
LC2	<---	LC	.543
LC3	<---	LC	.767
LC4	<---	LC	.541
RA5	<---	RA	.709
RA4	<---	RA	.624

Appendix 148

Covariances: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
SPS	<-->	WM	201.440	139.366	1.445	.148	par_20
SPS	<-->	RA	198.700	143.399	1.386	.166	par_21
SPS	<-->	EU	2783.633	1620.389	1.718	.086	par_22
SPS	<-->	LK	-342.930	305.841	-1.121	.262	par_23
WM	<-->	RA	.239	.248	.965	.334	par_24
WM	<-->	EU	16.709	4.908	3.404	***	par_25
WM	<-->	LK	2.797	.841	3.328	***	par_26
EU	<-->	RA	6.018	3.306	1.820	.069	par_27
RA	<-->	LK	.934	.559	1.670	.095	par_28
EU	<-->	LK	57.376	9.261	6.195	***	par_29

Appendix 149

Correlations: (Group number 1 - Default model)

			Estimate
SPS	<-->	WM	.133
SPS	<-->	RA	.158
SPS	<-->	EU	.139
SPS	<-->	LK	-.102
WM	<-->	RA	.102
WM	<-->	EU	.448
WM	<-->	LK	.445
EU	<-->	RA	.194
RA	<-->	LK	.178
EU	<-->	LK	.692

Appendix 150

Squared Multiple Correlations: (Group number 1 - Default model)

	Estimate
LC	.657
GK	.635
VK	.729
PK	.585
LC4	.293
LC3	.588
LC2	.295
LC1	.544
EU4	.919
EU3	.953
EU2	.969
EU1	.914
RA5	.503
RA4	.389
WM4	.850
WM2	.313
SPS4	.564
SPS3	.504
SPS2	.649
SPS1	.580
GK4	.314
GK3	.662
GK2	.778
GK1	.688
VK6	.675
VK5	.519
VK4	.616
VK3	.640
VK2	.663
VK1	.253
PK4	.628
PK3	.508

Appendix 151

Correlations: (Group number 1 - Default model)

			Estimate
EU	<-->	LC	.716
LC	<-->	LK	.734
WM	<-->	LC	.492
SPS	<-->	LC	.139
LC	<-->	RA	.250
SPS	<-->	WM	.133
SPS	<-->	RA	.158
SPS	<-->	EU	.139
SPS	<-->	LK	-.102
WM	<-->	RA	.102
WM	<-->	EU	.448
WM	<-->	LK	.445
EU	<-->	RA	.194
LK	<-->	RA	.178
EU	<-->	LK	.692

Appendix 152

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	74	490.921	332	.000	1.479
Saturated model	406	.000	0		
Independence model	28	3778.755	378	.000	9.997

Appendix 153

RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	2702.166	.856	.824	.700
Saturated model	.000	1.000		
Independence model	78920.153	.245	.189	.228

Appendix 154

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.870	.852	.954	.947	.953
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Appendix 155
RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.051	.041	.060	.441
Independence model	.220	.214	.226	.000

Appendix 156
Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
LC	<---	EU	.017	.005	3.203	.001	par_11
LC	<---	RA	.035	.037	.934	.350	par_12
LC	<---	WM	.026	.035	.740	.459	par_13
LC	<---	SPS	.000	.000	1.399	.162	par_14
LC	<---	LK	.181	.045	4.023	***	par_15
PK	<---	LK	.839	.101	8.315	***	par_16
VK	<---	LK	.820	.086	9.567	***	par_17
GK	<---	LK	1.000				
PK_av	<---	PK	.851				
VK_av	<---	VK	.939				
GK_av	<---	GK	.927				
SPS_av	<---	SPS	.911				
WM_av	<---	WM	.851				
EU_av	<---	EU	.992				
LC_av	<---	LC	.864				
RA_av	<---	RA	.785				

Appendix 157
Standardized Residual Covariances (Group number 1 - Default model)

	LC_av	EU_av	RA_av	WM_av	SPS_av	GK_av	VK_av	PK_av
LC_av	.000							
EU_av	.000	.000						
RA_av	.000	.000	.000					
WM_av	.000	.000	.000	.000				
SPS_av	.000	.000	.000	.000	.000			
GK_av	.162	-1.041	1.517	.224	.127	.000		
VK_av	-.265	.790	-1.058	-.219	.017	.068	.000	
PK_av	.333	-.091	-.115	.122	-.238	.660	-.589	.000

Appendix 158
CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	28	7.034	8	.533	.879
Saturated model	36	.000	0		
Independence model	8	436.208	28	.000	15.579

Appendix 159
RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	31.089	.991	.958	.220
Saturated model	.000	1.000		
Independence model	455.329	.546	.417	.425

Appendix 160
Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.984	.944	1.002	1.008	1.000
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Appendix 161
RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.000	.000	.079	.794
Independence model	.280	.257	.303	.000

Appendix 162

Regression Weights: (Group number 1 - Default model)

			M.I.	Par Change
GK	<---	RA	4.023	.424
LC2	<---	SPS4	4.553	.000
EU4	<---	VK5	4.051	.329
EU2	<---	SPS2	5.263	-.001
EU1	<---	SPS	5.793	.001
EU1	<---	SPS2	9.446	.001
RA5	<---	VK6	7.092	-.064
RA4	<---	GK4	4.826	.056
WM4	<---	VK2	4.204	-.061
SPS4	<---	RA5	5.710	-60.771
SPS4	<---	VK6	4.884	20.809
SPS2	<---	LK	5.917	-40.448
SPS2	<---	WM	4.908	-80.846
SPS2	<---	PK	10.035	-71.238
SPS2	<---	EU2	4.361	-5.296
SPS2	<---	WM4	4.019	-72.477
SPS2	<---	WM2	4.340	-39.609
SPS2	<---	GK3	5.412	-36.182
SPS2	<---	PK4	4.208	-32.384
SPS2	<---	PK3	16.630	-64.179
SPS1	<---	RA	12.919	188.786
SPS1	<---	RA5	12.816	106.107
SPS1	<---	RA4	5.549	64.398
GK4	<---	RA	4.025	.621
GK4	<---	RA4	6.492	.410
GK3	<---	VK1	6.136	.315
GK3	<---	PK3	8.954	.144
GK2	<---	LC2	5.698	-.366
VK6	<---	RA	7.040	-.588
VK6	<---	RA5	12.323	-.439
VK5	<---	RA	4.515	.325
VK5	<---	EU4	5.415	.017
VK5	<---	RA4	5.950	.194
VK5	<---	GK4	7.535	-.084
VK4	<---	SPS1	4.304	.000
VK3	<---	LC4	5.252	-.301
VK1	<---	GK	4.645	.049
VK1	<---	GK3	8.356	.070
VK1	<---	GK2	5.808	.046

			M.I.	Par Change
PK3	<---	SPS2	4.750	.000
PK3	<---	GK3	5.854	.139
PK3	<---	VK3	4.603	-.118

Appendix 163
CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	75	472.957	331	.000	1.429
Saturated model	406	.000	0		
Independence model	28	3778.755	378	.000	9.997

Appendix 164
RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	2723.155	.860	.829	.701
Saturated model	.000	1.000		
Independence model	78920.153	.245	.189	.228

Appendix 165
Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.875	.857	.959	.952	.958
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Appendix 166
RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.048	.038	.058	.623
Independence model	.220	.214	.226	.000

Appendix 167

Regression Weights: (Group number 1 - Default model)

			M.I.	Par Change
GK	<---	RA	4.129	.424
LC2	<---	SPS4	4.599	.000
EU4	<---	VK5	4.044	.329
EU2	<---	SPS2	5.201	-.001
EU1	<---	SPS	6.246	.001
EU1	<---	SPS2	9.421	.001
RA5	<---	VK6	7.098	-.064
RA4	<---	GK4	4.942	.056
WM4	<---	VK2	4.128	-.060
SPS4	<---	LC4	4.701	-77.186
SPS4	<---	LC2	4.669	-83.099
SPS4	<---	RA5	6.309	-63.446
SPS1	<---	RA	11.791	176.479
SPS1	<---	RA5	11.936	101.800
SPS1	<---	RA4	5.442	63.398
GK4	<---	LC1	4.007	.448
GK4	<---	RA4	6.505	.411
GK3	<---	PK3	9.183	.146
GK3	<---	VK1	6.105	.314
GK2	<---	LC2	5.640	-.364
VK6	<---	RA	7.386	-.593
VK6	<---	RA5	12.374	-.440
VK5	<---	RA	4.413	.316
VK5	<---	EU4	5.370	.017
VK5	<---	RA4	5.938	.194
VK5	<---	GK4	7.545	-.084
VK4	<---	SPS1	4.481	.000
VK3	<---	LC4	5.334	-.303
VK1	<---	GK	4.625	.049
VK1	<---	GK3	8.334	.069
VK1	<---	GK2	5.776	.046

Appendix 168

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	76	459.789	330	.000	1.393
Saturated model	406	.000	0		
Independence model	28	3778.755	378	.000	9.997

Appendix 169
RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	2716.484	.864	.833	.702
Saturated model	.000	1.000		
Independence model	78920.153	.245	.189	.228

Appendix 170
Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.878	.861	.962	.956	.962
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Appendix 171
RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.046	.035	.056	.742
Independence model	.220	.214	.226	.000

Appendix 172
Assessment of normality (Group number 1)

Variable	min	max	skew	c.r.	kurtosis	c.r.
LC	-1.840	1.875	.067	.374	-.575	-1.604
EU4	.000	100.000	1.532	8.555	1.012	2.824
EU3	.000	98.000	1.855	10.358	2.508	7.001
EU2	.000	92.000	1.770	9.882	2.114	5.902
EU1	.000	90.000	2.094	11.692	3.721	10.387
RA5	1.000	8.000	.011	.060	-.844	-2.355
RA4	4.000	12.000	.069	.387	-.732	-2.045
WM4	2.000	8.000	-.189	-1.056	-.749	-2.089
WM2	3.000	14.000	.082	.455	-.876	-2.446
SPS4	183.500	4767.000	.826	4.610	.245	.685
SPS3	.000	5096.000	.798	4.453	-.001	-.002
SPS2	232.667	5342.000	1.117	6.237	.754	2.105
SPS1	.000	5859.667	.836	4.665	.399	1.114
GK4	19.000	41.000	-.204	-1.141	-.626	-1.747
GK3	11.000	24.000	-.542	-3.023	-.664	-1.853

Variable	min	max	skew	c.r.	kurtosis	c.r.
GK2	11.000	28.000	-.437	-2.438	-.811	-2.263
GK1	15.000	28.000	-.326	-1.818	-.598	-1.669
VK6	7.000	28.000	.081	.452	-.860	-2.401
VK5	8.000	20.000	-.102	-.570	-.688	-1.919
VK4	6.000	19.000	.093	.518	-.616	-1.719
VK3	7.000	22.000	-.023	-.128	-.816	-2.277
VK2	13.000	24.000	-.310	-1.729	-.784	-2.188
VK1	19.000	24.000	-1.269	-7.086	.878	2.451
PK4	29.000	43.000	-.272	-1.521	-.606	-1.690
PK3	27.000	41.000	-.136	-.759	-.733	-2.046
Multivariate					23.016	4.283

Appendix 173

Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
PK	<---	LK	.543	.075	7.261	***	par_27
VK	<---	LK	1.000				
GK	<---	LK	.828	.095	8.688	***	par_28
PK3	<---	PK	.879	.122	7.209	***	par_1
PK4	<---	PK	1.000				
VK1	<---	VK	.159	.023	6.948	***	par_2
VK2	<---	VK	.582	.046	12.569	***	par_3
VK3	<---	VK	.708	.057	12.404	***	par_4
VK4	<---	VK	.606	.050	12.142	***	par_5
VK5	<---	VK	.522	.049	10.679	***	par_6
VK6	<---	VK	1.000				
GK1	<---	GK	.686	.047	14.464	***	par_7
GK2	<---	GK	1.000				
GK3	<---	GK	.715	.054	13.335	***	par_8
GK4	<---	GK	.751	.093	8.079	***	par_9
SPS1	<---	SPS	.935	.095	9.897	***	par_10
SPS2	<---	SPS	1.000				
SPS3	<---	SPS	.881	.092	9.618	***	par_11
SPS4	<---	SPS	.780	.077	10.082	***	par_12
WM2	<---	WM	1.000				
WM4	<---	WM	.805	.181	4.442	***	par_13
EU1	<---	EU	.850	.022	37.901	***	par_14
EU2	<---	EU	1.000				
EU3	<---	EU	.911	.020	46.501	***	par_15
EU4	<---	EU	.994	.026	38.755	***	par_16
RA5	<---	RA	.947	.580	1.634	.102	par_29

			Estimate	S.E.	C.R.	P	Label
RA4	<---	RA	1.000				
LC	<---	EU	-.022	.005	-4.388	***	par_30
LC	<---	RA	-.067	.066	-1.012	.312	par_31
LC	<---	WM	-.001	.050	-.011	.991	par_32
LC	<---	LK	.205	.036	5.666	***	par_33
LC	<---	SPS	.000	.000	3.394	***	par_34

Appendix 174

Standardized Regression Weights: (Group number 1 - Default model)

			Estimate
PK	<---	LK	.724
VK	<---	LK	.890
GK	<---	LK	.777
PK3	<---	PK	.704
PK4	<---	PK	.803
VK1	<---	VK	.503
VK2	<---	VK	.820
VK3	<---	VK	.804
VK4	<---	VK	.784
VK5	<---	VK	.714
VK6	<---	VK	.818
GK1	<---	GK	.833
GK2	<---	GK	.887
GK3	<---	GK	.804
GK4	<---	GK	.560
SPS1	<---	SPS	.765
SPS2	<---	SPS	.801
SPS3	<---	SPS	.712
SPS4	<---	SPS	.751
WM2	<---	WM	.581
WM4	<---	WM	.889
EU1	<---	EU	.956
EU2	<---	EU	.984
EU3	<---	EU	.976
EU4	<---	EU	.959
RA5	<---	RA	.674
RA4	<---	RA	.656
LC	<---	EU	-.516
LC	<---	RA	-.099
LC	<---	WM	-.001
LC	<---	LK	.838

			Estimate
LC	<---	SPS	.291

Appendix 175

Covariances: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
SPS	<-->	WM	217.626	147.167	1.479	.139	par_17
SPS	<-->	RA	195.361	132.335	1.476	.140	par_18
SPS	<-->	EU	2792.403	1612.376	1.732	.083	par_19
SPS	<-->	LK	-339.281	312.396	-1.086	.277	par_20
WM	<-->	RA	.271	.286	.947	.343	par_21
WM	<-->	EU	17.909	5.002	3.580	***	par_22
WM	<-->	LK	2.990	.873	3.426	***	par_23
EU	<-->	RA	5.981	3.207	1.865	.062	par_24
RA	<-->	LK	.904	.705	1.281	.200	par_25
EU	<-->	LK	59.795	9.255	6.461	***	par_26

Appendix 176

Correlations: (Group number 1 - Default model)

			Estimate
SPS	<-->	WM	.139
SPS	<-->	RA	.155
SPS	<-->	EU	.141
SPS	<-->	LK	-.098
WM	<-->	RA	.111
WM	<-->	EU	.463
WM	<-->	LK	.442
EU	<-->	RA	.192
RA	<-->	LK	.166
EU	<-->	LK	.696

Appendix 177

Squared Multiple Correlations: (Group number 1 - Default model)

	Estimate
GK	.604
VK	.791
PK	.524
LC	.354
EU4	.919
EU3	.953
EU2	.968
EU1	.915
RA5	.454
RA4	.431
WM4	.790
WM2	.337
SPS4	.564
SPS3	.507
SPS2	.642
SPS1	.584
GK4	.313
GK3	.647
GK2	.787
GK1	.693
VK6	.668
VK5	.510
VK4	.614
VK3	.647
VK2	.672
VK1	.253
PK4	.645
PK3	.495

Appendix 178
Correlations: (Group number 1 - Default model)

			Estimate
SPS	<-->	WM	.139
SPS	<-->	RA	.155
SPS	<-->	EU	.141
SPS	<-->	LK	-.098
WM	<-->	RA	.111
WM	<-->	EU	.463
WM	<-->	LK	.442
EU	<-->	RA	.192
RA	<-->	LK	.166
EU	<-->	LK	.696
SPS	<-->	LC	.121
WM	<-->	LC	.160
RA	<-->	LC	-.014
EU	<-->	LC	.088
LK	<-->	LC	.434

Appendix 179
CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	67	417.006	258	.000	1.616
Saturated model	325	.000	0		
Independence model	25	3471.448	300	.000	11.571

Appendix 180
RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	3101.644	.861	.825	.684
Saturated model	.000	1.000		
Independence model	88208.254	.265	.204	.244

Appendix 181
Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.880	.860	.951	.942	.950
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Appendix 182

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.058	.047	.068	.110
Independence model	.238	.231	.246	.000

Appendix 183

Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
PK	<---	LK	.818	.103	7.979	***	par_11
VK	<---	LK	.876	.087	10.120	***	par_12
GK	<---	LK	1.000				
PK_av	<---	PK	.852				
VK_av	<---	VK	.939				
GK_av	<---	GK	.927				
SPS_av	<---	SPS	.919				
WM_av	<---	WM	.844				
EU_av	<---	EU	.992				
RA_av	<---	RA	.783				
LC	<---	EU	-.022	.005	-4.560	***	par_13
LC	<---	RA	-.040	.034	-1.188	.235	par_14
LC	<---	WM	.007	.031	.222	.824	par_15
LC	<---	LK	.268	.045	5.930	***	par_16
LC	<---	SPS	.000	.000	3.612	***	par_17

Appendix 184

Standardized Residual Covariances (Group number 1 - Default model)

	LC	EU_av	RA_av	WM_av	SPS_av	GK_av	VK_av	PK_av
LC	.000							
EU_av	.000	.000						
RA_av	.000	.000	.000					
WM_av	.000	.000	.000	.000				
SPS_av	.000	.000	.000	.000	.000			
GK_av	.497	-.943	1.717	.371	.096	.000		
VK_av	-.066	.451	-.950	-.301	.040	-.035	.000	
PK_av	-.616	.131	.083	.318	-.280	1.115	-.524	.000

Appendix 185
CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	28	11.840	8	.158	1.480
Saturated model	36	.000	0		
Independence model	8	375.635	28	.000	13.416

Appendix 186
RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	13.406	.985	.932	.219
Saturated model	.000	1.000		
Independence model	455.300	.617	.508	.480

Appendix 187
Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.968	.890	.990	.961	.989
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Appendix 188
RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.051	.000	.108	.432
Independence model	.258	.235	.282	.000

Appendix 189
Regression Weights: (Group number 1 - Default model)

			M.I.	Par Change
GK	<---	RA	4.371	.448
LC	<---	GK3	4.394	-.035
EU2	<---	SPS2	5.241	-.001
EU1	<---	SPS	5.787	.001
EU1	<---	SPS2	9.504	.001
RA5	<---	SPS1	4.103	.000
RA5	<---	VK6	6.143	-.060
RA4	<---	GK4	5.146	.057

			M.I.	Par Change
WM4	<---	GK3	4.491	.054
WM4	<---	VK2	4.076	-.061
SPS4	<---	RA5	5.861	-61.513
SPS4	<---	VK6	4.921	20.868
SPS2	<---	LK	5.976	-38.884
SPS2	<---	WM	5.265	-83.145
SPS2	<---	PK	9.947	-70.311
SPS2	<---	EU2	4.332	-5.294
SPS2	<---	WM2	4.404	-40.015
SPS2	<---	GK3	4.958	-34.735
SPS2	<---	PK4	4.103	-32.075
SPS2	<---	PK3	16.301	-63.729
SPS1	<---	RA	11.868	181.017
SPS1	<---	RA5	12.705	105.274
SPS1	<---	RA4	5.693	64.999
GK4	<---	RA	4.069	.627
GK4	<---	LC	4.452	.767
GK4	<---	RA4	6.675	.416
GK3	<---	LC	7.583	-.513
GK3	<---	VK1	6.146	.319
GK3	<---	PK3	10.472	.157
VK6	<---	RA	6.603	-.574
VK6	<---	RA5	12.683	-.447
VK5	<---	RA	4.975	.344
VK5	<---	LC	4.911	-.400
VK5	<---	EU4	5.398	.017
VK5	<---	RA4	6.005	.196
VK5	<---	GK4	7.303	-.084
VK4	<---	SPS1	4.334	.000
VK2	<---	WM4	4.170	-.185
VK1	<---	GK	4.596	.048
VK1	<---	GK3	8.534	.070
VK1	<---	GK2	5.608	.045
PK3	<---	SPS2	4.876	.000
PK3	<---	GK3	8.064	.165
PK3	<---	VK3	4.142	-.113

Appendix 190
CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	68	399.436	257	.000	1.554
Saturated model	325	.000	0		
Independence model	25	3471.448	300	.000	11.571

Appendix 191
RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	3129.725	.866	.831	.685
Saturated model	.000	1.000		
Independence model	88208.254	.265	.204	.244

Appendix 192
Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.885	.866	.956	.948	.955
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Appendix 193
RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.055	.044	.065	.229
Independence model	.238	.231	.246	.000

Appendix 194

Regression Weights: (Group number 1 - Default model)

			M.I.	Par Change
GK	<---	RA	4.596	.473
GK	<---	PK3	5.938	.160
LC	<---	GK3	4.393	-.035
EU2	<---	SPS2	5.200	-.001
EU1	<---	SPS	6.221	.001
EU1	<---	SPS2	9.483	.001
RA5	<---	VK6	6.272	-.061
RA4	<---	GK4	5.322	.058
WM4	<---	GK3	4.592	.054
SPS4	<---	RA5	6.482	-64.237
SPS1	<---	RA	11.097	178.642
SPS1	<---	RA5	11.850	101.074
SPS1	<---	RA4	5.609	64.141
GK4	<---	LC	4.466	.768
GK4	<---	RA4	6.700	.417
GK3	<---	PK3	10.713	.159
GK3	<---	LC	7.539	-.512
GK3	<---	VK1	6.108	.318
VK6	<---	RA	7.208	-.615
VK6	<---	RA5	12.710	-.447
VK5	<---	RA	4.742	.345
VK5	<---	LC	4.937	-.402
VK5	<---	EU4	5.341	.017
VK5	<---	RA4	6.028	.197
VK5	<---	GK4	7.291	-.083
VK4	<---	SPS1	4.536	.000
VK2	<---	WM4	4.201	-.186
VK1	<---	GK	4.586	.048
VK1	<---	GK3	8.518	.070
VK1	<---	GK2	5.574	.045

Appendix 195

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	69	385.803	256	.000	1.507
Saturated model	325	.000	0		
Independence model	25	3471.448	300	.000	11.571

Appendix 196
RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	3125.379	.871	.836	.686
Saturated model	.000	1.000		
Independence model	88208.254	.265	.204	.244

Appendix 197
Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.889	.870	.960	.952	.959
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Appendix 198
RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.052	.041	.063	.357
Independence model	.238	.231	.246	.000

Appendix 199 The ethics forms

University of Reading
Institute of Education
Ethical Approval Form A (version May 2015)

University of Reading

Tick one:

Staff project: _____ PhD _X_ EdD ____

Name of applicant (s):Yun Wang

Title of project: Individual differences in listening comprehension among adult Chinese learners of English

Name of supervisor (for student projects): Jeanine Treffers-Daller, Suzanne Graham

Please complete the form below including relevant sections overleaf.

	YES	NO	
Have you prepared an Information Sheet for participants and/or their parents/carers that:			
a) explains the purpose(s) of the project	x		
b) explains how they have been selected as potential participants	x		
c) gives a full, fair and clear account of what will be asked of them and how the information that they provide will be used	x		
d) makes clear that participation in the project is voluntary	x		
e) explains the arrangements to allow participants to withdraw at any stage if they wish	x		
f) explains the arrangements to ensure the confidentiality of any material collected during the project, including secure arrangements for its storage, retention and disposal	x		
g) explains the arrangements for publishing the research results and, if confidentiality might be affected, for obtaining written consent for this	x		
h) explains the arrangements for providing participants with the research results if they wish to have them	x		
i) gives the name and designation of the member of staff with responsibility for the project together with contact details, including email . If any of the project investigators are students at the IoE, then this information must be included and their name provided	x		
k) explains, where applicable, the arrangements for expenses and other payments to be made to the participants	x		
j) includes a standard statement indicating the process of ethical review at the University undergone by the project, as follows: 'This project has been reviewed following the procedures of the University Research Ethics Committee and has been given a favourable ethical opinion for conduct'.	x		
k)includes a standard statement regarding insurance: "The University has the appropriate insurances in place. Full details are available on request".	x		
Please answer the following questions			
1) Will you provide participants involved in your research with all the information necessary to ensure that they are fully informed and not in any way deceived or misled as to the purpose(s) and nature of the research? (Please use the subheadings used in the example information sheets on blackboard to ensure this).	x		
2) Will you seek written or other formal consent from all participants, if they are able to provide it, in addition to (1)?	x		
3) Is there any risk that participants may experience physical or psychological distress in taking part in your research?		x	
4) Have you taken the online training modules in data protection and information security (which can be found here: http://www.reading.ac.uk/internal/imps/Staffpages/imps-training.aspx)?	x		
5) Have you read the Health and Safety booklet (available on Blackboard) and completed a Risk Assessment Form to be included with this ethics application?	x		
6) Does your research comply with the University's Code of Good Practice in Research?	x		
	YES	NO	N.A.
7) If your research is taking place in a school, have you prepared an information sheet and consent form to gain the permission in writing of the head teacher or other relevant supervisory professional?	x		

8) Has the data collector obtained satisfactory DBS clearance?			x
9) If your research involves working with children under the age of 16 (or those whose special educational needs mean they are unable to give informed consent), have you prepared an information sheet and consent form for parents/carers to seek permission in writing, or to give parents/carers the opportunity to decline consent?			x
10) If your research involves processing sensitive personal data[1], or if it involves audio/video recordings, have you obtained the explicit consent of participants/parents?	x		
11) If you are using a data processor to subcontract any part of your research, have you got a written contract with that contractor which (a) specifies that the contractor is required to act only on your instructions, and (b) provides for appropriate technical and organisational security measures to protect the data?	x		
12a) Does your research involve data collection outside the UK?	x		
12b) If the answer to question 12a is "yes", does your research comply with the legal and ethical requirements for doing research in that country?	x		
13a) Does your research involve collecting data in a language other than English?	x		
13b) If the answer to question 13a is "yes", please confirm that information sheets, consent forms, and research instruments, where appropriate, have been directly translated from the English versions submitted with this application.	x		
14a. Does the proposed research involve children under the age of 5?		x	
14b. If the answer to question 14a is "yes": My Head of School (or authorised Head of Department) has given details of the proposed research to the University's insurance officer, and the research will not proceed until I have confirmation that insurance cover is in place.			x
If you have answered YES to Question 3, please complete Section B below			x

Please complete **either** Section A **or** Section B and provide the details required in support of your application. Sign the form (Section C) then submit it with all relevant attachments (e.g. information sheets, consent forms, tests, questionnaires, interview schedules) to the Institute's Ethics Committee for consideration. Any missing information will result in the form being returned to you.

A: My research goes beyond the 'accepted custom and practice of teaching' but I consider that this project has no significant ethical implications. (Please tick the box.)	x
Please state the total number of participants that will be involved in the project and give a breakdown of how many there are in each category e.g. teachers, parents, pupils etc. The total number of participants will be 160 in the main study: 120 will be university students in China and 40 will be university students in the UK. The participants will be recruited on a voluntary basis.	
Give a brief description of the aims and the methods (participants, instruments and procedures) of the project in up to 200 words noting: 1. title of project 2. purpose of project and its academic rationale 3. brief description of methods and measurements 4. participants: recruitment methods, number, age, gender, exclusion/inclusion criteria 5. consent and participant information arrangements, debriefing (attach forms where necessary) 6. a clear and concise statement of the ethical considerations raised by the project and how you intend to deal with then. 7. estimated start date and duration of project Title of the PhD project :Individual differences in listening comprehension among adult Chinese learners of English This PhD research project is designed to explore individual differences in listening comprehension among adult Chinese learners of English and the extent to which the individual differences can explain the variance in L2 listening comprehension performance. The rationale for conducting the study is that individual differences in listening comprehension have not yet been studied extensively. It is important to do a listening comprehension study because listening is a source of frustration to learners and it is an area in which it seems difficult to make progress (Graham, 2011). Studies on the roles of linguistic knowledge components as well as speed or fluency factors have been undertaken quite often for reading, writing, and speaking, however, a limited number of comprehensive studies have been done on the roles of individual differences in listening comprehension. Although previous studies (Mecartty, 2000; Staehr, 2008, 2009) indicated that linguistic knowledge applied could explain some variance in listening comprehension, it is	

[1] Sensitive personal data consists of information relating to the racial or ethnic origin of a data subject, their political opinions, religious beliefs, trade union membership, sexual life, physical or mental health or condition, or criminal offences or record.

not clear whether individual differences in sentence processing efficiency with spoken input, working memory capacities and reasoning abilities will explain the variance in the success of listening comprehension. In this PhD research project, I plan to adopt a quantitative approach to further our understanding of L2/FL listening comprehension through assessing the roles of linguistic knowledge, sentence processing speed, working memory capacity, reasoning abilities and personal background of L2/FL learners.

A range of online and offline instruments will be used to measure learners' competence in above mentioned five factors. The data will be analyzed using structural equation modelling. All planned tasks will be completed outside of normal class time under supervision of the researcher. The tasks will take four hours in total. I will give written feedback to each participating student on how to improve his/her listening comprehension. Participant information sheets will be distributed to participating students and consent forms will be collected before any experiments are conducted. Participants' performance in the experiments will not be disclosed to any institute and only the people working on the project will know their answers. All data collected will be used only for the research purpose. The main study will start from March 2017 and end in November 2017. I plan to complete the PhD research project in December 2018.

B: I consider that this project may have ethical implications that should be brought before the Institute's Ethics Committee.	

Please state the total number of participants that will be involved in the project and give a breakdown of how many there are in each category e.g. teachers, parents, pupils etc.

Give a brief description of the aims and the methods (participants, instruments and procedures) of the project in up to 200 words.

1. title of project
2. purpose of project and its academic rationale
3. brief description of methods and measurements
4. participants: recruitment methods, number, age, gender, exclusion/inclusion criteria
5. consent and participant information arrangements, debriefing (attach forms where necessary)
6. a clear and concise statement of the ethical considerations raised by the project and how you intend to deal with then.
7. estimated start date and duration of project

C: SIGNATURE OF APPLICANT:

Note: a signature is required. Typed names are not acceptable.

I have declared all relevant information regarding my proposed project and confirm that ethical good practice will be followed within the project.

Signed: ….. [redacted] Print Name…王云. Date…09 September, 2016.

STATEMENT OF ETHICAL APPROVAL FOR PROPOSALS SUBMITTED TO THE INSTITUTE ETHICS COMMITTEE

This project has been considered using agreed Institute procedures and is now approved.

Signed: [redacted]. Print Name……Xiao Lan Curdt-Christiansen……Date: 9 September 2016
(IoE Research Ethics Committee representative)*

* A decision to allow a project to proceed is not an expert assessment of its content or of the possible risks involved in the investigation, nor does it detract in any way from the ultimate responsibility which students/investigators must themselves have for these matters. Approval is granted on the basis of the information declared by the applicant.

University of Reading

Researcher:
Ms Yun Wang
Tel:00-44-7778723910
Email:
y.wang19@pgr.reading.ac.uk

Supervisor:
Prof.Jeanine Caroline Treffers-Daller
Tel: 00-44-118-3782690
Email:j.c.treffers-daller
@reading.ac.uk

Student information sheet

Research Project: Individual differences in listening comprehension among adult Chinese learners of English

Project Team Members: Ms Yun Wang

We would like to invite you to take part in a research project exploring individual differences in listening comprehension among Chinese learners of English.

What is the study about?

This PhD research project is designed to explore individual differences in listening comprehension among adult Chinese learners of English and the extent to which the individual differences explain the variance in L2 listening comprehension performance. It is important to do a listening comprehension study because listening is a source of frustration to learners and it is an area in which it seems difficult to make progress. Studies on the roles of linguistic knowledge components as well as speed or fluency factors have been undertaken quite often for reading, writing, and speaking, however, a limited number of comprehensive studies have been done on the roles of individual differences in listening comprehension. Although previous studies indicate that linguistic knowledge applied could explain some variance in listening comprehension, it is not clear whether individual differences in sentence processing efficiency with spoken input, working memory capacities and reasoning abilities will explain the variance in the success of listening comprehension.

Why have I been chosen to take part?

You have been invited to take part because you are studying university English and will take CET4 test. Your participation will provide valuable information on individual differences in listening comprehension which we can refer to and will do great contributions to our study aiming to explore what individual differences affect L2/FL learners' listening comprehension. According to my teaching and research experiences, students are very cooperative in such studies. We believe you would like to cooperate as usual.

University of Reading

Researcher:
Ms Yun Wang
Tel:00-44-7778723910
Email:
y.wang19@pgr.reading.ac.uk

Supervisor:
Prof.Jeanine Caroline Treffers-Daller
Tel: 00-44-118-3782690
Email:j.c.treffers-daller
@reading.ac.uk

Do I have to take part?

No, not at all. Also, you can stop helping us with my project at any time, without giving any reason. Just tell me yourself by contacting me on the following email address: y.wang19@pgr.reading.ac.uk.

What will happen if I take part?

You will take tasks aiming to measure your listening comprehension proficiency, linguistic knowledge (such as phonological knowledge, vocabulary knowledge and grammatical knowledge), sentence processing speed, working memory capacities, and reasoning abilities and to help us know about your personal background. The tasks will be held outside of normal class time with the supervision of the researcher and will last four hours in total. I will give written feedback to you on how to improve your listening comprehension.

Will anyone know about my answers?

Only the people working on the project will know about your answers. I won't tell your university how you perform in the tasks.

I anticipate that the findings of the study will be useful for explaining individual differences in English listening comprehension.

What will happen with the data?

Any data collected will be held in strict confidence and no real names will be used in any subsequent publications. The records of this study will be kept private. No identifiers linking you, the students or the school to the study will be included in any sort of report that might be published. Participants will be assigned a number and will be referred to by that number in all records. Research records will be stored securely in a locked filing cabinet and on a password-protected computer and only the research team will have access to the records. The data will be destroyed securely once the findings of the study are written up, after five years. The results of the study may be presented at national and international conferences, and in written reports and articles. An electronic summary of the findings of the study will be made available to you by contacting the Principal Researcher.

University of Reading

Researcher:
Ms Yun Wang
Tel:00-44-7778723910
Email:
y.wang19@pgr.reading.ac.uk

Supervisor:
Prof.Jeanine Caroline Treffers-Daller
Tel: 00-44-118-3782690
Email:j.c.treffers-daller
@reading.ac.uk

Who has reviewed the study?

This project has been reviewed following the procedures of the University Research Ethics Committee and has been given a favourable ethical opinion for conduct. The University has the appropriate insurances in place. Full details are available on request.

What happens if I change my mind?

You can change your mind at any time you like. If you change your mind after data collection has ended, we won't use your data.

What happens if something goes wrong?

In the unlikely case of concern or complaint, you can contact Professor Jeanine Caroline Treffers-Daller at University of Reading by phone on 00-44-118-3782690or by email on j.c.treffers-daller@reading.ac.uk.

Where can I get more information?

If you would like more information, please contact Yun Wang by phone on 00-44-7778723910 or by email on y.wang19@pgr.reading.ac.uk.

What do I do next?

I do hope that you will agree to your participation in the study. If you do, please complete the attached consent form and return it to us.

Thank you for your time.

Yours sincerely,

Yun Wang

University of Reading

Researcher:
Ms Yun Wang
Tel:00-44-7778723910
Email:
y.wang19@pgr.reading.ac.uk

Supervisor:
Prof.Jeanine Caroline Treffers-Daller
Tel: 00-44-118-3782690
Email:j.c.treffers-daller
@reading.ac.uk

Research Project: Individual differences in listening comprehension among adult Chinese learners of English

IF YOU ARE HAPPY TO TAKE PART THEN PLEASE FILL IN THE FORM BELOW

Student Consent Form

I wish to take part in the research.

Name: ______________________________

Name of university: ______________________

Please tick as appropriate:

I consent to my involvement in the project as outlined in the information sheet. ☐

Signed:____________________

Date: ______________________

University of Reading

博士研究生
王云
Tel:00-44-7778723910
Email:
y.wang19@pgr.reading.ac.uk

指导老师
Jeanine Caroline Treffers-Daller 教授
Tel: 00-44-118-3782690
Email: j.c.treffers-daller
@reading.ac.uk

学生信息告知书

研究课题：中国成人英语学习者个性差异对听力理解的影响（Individual differences in listening comprehension among adult Chinese learners of English）

课题组成员：王云

亲爱的同学：

您好！

我诚挚地邀请您参加我的博士研究课题：中国成人英语学习者个性差异对听力理解的影响。

课题研究内容：

我的博士研究课题旨在探索中国成人英语学习者听力理解中的个性差异，以及个性差异对学习者英语听力理解的影响程度。听力理解是英语学习者的一大难点，许多英语学习者发现自己很难提高英语听力理解的水平，因此，英语听力理解研究显得尤为重要。虽然目前国内外研究者已经进行了大量的研究， 发现了第二语言/外语学习者的语言知识以及语言处理速度或流利度等因素在第二语言/外语阅读、写作及口语中的影响，但是关于学习者的个性差异对听力理解影响的综合性研究却非常有限。尽管研究显示学习者的第二语言/外语语言知识影响听力理解，但是其他的个性差异如：第二语言/外语听力语句处理速度、工作记忆容量、推理能力等对听力理解的影响仍不清楚。

邀请您参加我的博士研究课题的原因：

您正在学习大学英语，并将参加英语 4 级考试。您的参与将为本研究提供宝贵的参考数据，也将有助于我们研究第二语言/外语学习者听力理解中个性差异的影响。根据我多年的教学和科研经历，学生在类似的研究中一直具有合作精神。我们坚信您会一如既往地表现出合作精神。

您是否必须参加：

您随时可以通过以下邮箱联系我 y.wang19@pgr.reading.ac.uk 收回您的同意书。

University of Reading
博士研究生
王云
Tel:00-44-7778723910
Email:
y.wang19@pgr.reading.ac.uk

指导老师
Jeanine Caroline Treffers-Daller 教授
Tel: 00-44-118-3782690
Email: j.c.treffers-daller
@reading.ac.uk

您如何参加本研究课题：

您将完成一系列英语听力理解任务，旨在了解中国成人英语学习者在听力理解水平、语言知识（如语音知识、词汇知识及语法知识）、语句处理速度、工作记忆容量、推理能力等方面的个性差异，并回答一个关于学习者外语语言学习及使用的背景调查问卷。为了不影响您正常的上课，我将安排您利用课外时间完成听力任务和问卷回答，计划用时 4 小时。我会对您提供书面的反馈，并就如何提高您的听力水平提出建议。

您的信息不会被泄露：

只有课题组成员有权查看您的信息。

我们期待通过本研究能够发现影响第二语言/外语学习者听力理解的个性差异。

我将如何保护所收集的数据：

所有数据都将严格保密，在随后发表的相关论文或著作中，我不会透露任何学生的真实姓名。研究记录将被妥善保管。所有学生的记录都将被编号，并通过编号对记录进行查看。所有研究记录都将存放于保险柜中及有密码保护的电脑中，只有我本人有权查看记录。研究结果一旦被发表， 数据将在五年后被安全处理。研究结果可以在国内及国外的会议上进行交流，或以报告及论文形式公开发表。

研究计划已通过审核：

本博士研究课题已经通过了雷丁大学道德委员会的审核并获准实施。雷丁大学能够确保其安全性。详情可从雷丁大学获取。

如果您改变主意：

您可以在任何时间改变主意。如果您在数据收集工作结束后改变主意，我们将不采用您的数据。

如有其他相关事情：

如有其他关于本研究课题的诸如投诉之类的事情，您可以通过电话 00-44-118-3782690 或邮件 j.c.treffers-daller@reading.ac.uk 联系雷丁大学的 Jeanine Treffers-Daller 教授。

University of Reading

博士研究生
王云
Tel:00-44-7778723910
Email:
y.wang19@pgr.reading.ac.uk

指导老师
Jeanine Caroline Treffers-Daller 教授
Tel: 00-44-118-3782690
Email: j.c.treffers-daller
@reading.ac.uk

如何获取更多相关信息：

您可以通过电话 00-44-7778723910 或邮件 y.wang19@pgr.reading.ac.uk 联系我本人。

下一步工作：

我衷心希望您同意参加我的博士研究课题。如果您同意，请在随后附上的知情同意书上签名并将其发送给我。

此致！

敬礼！

王云
2016 年 9 月 9 日